The challenges and opportunities of assessing low-value care provision in the Netherlands

Joris Müskens

The challenges and opportunities of assessing low-value care provision in the Netherlands

Joris Müskens

Colofon

The challenges and opportunities of assessing low-value care provision in the Netherlands

The work presented in this thesis was carried out within the Radboud Institute for Health Sciences

Design/Lay-out Proefschriftenbalie, Nijmegen

Print Ipskamp Printing, Nijmegen

ISBN 978-94-6473-480-5

The challenges and opportunities of assessing low-value care provision in the Netherlands

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op vrijdag 7 juni 2024
om 12.30 uur precies

door

Joris Lambertus Jozef Maria Müskens
geboren op 8 september 1992
te Nijmegen

Promotoren
Prof. dr. R.B. Kool
Prof. dr. G.P. Westert

Copromotor
Dr. S.A. van Dulmen

Manuscriptcommissie
Prof. dr. M.G.M. Olde Rikkert
Prof. dr. R.A. Verheij (Tilburg University)
Prof. dr. D.L.M. Zwart (Universiteit Utrecht)

Table of Contents

CHAPTER 1

General introduction

Over the past few decades, the movement towards evidence-based medicine has become increasingly prominent in healthcare systems worldwide. [1, 2] Evidence-based medicine combines the best available scientific knowledge with the clinical experience of healthcare professionals and the values of patients to achieve the best possible medical decisions. [3] However, the implementation of evidence-based medicine into practice remains challenging. Despite physicians' best efforts to provide the best care possible for patients, studies indicate that low-value care is (still) ubiquitously provided across health care systems worldwide. [4-9] Moreover, they also show that significant variation in low-value care provision among healthcare providers exists. [10-12] Low-value care is generally defined as care which offers little to no net benefit for the patient with respect to its harms, costs, alternatives or the patients' preferences. [13-15] Estimates of the prevalence of low-value care range between 20% and 30% of total healthcare spending. [4, 6, 16] The provision of low-value care can contribute to over-medicalization of healthcare and can result in a cascade of low-value testing or treatments which could result in harm for patients. [17, 18] Its reduction is therefore considered an important step towards improving quality of healthcare. [7, 19, 20] In order to start reducing low-value care, one must first know if and where the problem is present. Obtaining insight into the presence of low-value care is often a first step in initiating a discussion regarding the necessity of reducing certain healthcare procedures. [21, 22]

International assessments of low-value care

Multiple international studies report large variations in low-value care provision between services, even among assessments conducted within the same countries and populations. [4-6, 11, 23-30] For example, in the study by Schwartz et al. the proportion of the low-value care counts among Medicaid beneficiaries ranged from 0 to 14% across the 26 services analysed. [6] Most insights into the presence of low-value care to date have been obtained from the United States [6, 10, 23, 31, 32], Canada [24, 27, 33, 34], and Australia. [11, 25, 29, 35] In the Netherlands, the number of assessments was limited at the start of this thesis.

The existing (international) assessments report largely different and varying outcomes. Various general and methodological factors could provide an explanation for the observed variation in assessment outcomes. First, it is important to recognize that the prevalence of both medical conditions and the opportunity of (low-value) healthcare provision greatly varies within and between countries. [36, 37] The present variation can partly be attributed to differences between healthcare systems, such as the insurance models in effect or general accessibility of healthcare. [37-42] Second, the observed variation in low-value care provision could also be caused by differences in population characteristics such as age, gender or socioeconomic status. Previous research has already indicated that gender, socioeconomic status and age of an individual all affect the amount of care patients require, receive, and have access to. [43-49] Third, methodological differences between studies are probably the main source which could explain the majority of the

variation observed among both international and local assessments. Decisions regarding the assessment methods, the included population, and operationalization are generally the factors that have the largest impact on the reported assessment outcomes. The use and existence of a wide range of different assessment methods make the assessment of low-value care prone to several challenges. These should be carefully considered when performing an assessment of low-value care.

Challenges in the assessment of low-value care

There are three factors that make the assessment of low-value care challenging: the definitions of low-value services used, the available data set and assessment methods used. Below these three main challenges are discussed in more depth.

1 | Data definitions of low-value care used

Generally, the assessment of low-value care starts with picking and defining the type of low-value care you aim to examine. It is important to keep in mind that not all types of (low-value) care are equally accessible for assessment. Only a few types of tests or treatments are universally beneficial or entirely ineffective and post a risk to all patients should they be delivered. In case of this small group, definitions of low-value care are easily defined. [5] However, most types of care fall into the so-called '*grey zone*'. The grey zone includes services that offer little benefit to most patients, those for which the balance between benefits and harms varies substantially among patients and/or the many treatments and services that are backed by little evidence. [5] This nuanced character of low-value care makes it difficult to obtain an accurate definition of when a treatment or service should be considered as low-value. This is specifically because the appropriateness of a service is often defined by the clinical scenario in which it is used, rather than being ubiquitously coupled to the use of an individual test or procedure. [50] Almost no healthcare service can be considered of low-value in all clinical scenarios, but only for a part of the patients receiving it. These nuances should preferably all be taken into account when conducting and interpreting an assessment. Additionally, the patients' values and preferences should also be taken into consideration in defining low-value care. [51-54] Research shows that patients who are presented with a selection of treatments that may offer potential benefits will have differing views or opinions regarding the advantages and disadvantages of each option. [5, 52]

Most assessments of low-value care use (clinical) guidelines to obtain their definitions. Although, these guidelines are a useful source for finding a detailed definition of low-value care, unfortunately most of the recommendations presented are unfit for use in assessments of low-value care. These recommendations often include a highly specific description of the relevant population, requiring information which is not present in the available databases to accurately select the relevant patients. Examples are details

regarding the presence of "*locking*" or "*catching*" sensation in the knee, information which is required to distinguish whether or not a knee arthroscopy is indicated in case of knee arthrosis in patients above 50 years of age. Other examples of specific description are subjective information regarding the severity of a condition, the patient's preferences, symptom duration or general patient characteristics like the patient's smoking habits. [55-58] Additionally, recommendations often contain terms that do not map directly to data variables; also, diagnosis and procedure codes may not precisely identify patients for whom care is of low value. This introduces an inherent uncertainty in identifying if a treatment would be of low-value. [29]

2 | Data(bases) used in the assessment of low-value care

In order for data to be suitable for the assessment of low-value care, they need to contain sufficient clinical information. [56] However, the level of clinical information differs among the data sources used in the assessment of low-value care and not all available data are equally suitable for the assessment of low-value care. [59] Commonly, two types of data are distinguished and used within the assessment of low-value care: registration and claims data. Registration data are characterised by the presence of sufficient clinical information to distinguish appropriate from inappropriate care, and can be further subdivided into clinical and administrative data. The difference between clinical and administrative databases lies within the purpose for which they were collected. Clinical databases contain information that is directly extracted from electronic patient records, and aim to provide an extensive overview of the diagnosis and treatment history of patients. Administrative data are obtained from records of service utilisation and payments for payer or hospital billing purposes. Administrative data contain large amounts of information on diagnoses, medical procedures, resource utilisation, and costs or charges and are primarily collected for financial and administrative management. [60, 61] Although both clinical and administrative databases contain the required clinical information for the accurate assessment of low-value care, neither were originally designed for it. [21, 62] It is therefore often difficult to gain access to such data, while these often contain highly detailed personal information which should not be easily accessible to the public (such as information regarding a patient's health status or financial situation). [63, 64]

Conversely, claims data are often more readily available and encompass a large proportion of the population. Nevertheless, claims data are limited with respect to the amount of clinical detail they contain, which is too limited to accurately distinguish appropriate from inappropriate care. Claims data are therefore generally used to assess the utilisation rates of different types of care, or to indirectly assess low-value care. [21, 56]

Although a substantial amount of information is available for and used in the assessment of low-value care, in most cases it is not enough to be able to perform an accurate assessment. Not all of the required information to perform an assessment is recorded in

the available databases. Therefore, researchers often have to make assumptions in order to perform their assessments of low-value care.

3 | Methods used to assess the volume of low-value care

Assessments of low-value care can be divided into two distinct categories: indirect and direct assessments of low-value care. [5, 32] Indirect measures of low-value care examine unwarranted geographical variations in prevalence of procedures and care intensity without evaluating the appropriateness of the care delivered. The observed variations often do not directly demonstrate the provision of low-value care, but rather give an indication as to how treatment or healthcare use differs geographically. [15, 56] For example, geographic variation analysis examines rates of services in different areas, with the interpretation that higher than average rates may partly indicate overuse or low-value care. In 2014 the Organisation for Economic Co-operation and Development (OECD) examined geographic variations in the utilisation of 10 health services among 13 countries. The report showed that large variation in hospital medical admission rates existed, even after adjusting for population differences. For example, knee replacement rates were found to be four times higher in Australia, Switzerland, Finland, Canada and Germany when compared to Israel or Portugal (e.g. 200 knee replacements per 100,000 vs. 56 and 75 per 100,000 people). [42] However, as no judgement can be made regarding the appropriateness of the individual tests or treatments, it remains uncertain whether this variation actually represents the provision of low-value care. Conversely, direct measures of low-value care most often require patient-level data containing information on both the service and diagnosis, enabling the distinction between appropriate and inappropriate care. When the patient characteristics do not comply with the indications for a treatment, the provided care can be considered of low-value. Direct measures enable low-value care to be mapped on a detailed level, thereby providing reliable insight into its presence.

For direct measures of low-value care to be reliable generally two requirements must be met. First, a clear and unambiguous definition of low-value care must be present, which ideally is widely accepted and understood by all relevant stakeholders. Definitions of low-value services are therefore often derived from evidence-based or consensus-based guidelines, or through a multidisciplinary iterative process with the involved (medical) stakeholders. [5] Second, as previously discussed, the information required to distinguish low-value care should be present in the data. [5, 21, 58]

The use of a clear and unambiguous definition of low-value care is important because it determines the numerator and the type of assessment lens used. In 2017, Chalmers et al., proposed that assessments quantifying low-value care can be categorised into two types based on the denominator that is used. [65] Assessment using a patient lens report the proportion of patients receiving the low-value service, while the service lens reports the proportion of services considered to be of low-value. Depending on whether only

patients who had the potential to receive the examined type of low-value care or the entire cohort are included in the denominator, the patient lens can be further subdivided into the patient-indication and patient-population lens. For example, when examining the low-value use of imaging in case of lower back pain among the entire population of a GP practice, a patient-population lens is applied as the entire patient cohort of the practice is included in the denominator. When an assessment examines the use of low-value imaging among patients with lower back pain, only the patients with lower back pain are included in the denominator. In this case a patient-indication lens is used, while only patients with a specific indication are included in the denominator (e.g. patients with lower back pain). When using the service lens, all services (in this case imaging procedures) within said practice are included and examined with respect to the appropriate indications. Implying that when one patient has undergone multiple imaging studies, each of these will be included separately in the denominator of the assessment. Figure 1 provides a schematic depiction of the concept of assessment lenses.

The use of these different assessment lenses can have a significant impact on the outcomes reported as they affect the magnitude of the denominator. Thereby warranting cautiousness when comparing different assessments of low-value care, even when they examine the same type of care within similar settings. [65]

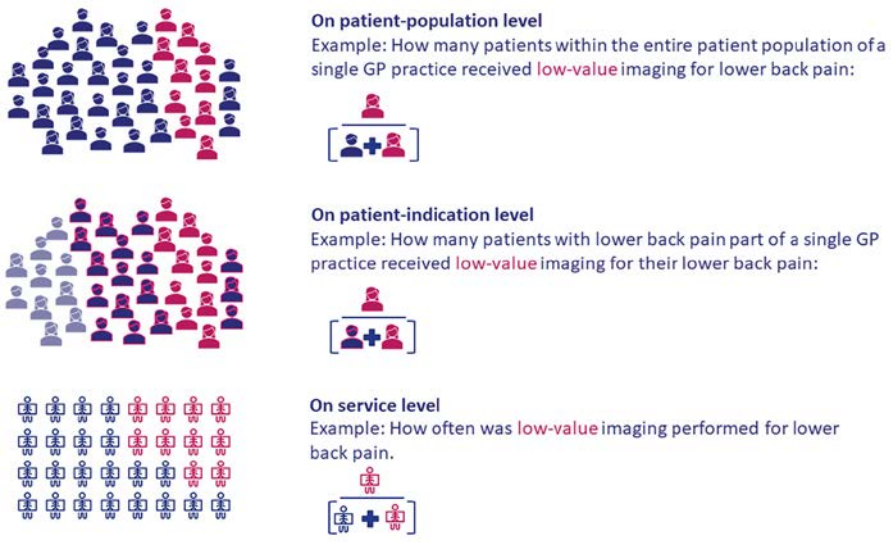


Figure 1 | Overview of the different assessment lenses used in the assessment of low-value care

Overall, the abovementioned factors make the assessment of low-value care a complicated endeavour. To emphasise the impact of the abovementioned factors, table 1 shows a breakdown of the abovementioned factors for three studies regarding low-value imaging patients with lower back pain.

Table 1 | Breakdown of the used definition of low-value care, included population, data source, applied lens and their assessment outcomes of three assessments examining low-value imaging in case of lower back pain.

	Bouck et al., 2019 [24]	**Mafi et al., 2017** [31]	**Schwartz et al., 2014** [6]
Country	Canada	United States	United States
Definition of LVC service	Imaging for lower back pain within the first six weeks of symptom onset, in the absence of red flags. Red flags include suspected epidural abscess or hematoma presenting with acute pain, but no neurological symptoms; suspected cancer; suspected infection; cauda equina syndrome; severe or progressive neurologic deficit; and suspected compression fracture. In patients with suspected uncomplicated herniated disc or spinal stenosis, imaging is only indicated after at least a six-week trial of conservative management and if symptoms are severe enough that surgery is being considered.	Do not do imaging for lower back pain within the first six weeks, unless red flags are present. In this study an imaging procedure was deemed to be low value if it was either potentially or very likely low value, according to the expert guidelines.	No diagnoses in claim warranting imaging (e.g. radiologic, CT, and MRI imaging of spine); imaging occurred within 6 week of the first diagnosis of lower back pain. (Exclusion diagnoses include cancer, trauma, intravenous drug abuse, neurological impairment, endocarditis, septicaemia, tuberculosis, osteomyelitis, fever, weight loss, loss of appetite, night sweats, and anaemia)
Included population	Adult patients (≥18 years old) in Alberta with non-persistent lower back. Excluding patients with persistent lower back pain or with a history of lower back pain, related imaging or surgery, or other red flags within 12 months prior to the index visit.	All members 65 years of age and younger with a low back imaging service and diagnosis of lower back pain within 6 weeks prior to the low back imaging.	Patients with lower back pain continuously enrolled in Parts A and B of traditional fee-for-service Medicare that were living in the United States or Washington, DC, and were at least 65 years old.
Data source	Claims data	Administrative data	Claims data
Lens	Patient-indication	Service	Patient-population
Assessment outcome	30.7% of patients underwent potentially unnecessary imaging (X-ray and/or CT/MRI) within six months of their initial visit regarding lower back pain despite the absence of red flags.	86.2% of the imaging procedures examined were considered of low-value.	14% of the identified imaging procedures related to non-specific lower back pain were deemed to be of low-value.

Existing frameworks for the assessments of low-value care

Although the assessment of low-value care plays an important role in its reduction, standardisation of assessment methodologies is lacking. To date, most frameworks focus on measuring waste within healthcare systems [66, 67] or on addressing low-value care through changing patients' and physicians' behaviour. [22, 68, 69] Although most of these frameworks emphasise the necessity of measuring the extent of the problem, only one framework from Miller et al. provides a method of assessing low-value care. [56] Even though this framework provides a good first description of how assessment of low-value care could be conducted, its application is limited. This framework aims to provide estimates of the total expenditure of low-value care within a population, rather than providing tools and methods to assess the volume of low-value care provision. It does contain some description of three approaches through which the volume of low-value practices could be assessed, but does not elaborate on how these should be performed. Nor does it elaborate on the potential pitfalls to which each of these methods are prone. Second, it proposes that the identified proportion of low-value treatment (e.g. expenditure) among the high-expenditure services should be considered as a predictor of the proportion of low-value care among low-expenditure services. However, as previously mentioned, low-value utilisation varies largely among services. It is therefore unlikely that by assuming similar low-value use of both high- and low-value services would yield a reliable assessment outcome. Furthermore, the frameworks' scope is focussed on performing assessments using claims data, which are not the most suitable data source available to distinguish appropriate from inappropriate care. Overall, it appears that no framework exists that provides tools to correctly assess low-value care. The lack of a framework has resulted in studies performing assessments of low-value care according to their own methodologies, and therefore yielding their own distinct outcomes.

Goal and outline of this thesis

We conclude that several challenges exist in the assessment of low-value care, that have not been taken into account in most assessments performed to date. The general aims of this thesis, were to explore the methods used to assess low-value care and gain insight into the presence of low-value care in the Netherlands, considering those challenges. With these aims in mind, we conducted several assessments of low-value care within the Dutch healthcare setting. These aims resulted in the following research questions:

1 How is the volume of low-value care assessed in the current literature, and how do differences impact assessment outcomes? To gain insight into the methods used, we performed an extensive examination of the methods used in different studies assessing the volume of low-value diagnostic testing in the international literature and compared their outcomes. The outcomes of this examination are presented in chapter 2.

2 What is the actual volume of different low-value care diagnostic practices and treatments in the Netherlands? Following our examination of the methods used to assess the volume of low-value care, we have put our newly gained insight into practice. We performed multiple assessments among Dutch general practitioners (chapter 3, 4 and 6) and medical specialists (chapter 5) in order to gain insight into the magnitude of low-value care in the Netherlands. The assessed clinical procedures and problems were chosen in collaboration with medical professionals from the field, based on their perceived volume and relevance for practice.
3 What is the effect of an intervention on the volume of low-value care? Following the findings from our assessment among Dutch general practitioners, we developed an intervention to improve the (low-value) chronic prescription of acid reducing medication (chapter 7).

References

1. Claridge JA, Fabian TC. History and development of evidence-based medicine. World J Surg. 2005;29(5):547-53.
2. Smith R, Rennie D. Evidence based medicine—an oral history. BMJ : British Medical Journal. 2014;348:g371.
3. Tenny S, Varacallo M. Evidence Based Medicine. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.; 2023.
4. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System: Estimated Costs and Potential for Savings. JAMA. 2019;322(15):1501-9.
5. 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.
6. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring Low-Value Care in Medicare. JAMA Internal Medicine. 2014;174(7):1067-76.
7. Oakes AH, Radomski TR. Reducing Low-Value Care and Improving Health Care Value. Jama. 2021;325(17):1715-6.
8. Oakes AH, Sen AP, Segal JB. The impact of global budget payment reform on systemic overuse in Maryland. Healthc (Amst). 2020;8(4):100475.
9. Powers BW, Jain SH, Shrank WH. De-adopting Low-Value Care: Evidence, Eminence, and Economics. JAMA. 2020;324(16):1603-4.
10. Schwartz AL, Zaslavsky AM, Landon BE, Chernew ME, McWilliams JM. Low-Value Service Use in Provider Organizations. Health Serv Res. 2018;53(1):87-119.
11. Badgery-Parker T, Feng Y, Pearson S-A, Levesque J-F, Dunn S, Elshaug AG. Exploring variation in low-value care: a multilevel modelling study. BMC Health Services Research. 2019;19(1):345.
12. Do LA, Koethe BC, Daly AT, Chambers JD, Ollendorf DA, Wong JB, et al. State-Level Variation In Low-Value Care For Commercially Insured And Medicare Advantage Populations. Health Aff (Millwood). 2022;41(9):1281-90.
13. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. Med Care Res Rev. 2017;74(5):507-50.
14. Elshaug AG, Rosenthal MB, Lavis JN, Brownlee S, Schmidt H, Nagpal S, et al. Levers for addressing medical underuse and overuse: achieving high-value health care. Lancet. 2017;390(10090):191-202.
15. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. Med J Aust. 2015;203(4):179-81.
16. Berwick DM, Hackbarth AD. Eliminating Waste in US Health Care. JAMA. 2012;307(14):1513-6.
17. Ganguli I, Simpkin AL, Lupo C, Weissman A, Mainor AJ, Orav EJ, et al. Cascades of Care After Incidental Findings in a US National Survey of Physicians. JAMA Netw Open. 2019;2(10):e1913325.
18. Davenport MS. Incidental Findings and Low-Value Care. AJR Am J Roentgenol. 2023:1-7.

19. Grimshaw JM, Patey AM, Kirkham KR, Hall A, Dowling SK, Rodondi N, et al. De-implementing wisely: developing the evidence base to reduce low-value care. BMJ Quality & Safety. 2020;29(5):409-17.
20. Nilsen P, Ingvarsson S, Hasson H, von Thiele Schwarz U, Augustsson H. Theories, models, and frameworks for de-implementation of low-value care: A scoping review of the literature. Implementation Research and Practice. 2020;1:2633489520953762.
21. Marcotte LM, Schuttner L, Liao JM. Measuring low-value care: learning from the US experience measuring quality. BMJ Quality & Safety. 2020;29(2):154-6.
22. Bhatia RS, Levinson W, Shortt S, Pendrith C, Fric-Shamji E, Kallewaard M, et al. Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care. BMJ Qual Saf. 2015;24(8):523-31.
23. Mafi JN, Reid RO, Baseman LH, Hickey S, Totten M, Agniel D, et al. Trends in Low-Value Health Service Use and Spending in the US Medicare Fee-for-Service Program, 2014-2018. JAMA Network Open. 2021;4(2):e2037328-e.
24. Bouck Z, Pendrith C, Chen XK, Frood J, Reason B, Khan T, et al. Measuring the frequency and variation of unnecessary care across Canada. BMC Health Serv Res. 2019;19(1):446.
25. de Oliveira Costa J, Pearson S-A, Elshaug AG, van Gool K, Jorm LR, Falster MO. Rates of Low-Value Service in Australian Public Hospitals and the Association With Patient Insurance Status. JAMA Network Open. 2021;4(12):e2138543-e.
26. O'Sullivan JW, Albasri A, Nicholson BD, Perera R, Aronson JK, Roberts N, Heneghan C. Overtesting and undertesting in primary care: a systematic review and meta-analysis. BMJ Open. 2018;8(2):e018557.
27. Chami N, Simons JE, Sweetman A, Don-Wauchope AC. Rates of inappropriate laboratory test utilization in Ontario. Clin Biochem. 2017;50(15):822-7.
28. Charlesworth CJ, Meath TH, Schwartz AL, McConnell KJ. Comparison of Low-Value Care in Medicaid vs Commercially Insured Populations. JAMA Intern Med. 2016;176(7):998-1004.
29. Badgery-Parker T, Pearson S-A, Chalmers K, Brett J, Scott IA, Dunn S, et al. Low-value care in Australian public hospitals: prevalence and trends over time. BMJ Quality & Safety. 2019;28(3):205-14.
30. Chalmers K, Smith P, Garber J, Gopinath V, Brownlee S, Schwartz AL, et al. Assessment of Overuse of Medical Tests and Treatments at US Hospitals Using Medicare Claims. JAMA Network Open. 2021;4(4):e218075-e.
31. Mafi JN, Russell K, Bortz BA, Dachary M, Hazel WA, Jr., Fendrick AM. Low-Cost, High-Volume Health Services Contribute The Most To Unnecessary Health Spending. Health Aff (Millwood). 2017;36(10):1701-4.
32. Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB. Choosing wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2015;30(2):221-8.
33. Bouck Z, Calzavara AJ, Ivers NM, Kerr EA, Chu C, Ferguson J, et al. Association of Low-Value Testing With Subsequent Health Care Use and Clinical Outcomes Among Low-risk Primary Care Outpatients Undergoing an Annual Health Examination. JAMA Intern Med. 2020;180(7):973-83.

34. Pendrith C, Bhatia M, Ivers NM, Mecredy G, Tu K, Hawker GA, et al. Frequency of and variation in low-value care in primary care: a retrospective cohort study. CMAJ Open. 2017;5(1):E45-E51.
35. Chalmers K, Pearson SA, Badgery-Parker T, Brett J, Scott IA, Elshaug AG. Measuring 21 low-value hospital procedures: claims analysis of Australian private health insurance data (2010-2014). BMJ Open. 2019;9(3):e024142.
36. Krumholz HM. Variations in Health Care, Patient Preferences, and High-Quality Decision Making. JAMA. 2013;310(2):151-2.
37. Weaver MR, Nandakumar V, Joffe J, Barber RM, Fullman N, Singh A, et al. Variation in Health Care Access and Quality Among US States and High-Income Countries With Universal Health Insurance Coverage. JAMA Netw Open. 2021;4(6):e2114730.
38. Naylor KB, Tootoo J, Yakusheva O, Shipman SA, Bynum JPW, Davis MA. Geographic variation in spatial accessibility of U.S. healthcare providers. PLoS One. 2019;14(4):e0215016.
39. Verkerk EW, Van Dulmen SA, Born K, Gupta R, Westert GP, Kool RB. Key Factors that Promote Low-Value Care: Views of Experts From the United States, Canada, and the Netherlands. Int J Health Policy Manag. 2021;11(8):1514-21.
40. Saini V, Garcia-Armesto S, Klemperer D, Paris V, Elshaug AG, Brownlee S, et al. Drivers of poor medical care. Lancet. 2017;390(10090):178-90.
41. Ganguli I, Morden NE, Yang C-WW, Crawford M, Colla CH. Low-Value Care at the Actionable Level of Individual Health Systems. JAMA Internal Medicine. 2021;181(11):1490-500.
42. OECD. Geographic Variations in Health Care2014.
43. Arpey NC, Gaglioti AH, Rosenbaum ME. How Socioeconomic Status Affects Patient Perceptions of Health Care: A Qualitative Study. J Prim Care Community Health. 2017;8(3):169-75.
44. Loef B, Meulman I, Herber G-CM, Kommer GJ, Koopmanschap MA, Kunst AE, et al. Socioeconomic differences in healthcare expenditure and utilization in The Netherlands. BMC Health Services Research. 2021;21(1):643.
45. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. Front Public Health. 2020;8:231.
46. Kalseth J, Halvorsen T. Health and care service utilisation and cost over the life-span: a descriptive analysis of population data. BMC Health Services Research. 2020;20(1):435.
47. Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
48. David JL, Kaplan HB. Gender, social roles and health care utilization. Applied Behavioral Science Review. 1995;3(1):39-64.
49. National Academies of Sciences E, Medicine, Health, Medicine D, Board on Health Care S, Committee on Health Care U, Adults with D. Chapter 2. Factors That Affect Health-Care Utilization. Health-Care Utilization as a Proxy in Disability Determination. Washington (DC): National Academies Press (US) Copyright 2018 by the National Academy of Sciences. All rights reserved.; 2018.

50. Soong C, Wright SM. Measuring overuse: a deceptively complicated endeavour. BMJ Quality & Safety. 2022;31(1):8-10.
51. van der Weijden T, van der Kraan J, Brand PLP, van Veenendaal H, Drenthen T, Schoon Y, et al. Shared decision-making in the Netherlands: Progress is made, but not for all. Time to become inclusive to patients. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2022;171:98-104.
52. Verkerk EW, Boekkooi JAH, Pels EGM, Kool RB. Exploring patients' perceptions of low-value care: An interview study. Patient Education and Counseling. 2023;111:107687.
53. Mulley AG, Trimble C, Elwyn G. Stop the silent misdiagnosis: patients' preferences matter. BMJ : British Medical Journal. 2012;345:e6572.
54. Sypes EE, de Grood C, Clement FM, Parsons Leigh J, Whalen-Browne L, Stelfox HT, Niven DJ. Understanding the public's role in reducing low-value care: a scoping review. Implementation Science. 2020;15(1):20.
55. Kim DD, Do LA, Daly AT, Wong JB, Chambers JD, Ollendorf DA, Neumann PJ. An Evidence Review of Low-Value Care Recommendations: Inconsistency and Lack of Economic Evidence Considered. J Gen Intern Med. 2021;36(11):3448-55.
56. Miller G, Rhyan C, Beaudin-Seiler B, Hughes-Cromwick P. A Framework for Measuring Low-Value Care. Value Health. 2018;21(4):375-9.
57. Chalmers K, Badgery-Parker T, Pearson S-A, Brett J, Scott IA, Elshaug AG. Developing indicators for measuring low-value care: mapping Choosing Wisely recommendations to hospital data. BMC Research Notes. 2018;11(1):163.
58. Elshaug AG, McWilliams JM, Landon BE. The Value of Low-Value Lists. JAMA. 2013;309(8):775-6.
59. Isaac T, Rosenthal MB, Colla CH, Morden NE, Mainor AJ, Li Z, et al. Measuring overuse with electronic health records data. Am J Manag Care. 2018;24(1):19-25.
60. Boncyk CS, Jelly CA, Freundlich RE. The Blessing and the Curse of the Administrative Database. Ann Am Thorac Soc. 2020;17(2):174-5.
61. Hashimoto RE, Brodt ED, Skelly AC, Dettori JR. Administrative database studies: goldmine or goose chase? Evid Based Spine Care J. 2014;5(2):74-6.
62. Marcotte LM, Zech JM, Liao JM. Key Features Underlying Low-Value Care Recommendations. Am J Med Qual. 2021;36(2):99-102.
63. Hübner UH, Egbert N, Schulte G. Clinical Information Systems - Seen through the Ethics Lens. Yearb Med Inform. 2020;29(1):104-14.
64. Price WN, 2nd, Cohen IG. Privacy in the age of medical big data. Nat Med. 2019;25(1):37-43.
65. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
66. Bentley TG, Effros RM, Palar K, Keeler EB. Waste in the U.S. Health care system: a conceptual framework. Milbank Q. 2008;86(4):629-59.
67. Smith GL, Hannah S, Mountford J, Bisognano M. Model to reduce waste in healthcare and add value. BMJ Open Qual. 2022;11(1).

68. Morgan DJ, Leppin AL, Smith CD, Korenstein D. A Practical Framework for Understanding and Reducing Medical Overuse: Conceptualizing Overuse Through the Patient-Clinician Interaction. J Hosp Med. 2017;12(5):346-51.
69. Parchman ML, Henrikson NB, Blasi PR, Buist DS, Penfold R, Austin B, Ganos EH. Taking action on overuse: Creating the culture for change. Healthc (Amst). 2017;5(4):199-203.

CHAPTER 2

Overuse of diagnostic testing in healthcare: a systematic review

Joris L.J.M. Müskens
Rudolf B. Kool
Simone A. van Dulmen
Gert P. Westert

Published in BMJ Quality & Safety, Volume 31, Issue 1, 2022

Abstract

Background: Overuse of diagnostic testing substantially contributes to healthcare expenses and potentially exposes patients to unnecessary harm. Our objective was to systematically identify and examine studies that assessed the prevalence of diagnostic testing overuse across healthcare settings to estimate the overall prevalence of low-value diagnostic overtesting.

Methods: PubMed, Web of Science, and Embase were searched from inception until February 18, 2020 to identify articles published in the English language that examined the prevalence of diagnostic testing overuse using database data. Each of the assessments was categorized as using a patient-indication lens, a patient-population lens or a service lens.

Results: 118 assessments of diagnostic testing overuse, extracted from 35 studies, were included in this study. Most included assessments used a patient-indication lens (n=67, 57%), followed by the service lens (n=27, 23%) and patient-population lens (n=24, 20%). Prevalence estimates of diagnostic testing overuse ranged from 0.09% to 97.5% (median prevalence of assessments using a patient-indication lens: 11.0%, patient-population lens: 2.0%, service lens: 30.7%). The majority of assessments (n=85) reported overuse of diagnostic testing to be below 25%. Overuse of diagnostic imaging tests was most often assessed (n=96). Among the 33 assessments reporting high levels of overuse (≥25%), the most frequently examined tests were preoperative testing (n=7) and imaging for uncomplicated lower back pain (n=6). For assessments of similar diagnostic tests, major variation in the prevalence of overuse was observed. Differences in the definitions of low-value tests used, their operationalization and assessment methods likely contributed to this observed variation.

Conclusion: Our findings suggest that substantial overuse of diagnostic testing is present with wide variation in overuse. Preoperative testing and imaging for nonspecific lower back pain are the most frequently identified low-value diagnostic tests. Uniform definitions and assessments are required in order to obtain a more comprehensive understanding of the magnitude of diagnostic testing overuse.

Introduction

In modern medicine, diagnostic tests, including laboratory tests, imaging, and more invasive procedures, figure prominently in clinical decision making surrounding a new diagnosis. [1, 2] However, the use of a diagnostic test is not always appropriate, as it may generate false positives, produce downstream cascades of more testing, expose patients to radiation or other harms, and create unnecessary patient anxiety, and could therefore be considered of low-value. [3-7] Recent studies show that low-value diagnostic tests are still widely used and account for a substantial portion of the total amount of low-value healthcare expenses. [8-12] However, despite the potential avoidance of both costs and patient harms, the full quantification of low-value diagnostic testing has been difficult to achieve.

Understanding the prevalence of low-value diagnostic testing is essential to spur doctors, health systems, and policymakers to take action to reduce its use. Most assessments of low-value diagnostic testing to date have been performed in the USA, Canada and Australia. [5, 13-17] Only a few assessments have been completed in Europe. [18-21] Although multiple assessments of diagnostic testing overuse exist, only a small fragment of the problem has been uncovered.

One systematic review assessing the prevalence of diagnostic testing overuse and underuse in the primary care setting has previously been published by O'Sullivan et al. [22] Previous assessments demonstrate that overtesting is not limited to the primary care setting. [16, 18, 19] We therefore chose not to limit our study to one healthcare setting, but rather to include all assessments of overtesting irrespective of the healthcare setting in which they were conducted. Furthermore, it is often hard to distinguish primary from secondary care practices due to differences in definitions of primary and secondary care procedures between countries and healthcare systems. In the present study, we therefore chose a more specific approach to the examination of the problem of low-value diagnostic testing compared with O'Sullivan. We narrowed down our scope to studies of similar study design that only quantified the overuse, and not underuse, of diagnostic testing (e.g. overtesting) using database data and used guidelines to distinguish appropriate from inappropriate testing to obtain a more uniform overview of the problem. This review might help policy makers and healthcare providers in their efforts to reduce overuse of diagnostic testing and can also help identify new knowledge gaps.

Methods

This systematic review was performed and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [23] (Supplementary file 1) and Meta-analyses Of Observational Studies in Epidemiology statements, [24] no protocol has been registered. PubMed, Web of Science, and EMBASE

were searched on February 18, 2020, for studies, of any design, assessing overuse of diagnostic tests. We did not restrict our search with respect to publication start date. The search can be summarised as: (Medical Overuse OR Low-value care OR wasteful care OR wasteful healthcare) AND (Diagnosis) AND (Variation OR Volume OR Prevalence OR Frequency) (see Supplementary file 2 for the full strategy). We limited our search to human studies and studies published in English. The reference list of each included study was also searched for potentially relevant studies.

Study selection

Full texts were independently screened for eligibility by two reviewers. We included studies that quantified the overuse of diagnostic tests using database data, described a prevalence assessment and mentioned the relevant guideline(s) used to distinguish appropriate from inappropriate diagnostic testing. For the purpose of this study we defined low-value diagnostic testing (or overtesting) as; the overuse of diagnostic practices which are unlikely to benefit the patient given the harms, cost, available alternatives, or preferences of the patient. [25] We excluded studies that did not quantify or assess provision of low-value diagnostic services; measured against a local guideline only (e.g. did not use a guideline published by a government or professional society, but rather used a guideline that is only applicable locally); used survey data as the principal data source; were not published in English; used data derived from countries not in the Organization for Economic Co-operation and Development (OECD); were an intervention study; or assessed (non-diagnostic) routine (population) screening tests as defined by Wald and Law. [26] We only included studies using data from countries that are part of the OECD because of the comparability of the social-economic characteristics of the populations. Disagreements regarding eligibility of studies were discussed by 3 of the study authors until consensus was reached.

Data extraction

The following information was extracted from each article: author and year, country, study population demographics (age/sex), the guideline used to determine the (in) appropriateness, the low-value care definition used, data sources, collected parameters, healthcare setting in which the assessment was conducted (primary/secondary care/ both/unclear), type of low-value diagnostic test examined, and the study outcome (prevalence estimate(s)). Assessments of diagnostic imaging procedure(s) were assigned to one of six categories based on the imaging modality they examined: Cardiac test, Combination, Endoscopy, Scan, Ultrasound and X-ray (see supplementary file 3 for an overview of the different imaging modalities in each category). The Combination category contains assessments of multiple imaging modalities, but which did not report the individual outcomes for each included modality. For example, some studies examined the use of X-ray, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) in the examination of lower back pain, but did not report the individual outcome measures

for the different modalities, but solely reported the combined outcome. When studies contained assessments for more than one unique diagnostic test, data for each test was collected and presented as an individual assessment. In case assessments were carried out over multiple time periods, only the data from the most recent time period were extracted. Each of the extracted assessments was assigned an assessment lens based on the classification proposed by Chalmers et al in 2017. [27] Chalmers et al. concluded that different lenses are used to assess low-value care, each of which produces distinct outcomes. In general, Chalmers distinguishes two types of lenses that are used: service- and patient-centric lenses. Assessments using the service lens focus on the proportion of diagnostic tests that are of low-value, while assessments using the patient lens focus on the proportion of the patient population that received the low-value diagnostic tests. Assessments using a patient centric lens can be further subdivided into assessments using either a patient-indication or patient-population lens. Which of the two patient centric lenses is applicable depends on the type of denominator that is used. [27] Assessment using a *Patient-indication* lens only include patients with a specific indication in their denominator, while assessments using a *patient-population* lens include the entire cohort in their denominator.

The process of assigning lenses to the different assessments was performed in the following manner. One of the authors drafted an initial proposal regarding the applicable lens for each of the included assessments. This proposal was then critically appraised by two other authors, which was followed by multiple rounds of discussion until all authors agreed on the lens used.

Quality assessment

Risk of bias was assessed by three researchers using a modified version of the Hoy risk of bias tool. The Hoy risk of bias tool is a validated tool for the assessment of both internal and external validity of prevalence studies. [28] The tool was modified in the following manner: 1) Three domains (points 4, 7 and 9) from the original tool were found to be not applicable with respect to the identified studies. These domains either required information which is not applicable to retrospective research involving (electronic) database data or examined study designs which were not included in our study. Domain 7 was considered to be not applicable, since we did not grade the underlying evidence of the guidelines used in each of the included assessments. These domains were therefore removed after internal discussion among three authors. Supplementary file 4 contains the original and modified tool, including more detailed reasoning for removal of each of the three domains.; 2) The wording was adjusted to reflect the prevalence of low-value diagnostic testing instead of the prevalence of disease. Studies were considered at high risk of bias when they scored at least two '*high*' and one '*unclear*' among the seven Risk of bias criteria. The original Hoy risk of bias tool does not provide a definition of high risk of bias. We therefore decided to use the abovementioned cut-off value for high risk after internal discussion among the authors. The process of grading risk of bias was similar

to the one we used to assign lenses to the different assessments. One author drafted an initial proposal regarding the risk of bias scores of the included studies, which was followed by critical appraisal by two other authors, and followed by multiple rounds of discussion until consensus was reached regarding the risk of bias score for each of the studies.

Statistical analysis

The primary outcome of this study is the prevalence of overuse of diagnostic tests across all healthcare settings. Descriptive statistics and median prevalences were calculated across all assessments, for diagnostic imaging, laboratory testing and electroencephalogram categories, and for the different assessment lenses within those categories. Analysis was performed using R V3.6.3, [29] and data visualization was done using the R package ggplot2. [30] Random-effect meta-analysis with 95% Confidence Intervals (CIs) (Clopper-Pearson), according to the DerSimonian and Laird method, was performed on similar assessments applying the same lens using the Meta, [31] and Metafor package. [32] Variance was stabilized using the double arcsine transformation. The among-study heterogeneity was assessed using the I^2 statistic. The I^2 statistic represents the percentage of total variation across studies that is attributable to heterogeneity rather than change. [33, 34] When applicable, data from similar assessments was pooled based on the lens that was used to assess the prevalence.

Results

Article characteristics

Our search strategy identified 2,542 articles. Of these, 2,459 were excluded based on the title or abstract. Thirty-four studies met the eligibility criteria and were included. One additional eligible study was identified through screening of the reference lists of the included studies. A PRISMA flow diagram of the selection procedure is shown in Figure 1. From the included studies, seven conducted their assessments in the primary care setting (7/35), five in the secondary setting (5/35) and nine in both settings (9/35). The remaining fourteen studies (14/35), did not provide a clear indication as to the setting in which their assessments were conducted and therefore labelled as '*unclear*' (also see supplementary file 6). The included studies were conducted in 8 different countries and contained 118 assessments of low-value diagnostic tests. Most studies were conducted in the U.S. (N=23). The 118 identified assessments are divided into imaging procedures (N=96), and other diagnostic tests (N=22), which included laboratory tests (N=19), and electro-encephalography procedures (N=3) (as shown in Table 1). The majority of the assessments used a patient-indication lens (N=67, 57%), followed by the service lens (N=27, 23%) and patient-population lens (N=24, 20%). Among the studies included, three studies assessed overuse among different insurance populations, [35-37] and one study

assessed overuse across two different time periods. [38] Of note, since we were interested in the most recent measurements of low-value diagnostic overtesting, we decided to only include the most recent measurements from the study by Flaherty et al. [38]

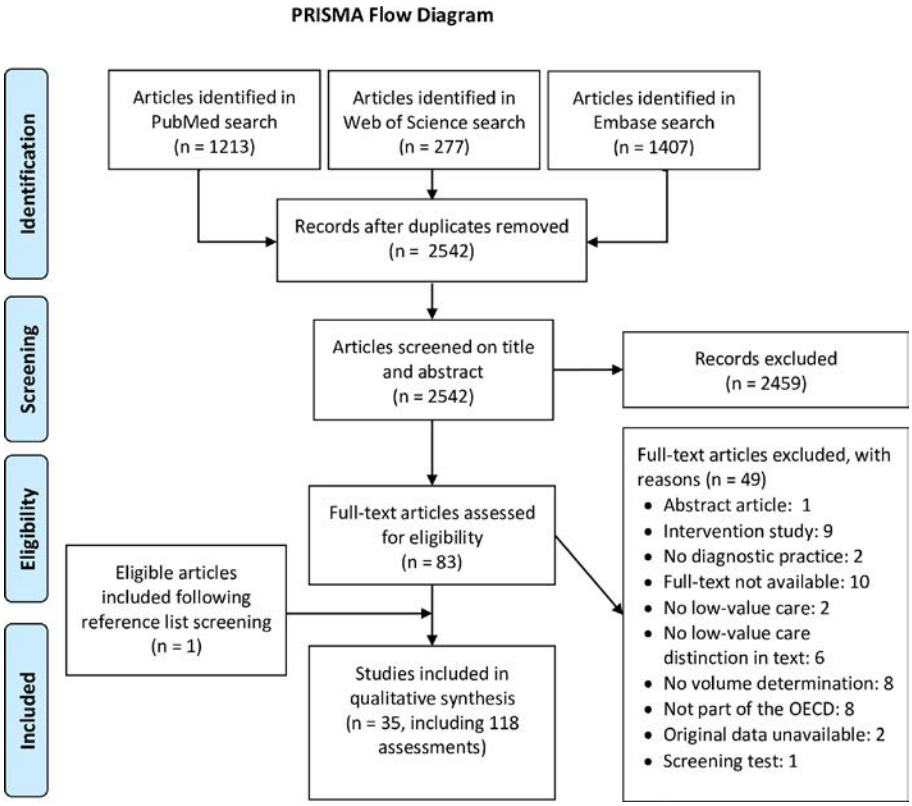


Figure 1 | PRISMA Flow-diagram. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

Risk of bias

Using the Hoy-Risk of bias tool, we assessed the risk of bias of the included studies based on 8 criteria (supplementary 4 contains the used modified Hoy-risk of bias tool). Assessment of Risk of bias revealed 25 studies as low risk of bias and 10 studies as high risk of bias (e.g. scoring at least two categories '*high*' and one '*unclear*'). Almost all studies graded as high risk of bias, scored as being of high risk on the following two criteria: '*the examined population being a close representation of the national population' and 'the use of a clear case definition of the low-value diagnostic test examined*'. Supplementary file 5 contains a detailed description of the risk of bias assessment outcome.

Table 1 | Overview of study characteristics.

Countries where the studies were conducted	**Number of studies**
Australia	3 (9%)
Austria	1 (3%)
Canada	4 (11%)
Italy	1 (3%)
Netherlands	1 (3%)
Spain	1 (3%)
Switzerland	1 (3%)
United States	23 (66%)
Total	35 (100%)
Type of diagnostic test	**Number of assessments**
Imaging	96 (81%)
Cardiac test	14 (12%)
Combination	14 (12%)
Endoscopy	11 (9%)
Scan	34 (29%)
Ultrasound	6 (5%)
X-ray	17 (14%)
Other diagnostic tests	22 (19%)
Laboratory tests	19 (16%)
Electroencephalography	3 (3%)
Total	118 (100%)
Type of assessment lens used	**Number of assessments**
Patient-indication	67 (57%)
Patient-population	24 (20%)
Service	27 (23%)
Total	118 (100%)

Overuse of diagnostic tests

Supplementary file 6 provides an overview of the studies, characteristics, and outcomes. Prevalence estimates of diagnostic testing overuse ranged from 0.09% to 97.5% (median prevalence of assessments using a patient-indication lens: 11.0%, a patient-population lens: 2.0%, a service lens: 30.7%). The majority of included assessments of low-value diagnostic testing (N=85) report overuse to be below 25%. Among the 33 assessments reporting high levels of overuse (≥25%), imaging for uncomplicated lower back pain (N=6) and preoperative testing (N=7), such as preoperative baseline lab tests, echocardiography or (cardiac) stress tests ,were most commonly assessed. Overuse of diagnostic imaging procedures was most often assessed (N=96), with prevalence of overuse varying between 0.09% and 97.5% (median prevalence of assessments using a patient-indication lens: 11.2%, a patient-population lens: 1.2%, a service lens: 22.0%) as is shown in figure 2. Prevalence assessments in the 'Other diagnostic tests' category (N=22)

varied between 0.10% and 78.6% as is shown in figure 3A-B. This category contained two distinct categories: laboratory tests (N=19, median prevalence of assessments using a patient-indication lens: 16.3%, a patient-population lens: 3.5 %, a service lens: 47.5%) and electroencephalography (N=3, median prevalence of assessments using a patient-indication lens: 0.2%, a patient-population lens: 0.1%).

The highest prevalence of overuse was reported in the following five diagnostic practices: use of electrocardiograms, chest x-rays, or pulmonary function tests in low-risk patients having low-risk surgery (97.5%); imaging for lower back pain within the first six weeks of symptom onset in absence of red flags (86.2%); knee arthroscopy for meniscal derangements (81.7%); baseline lab tests for low-risk patients receiving low risk surgery (78.6%); and knee arthroscopy for osteoarthritis (71.7%). Overall, imaging in case of nonspecific low back pain (15/118) and preoperative tests (14/118), such as preoperative baseline lab tests, echocardiography or exercise stress tests , were most often assessed diagnostic practices identified in this study. Figure 2 and 3 show that large variation in assessment outcomes of similar diagnostic tests, irrespective of assessment lens used, exists. For example, Bouck et al., [39] Schwartz et al., [13] and Mafi et al., [17] yield vastly different results in their respective studies. Bouck et al. [39] used a patient-indication lens and reported 30.70% of the identified imaging procedures to be considered as overuse, while Schwartz et al., [13] used a patient-population lens and found 4.1% to be considered as overuse. On the other hand, Mafi et al., used a service lens in their assessment and report the level of overuse to be 86.2%.

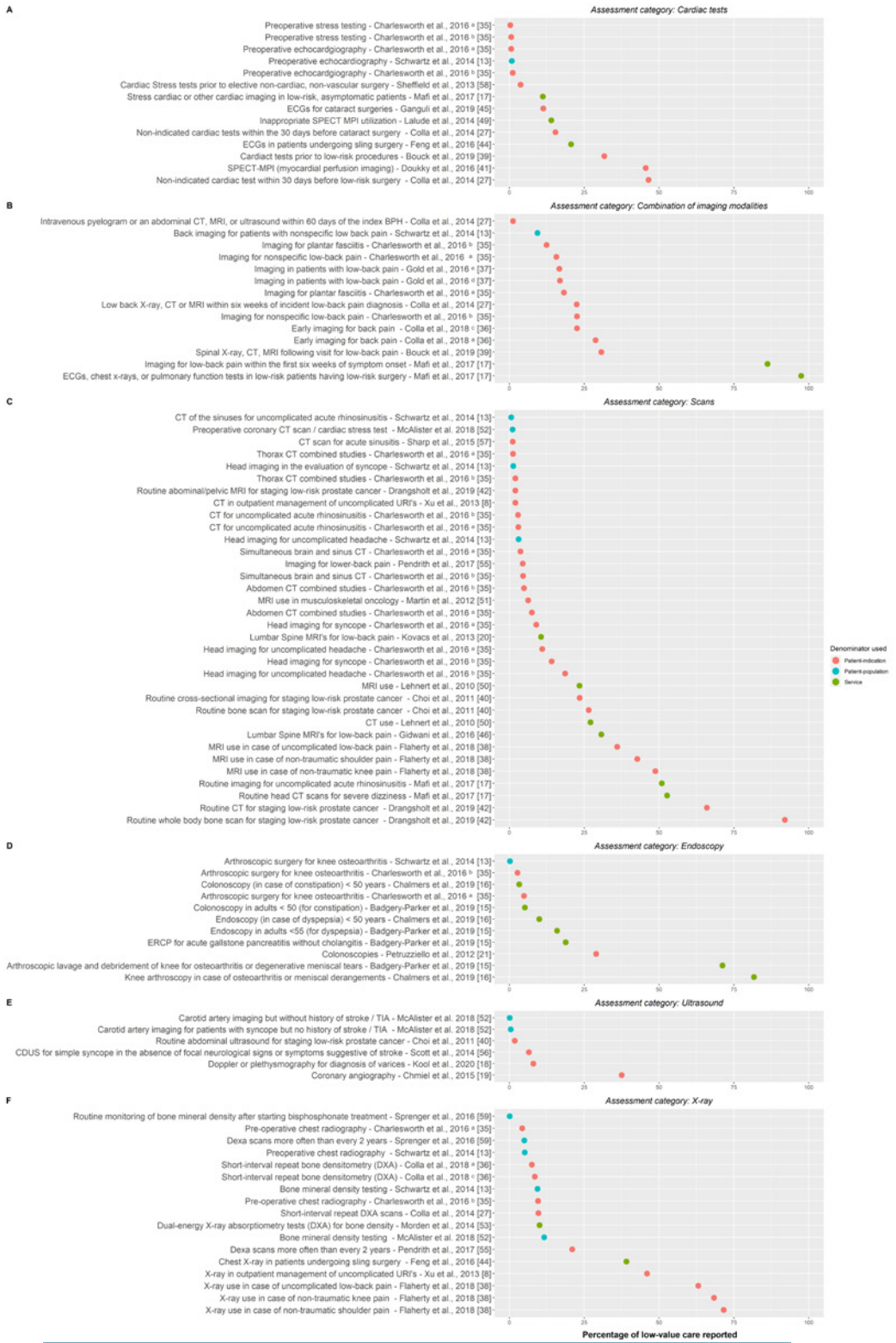


Figure 2 | Assessment outcomes regarding the prevalence of low-value diagnostic tests for all assessments included in the diagnostic imaging category: A) cardiac tests, B) combination, C) scans, D) endoscopy, E) ultrasound, F) x-ray

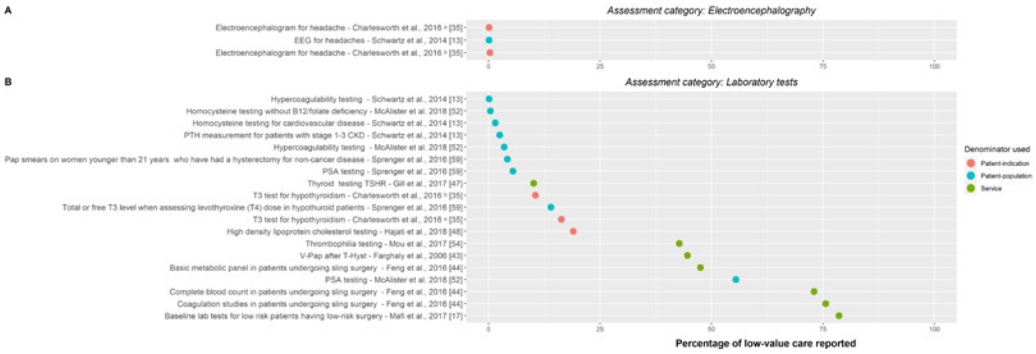


Figure 3 | Assessment outcomes regarding the prevalence of low-value diagnostic testing for all assessments included in the other diagnostic tests category: A) laboratory tests B) electroencephalography tests.

Variation among assessments of similar procedures

For two types of diagnostic tests, multiple assessments using similar lenses were identified among the included studies. These included: short-interval repeat bone densitometry testing (Dual-Energy X-ray absorptiometry) and the use of imaging procedures for nonspecific lower back pain. Considerable heterogeneity was observed between the extracted assessments for both groups ($I^2 \geq 100\%$) (see supplementary file 7 for the generated forest plot). We therefore chose to forgo generating pooled estimates, since pooling heterogenous studies could lead to invalid results. In particular, assessments of overuse of imaging for nonspecific lower back pain showed substantial variation, irrespective of the assessment lens used.

Discussion

In this systematic review, we identified and summarised the outcomes of studies assessing the prevalence of overuse of diagnostic tests. The majority of the 118 identified assessments examined the overuse of diagnostic imaging procedures (N=96), followed by the category 'Other diagnostic tests' which included laboratory tests (N=19) and electroencephalography tests (N=3). Assessments of low-value diagnostic testing using a patient-indication lens were most common (N=67, 57%), followed by assessments that used a service lens (N=27, 23%) and the patient-population lens (N=24, 20%). Major variation between prevalence estimates was observed, irrespective of assessment lens used. Prevalence estimates of diagnostic testing overuse ranged from 0.09% to 97.5% (median prevalence of assessments using a patient-indication lens: 11.0%,

a patient-population lens: 2.0%, a service lens: 30.7%) although eighty-four of the included assessments reported the prevalence of overuse to be below 25%. Among the 33 assessments reporting high levels of overuse (i.e., ≥25%), multiple assessments exploring the overuse of imaging for uncomplicated lower back pain (5 assessments) or preoperative tests (7 assessments) were present. Additionally, eleven of the 33 measurements reporting high levels of overuse were extracted from 8 studies considered at high risk of bias. Similar to the review of O'Sullivan [22], we found substantial variation in overuse among diagnostic services. However, our study adds to this finding by illustrating that variation is not limited to the primary care setting. Substantial overuse of diagnostic testing was also observed among diagnostic services often used in the secondary care setting, such as short-interval of bone mineral density testing or non-indicated cardiac testing before low-risk surgery. Through implementation of the concept of the assessment lenses to the included assessments, as proposed by Chalmers et al., [27] we were able to better compare the different assessment outcomes for similar diagnostic tests. Comparison of the different assessments outcomes regarding similar tests revealed that the observed variation could in part be explained by the use of different assessment lenses, an aspect which O'Sullivan et al. did not account for in their study. [22] Furthermore, we found that distinguishing primary from secondary care practices is often difficult and not always straight forward. Reasons are that many diagnostic practices are often provided in both the primary and secondary setting, and the setting in which these practices are provided often differs between countries and their respective healthcare systems.

For two types of low-value diagnostic testing, i.e. short-interval repeat bone densitometry testing and imaging for non-specific lower back pain without the presence of red flags, several similar assessments were extracted from the included literature. We tried to pool those similar assessments, however refrained from doing so after observing significant among-study heterogeneity ($I^2 \geq 100\%$). We therefore chose to report the results of the individual studies instead. The high levels of heterogeneity observed warrant further examination through means of subgroup analysis. However, the examination of potential sources of heterogeneity was hampered by the limited number of assessments present in each group. The limited number of assessments in each group also prevented us from reliably testing for publication bias. [40] Although we could not examine the heterogeneity through means of statistical subgroup analysis, we have tried to find possible explanations for the observed heterogeneity in the available literature and comparison of the studies. As mentioned before substantial variation among the extracted assessments of overuse was observed among the assessments included in our study. This variation could be caused by differences in study design, cohort size or operationalization of guidelines. Additionally, previous research has shown that factors such as population characteristics, healthcare systems, and insurance systems can greatly affect the amount of overuse. [7, 13, 15, 16, 18, 25, 35, 36, 39] For example, both the study by Bouck et al., [39] and Pendrith et al., [41] examined the overuse of imaging for lower back pain in Canada. However, each study used different data sources (Patient-Level

Physician Billing Repository, Discharge Abstract Database, National Ambulatory Care Reporting System vs. Ontario Health Insurance Plan claims database, respectively) and therefore used different codes to identify the included cohort. Furthermore, Pendrith et al., [41] included all visits to the primary care physician of adult patients (age>18 years) in their examination, while Bouck et al., [39] included only the first family physician visit. Although such differences appear small, they can drastically alter the patients included in the cohort and therefore influence the final prevalence estimate. The observed differences in estimates could also be caused by differences in definitions of low-value diagnostic testing. Most assessments are based on recommendations derived from initiatives such as the National Institute for Health and Care Excellence (NICE England) or Choosing Wisely (CW). However, no standardized definitions of low-value procedures or assessments, for the specific countries, exist. The absence of standardized definitions for specific countries could result in different cohorts and thus different prevalence estimates.

Finally, the use of different methods to assess overuse can explain the observed differences in outcomes. Some articles used the method as proposed by Schwartz et al., [13] which proposes the use of narrow (high specificity, low sensitivity) and broad indicators (low specificity, high sensitivity) to assess low-value care. [13] Narrow definitions are more tightly formulated, resulting in a more distinct cohort of patients/services that is included as compared to the cohorts created using broad definitions. Through a combination of both assessments, a more complete understanding of the problem is obtained. However, while using both narrow and broad indicators appears to be a good way to provide an estimate of the amount of overuse of low-value practices, it was only employed in 3 of the included articles. [13, 15, 16] In our analysis, we only used the broad assessments from those studies, since the underlying definitions of those more closely resembled the original recommendations. Therefore, broad assessment outcomes are more suitable for comparison to the outcomes of studies that directly used the relevant original recommendations in their assessments.

Strengths & Limitations

A strength of this study is that we did not limit our review to a single type of diagnostic testing or disease. Additionally, we did not limit the search to a particular setting; as a result, we present prevalence estimates for a wide range of diagnostic tests across all healthcare settings. Furthermore, we only included direct measures of diagnostic testing overuse acquired from data collected in databases.

Our study also faces some limitations. First, we recognize that the measurement of low-value care is often biased. Most existing measurements of low-value care target practices that are easily measured using existing data. These measurements clearly distinguish high-value from low-value services. However, most guidelines do not provide such a clear distinction. Detailed clinical information is often required to accurately distinguish high-

value from low-value care, but is often not present in the available data. [13, 36, 42-45] Because of these reasons, only a relatively small part of the total amount of low-value services has been examined so far. Unfortunately, we were unable to reliably test for publication bias due to the limited number of similar assessments which used the same scope present in our study. [40] Publication bias might be present among assessments of low-value practices because reports of the presence of substantial overuse is undesirable for most parties involved in such assessments. However, while our overview contains such a wide range of assessment outcomes we have attempted to reduce the publication bias where possible. Second, although we attempted to include all relevant keywords in our search strategy, our strategy may have missed some relevant terms and thus overlooked some studies assessing overuse of diagnostic services. Additionally, we incorporated several terms, such as overuse and low-value care in our search, which have been added to the lexicon relatively recently. Also, our search strategy only identifies studies that explicitly acknowledge the examined tests as representing overuse or low-value care. It is therefore possible our search might have missed studies which did not use these terms yet or that included some appropriate services alongside inappropriate ones in their assessment. Third, we only included studies that assessed overuse in relation to a specific guideline. Although this is a commonly used criterion and seen as an objective method to assess overuse, it is prone to underestimation of the actual prevalence of the problem. Yet, there is a risk of missing patients who do not exactly fit the specific guideline(s) used, or falsely classifying a test as (in)appropriate due to the clinical complexity of the patient involved. Furthermore, by requiring an assessment to be performed against a guideline, we did not capture all assessments of low-value diagnostic practices. Different methods are also used to distinguish appropriate from inappropriate care, such as expert opinion, Delphi or RAND appropriateness methods [46]. Because we only included assessment to require a guideline, our study therefore does not capture the full scope of assessments of low-value diagnostic overtesting. Fourth, we used a modified version of the Hoy risk of bias tool. [28] This is a validated tool for the assessment of risk of bias in prevalence studies, although we had to slightly adjust it to make it suitable to our research. However, while we tried to keep the tool as original as possible, we do need to consider that the modifications made to the original tool might have affected the outcome of our risk of bias assessment. Lastly, each of the included studies used their own definition of overuse in their assessments. Due to these differences in definitions of overuse it is often difficult to directly compare assessments of similar procedures, since these differences are in part responsible for the differences in outcome. However, by assigning assessment lenses to the included assessment of similar practices, we were able to group assessments using similar definitions of overuse and compare those to one another.

Implications for practice and future research

Most studies included in our review were conducted in the U.S. and only a few studies examining diagnostic testing overuse have been conducted in Europe. Findings from

one country (such as the U.S.) are often not generalizable to other countries, due to differences in (patient) population characteristics, healthcare and insurance systems. Additional assessments of overuse from different countries are needed to gain further insight into the magnitude of the overuse problem. Insight into the prevalence of diagnostic testing overuse is required to create a sense of urgency among (local) physicians and policy makers and to help develop effective strategies to tackle low-value diagnostic overtesting. [47, 48] Assessments should be repeated to monitor the problem of overuse of diagnostic testing over time and the effects of implemented strategies and interventions. In our review, only one study assessed overuse across multiple time periods. [38] The overview of assessment outcomes generated in this review could be used by both policy makers and care providers as a source of inspiration for (future) assessments in their own organisation(s) and (subsequently) as comparison material for their assessment outcomes.

International agreement on low-value service definitions and standardization assessment methods (e.g., identical denominators, similar lenses and scopes) could contribute to prevalence estimates that are comparable across countries. An example of which would be the recently completed study which compared the overuse of laboratory testing in U.S. to Canada. [49] However, while it would certainly help to have unified definitions and methods for the assessment of low-value care, it would certainly be an ambitious goal to set. Hence, each of the different assessments included in this study were conducted in different contexts and with slightly different purposes in mind. However, they all do have in common that they were performed to gain insight into the (local) problem of low-value diagnostic practices. These assessments therefore are crucial first steps in the process of reducing low-value diagnostic practices (locally).

Lastly, it might be of interest to include cost estimates in future assessments, because it is known that costs differences exist across countries and healthcare systems. Another reason why costs estimates would be of interest would be that previous research has indicated that low-cost services are predominantly overused. [17] We therefore suggest that future studies should include the associated costs of low-value diagnostic tests (possibly including additional down-stream costs due to performance of low-value diagnostic tests) in their assessments. However, we would like to emphasise that while cost is an important argument in the discussion of addressing low-value testing, it is not the only and certainly not the most important potential harm of unnecessary testing.

Conclusion

This study shows that there is substantial overuse of diagnostic testing present across all healthcare settings, with much variation among similar diagnostic services. Preoperative testing and imaging for nonspecific low back pain are the most frequently assessed and overused low-value diagnostic tests. Effective strategies to tackle the overuse of

diagnostic testing must be developed and implemented by health systems, providers, policy makers, and others. Additionally, more uniform definitions and assessments of low-value diagnostic tests are required in order to obtain a better understanding of the magnitude of diagnostic testing overuse.

References

1. Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, et al. Diagnostic strategies used in primary care. Bmj. 2009;338:b946.
2. Koch H, van Bokhoven MA, ter Riet G, van Alphen-Jager JT, van der Weijden T, Dinant GJ, et al. Ordering blood tests for patients with unexplained fatigue in general practice: what does it yield? Results of the VAMPIRE trial. The British journal of general practice : the journal of the Royal College of General Practitioners. 2009;59(561):e93-100.
3. McAlister FA, Lin M, Bakal J, Dean S. Frequency of low-value care in Alberta, Canada: a retrospective cohort study. BMJ Quality & Safety. 2018;27(5):340.
4. Berlin L. Overdiagnosed: Making People Sick in Pursuit of Health. JAMA. 2011;305(13):1356-9.
5. 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.
6. Brodersen J, Schwartz LM, Heneghan C, O'Sullivan JW, Aronson JK, Woloshin S. Overdiagnosis: what it is and what it isn't. BMJ Evidence-Based Medicine. 2018;23(1):1-3.
7. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. Med Care Res Rev. 2017;74(5):507-50.
8. Xu KT, Roberts D, Sulapas I, Martinez O, Berk J, Baldwin J. Over-prescribing of antibiotics and imaging in the management of uncomplicated URIs in emergency departments. BMC Emerg Med. 2013;13:7.
9. Xu S, Hom J, Balasubramanian S, Schroeder LF, Najafi N, Roy S, et al. Prevalence and Predictability of Low-Yield Inpatient Laboratory Diagnostic Tests. JAMA network open. 2019;2(9):e1910967.
10. Bruce Alexander C. Message from the President: Reducing Healthcare Costs through Appropriate Test Utilization. Critical Values. 2015;5(2):6-9.
11. 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.
12. Yeh DD. A clinician's perspective on laboratory utilization management. Clinica chimica acta; international journal of clinical chemistry. 2014;427:145-50.
13. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring low-value care in Medicare. JAMA Intern Med. 2014;174(7):1067-76.
14. Elshaug AG. Over 150 potentially low-value health care practices: an Australian study. Med J Aust. 2013;198(2):85.
15. Badgery-Parker T, Pearson SA, Chalmers K, Brett J, Scott IA, Dunn S, et al. Low-value care in Australian public hospitals: prevalence and trends over time. BMJ Qual Saf. 2019;28(3):205-14.
16. Chalmers K, Pearson SA, Badgery-Parker T, Brett J, Scott IA, Elshaug AG. Measuring 21 low-value hospital procedures: claims analysis of Australian private health insurance data (2010-2014). BMJ Open. 2019;9(3):e024142.

17. Mafi JN, Russell K, Bortz BA, Dachary M, Hazel WA, Jr., Fendrick AM. Low-Cost, High-Volume Health Services Contribute The Most To Unnecessary Health Spending. Health Aff (Millwood). 2017;36(10):1701-4.
18. Kool RB, Verkerk EW, Meijs J, van Gorp N, Maessen M, Westert G, et al. Assessing volume and variation of low-value care practices in the Netherlands. European journal of public health. 2020.
19. Chmiel C, Reich O, Signorell A, Tandjung R, Rosemann T, Senn O. Appropriateness of diagnostic coronary angiography as a measure of cardiac ischemia testing in non-emergency patients - a retrospective cross-sectional analysis. PLoS One. 2015;10(2):e0117172.
20. Kovacs FM, Arana E, Royuela A, Cabrera A, Casillas C, Pinero P, et al. Appropriateness of lumbar spine magnetic resonance imaging in Spain. Eur J Radiol. 2013;82(6):1008-14.
21. Petruzziello L, Hassan C, Alvaro D, Kohn A, Rossi Z, Zullo A, et al. Appropriateness of the indication for colonoscopy: is the endoscopist the 'gold standard'? J Clin Gastroenterol. 2012;46(7):590-4.
22. O'Sullivan JW, Albasri A, Nicholson BD, Perera R, Aronson JK, Roberts N, et al. Overtesting and undertesting in primary care: a systematic review and meta-analysis. BMJ Open. 2018;8(2):e018557.
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
24. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama. 2000;283(15):2008-12.
25. Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB. Choosing Wisely: Prevalence and Correlates of Low-Value Health Care Services in the United States. Journal of General Internal Medicine. 2014;30(2):221-8.
26. Nicholas W, Malcolm L. Medical screening Oxford Textbook of Medicine. Oxford, UK: Oxford University Press.
27. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
28. Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol. 2012;65(9):934-9.
29. R Core Team. R: A language and environment for statistical computing. v3.6.3 ed. Vienna, Austria: R foundation for Statistical Computing; 2019.
30. Wickham H. ggplot2: elegant graphics for data analysis. In: Springer New York, editor. 2009.
31. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evidence Based Mental Health. 2019;22(4):153-60.
32. Viechtbauer W. Conducting meta-analyses in {R} with the {metafor} package. Journal of Statistical Software. 2010;36(3):1 - 48.
33. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.

34. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003;327(7414):557-60.
35. Charlesworth CJ, Meath TH, Schwartz AL, McConnell KJ. Comparison of Low-Value Care in Medicaid vs Commercially Insured Populations. JAMA Intern Med. 2016;176(7):998-1004.
36. Colla CH, Morden NE, Sequist TD, Mainor AJ, Li Z, Rosenthal MB. Payer Type and Low-Value Care: Comparing Choosing Wisely Services across Commercial and Medicare Populations. Health Services Research. 2018;53(2):730-46.
37. Gold R, Esterberg E, Hollombe C, Arkind J, Vakarcs PA, Tran H, et al. Low Back Imaging When Not Indicated: A Descriptive Cross-System Analysis. Perm J. 2016;20(2):25-33.
38. Flaherty S, Zepeda ED, Mortele K, Young GJ. Magnitude and financial implications of inappropriate diagnostic imaging for three common clinical conditions. Int J Qual Health Care. 2018.
39. Bouck Z, Pendrith C, Chen XK, Frood J, Reason B, Khan T, et al. Measuring the frequency and variation of unnecessary care across Canada. BMC Health Serv Res. 2019;19(1):446.
40. Page MJ HJ, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor: Cochrane; 2020 2020.
41. Pendrith C, Bhatia M, Ivers NM, Mecredy G, Tu K, Hawker GA, et al. Frequency of and variation in low-value care in primary care: a retrospective cohort study. CMAJ Open. 2017;5(1):E45-E51.
42. Elshaug AG, McWilliams JM, Landon BE. The value of low-value lists. Jama. 2013;309(8):775-6.
43. Bhatia RS, Levinson W, Shortt S, Pendrith C, Fric-Shamji E, Kallewaard M, et al. Measuring the effect of Choosing Wisely: an integrated framework to assess campaign impact on low-value care. BMJ Qual Saf. 2015;24(8):523-31.
44. Morgan DJ, Leppin AL, Smith CD, Korenstein D. A Practical Framework for Understanding and Reducing Medical Overuse: Conceptualizing Overuse Through the Patient-Clinician Interaction. J Hosp Med. 2017;12(5):346-51.
45. Chalmers K, Badgery-Parker T, Pearson S-A, Brett J, Scott IA, Elshaug AG. Developing indicators for measuring low-value care: mapping Choosing Wisely recommendations to hospital data. BMC Res Notes. 2018;11(1):163-.
46. Kathryn F. The Rand/UCLA appropriateness method user's manual: Santa Monica : Rand, 2001.; 2001.
47. Parchman ML, Henrikson NB, Blasi PR, Buist DS, Penfold R, Austin B, et al. Taking action on overuse: Creating the culture for change. Healthc (Amst). 2017;5(4):199-203.
48. Grimshaw JM, Patey AM, Kirkham KR, Hall A, Dowling SK, Rodondi N, et al. De-implementing wisely: developing the evidence base to reduce low-value care. BMJ Qual Saf. 2020.
49. Henderson J, Bouck Z, Holleman R, Chu C, Klamerus ML, Santiago R, et al. Comparison of Payment Changes and Choosing Wisely Recommendations for Use of Low-Value Laboratory Tests in the United States and Canada. JAMA Intern Med. 2020.

Supplementary materials

Supplementary file 1: PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	N.A.
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	No review protocol has been registered
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5 & Supplementary file 2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5-7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7-8 & Supplementary file 4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7-8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N.A.
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8-9 & Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-12 & Supplementary file 6
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10 & Supplementary file 4 & 5
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	10-11, Table 2 & Figure 2 & Figure 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Supplementary file 7

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	10 & Supplementary file 5
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N.A.
DISCUSSION			
Summary of evidence	24	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	13-15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15-17
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17-18
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	20

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Supplementary file 2: Search strategy

1 | Pubmed

("Medical Overuse"[Mesh] OR Low-value care [tiab] OR low-value hospital care [tiab] OR low-value healthcare [tiab] OR wasteful care [tiab] OR wasteful healthcare [tiab] OR wasteful hospital care [tiab] OR overuse of healthcare [tiab] OR overuse procedure* [tiab] OR medical overuse [tiab] OR inappropriate healthcare [tiab] OR inappropriate care [tiab] OR unwanted healthcare [tiab] OR unwanted care [tiab] OR unnecessary healthcare [tiab] OR unnecessary care [tiab] OR Overdiagnos* [tiab] OR Over diagnos* [tiab] OR ineffective care [tiab] OR ineffective healthcare [tiab]) AND ("Diagnosis"[Mesh] OR diagnos* [tiab]) AND "statistics and numerical data" [Subheading] OR Variation* [tiab] OR Volume* [tiab] OR Prevalence* [tiab] OR Cost [tiab] OR costs [tiab] OR Frequenc* [tiab] AND "Guidelines as Topic"[Mesh] OR Guideline* [tiab] OR Choosing Wisely [tiab] OR policy [tiab] OR policies [tiab]

2 | Web of science:

TOPIC: ("Low-value care" OR "low-value hospital care" OR "low-value healthcare" OR "wasteful care" OR "wasteful healthcare" OR "wasteful hospital care" OR "overuse of healthcare" OR "overuse procedure*" OR "medical overuse" OR "inappropriate healthcare" OR "inappropriate care" OR "unwanted healthcare" OR "unwanted care" OR "unnecessary healthcare" OR "unnecessary care" OR "Overdiagnos*" OR "Over diagnos*" OR "ineffective care" OR "ineffective healthcare") AND (diagnos*) AND (Variation* OR Volume* OR Prevalence* OR Cost OR costs OR frequenc*) AND (Guideline* OR "Choosing Wisely" OR policy OR policies)

3 | Embase:

(exp clinical effectiveness/ OR (Low-value care OR low-value hospital care OR low-value healthcare OR wasteful care OR wasteful healthcare OR wasteful hospital care OR overuse of healthcare OR overuse procedure* OR medical overuse OR inappropriate healthcare OR inappropriate care OR unwanted healthcare OR unwanted care OR unnecessary healthcare OR unnecessary care OR Overdiagnos* OR Over diagnos* OR ineffective care OR ineffective healthcare).ti,ab,kw.) AND (exp diagnosis/ OR (diagnos*).ti,ab,kw.) AND frequency/ OR exp health statistics/ OR (Variation* OR Volume* OR Prevalence* OR Cost OR costs OR frequenc*).ti,ab,kw. AND exp Health statistics/ AND exp Practice guideline/ OR (Guideline* OR Choosing Wisely OR policy OR policies).ti,ab,kw.

Supplementary file 3: Overview of imaging modalities per category

Cardiac test	Endoscopy	Scan	Ultrasound	X-ray	Combination
Stress electrocardiogram	Arthroscopy	CT	Carotid artery imaging	X-ray	Combination of multiple modality categories
Echocardiogram	Endoscopy	MRI	Doppler	Dual Energy X-ray absorptiometry	
Cardiac nuclear medicine imaging	Colonoscopy	Whole body scan	Ultrasound		
Cardiac MRI/CT angiography		Bone scan	Plethysmography		
Single photon emission computed tomography myocardial perfusion imaging (SPECT-MPI)					

Name of author(s): ..... Year of publication: .....

Name of paper/study: .....

This tool is designed to assess the risk of bias in population-based prevalence studies. Please read the additional notes for each item when initially using the tool. Note: If there is insufficient information in the article to permit a judgement for a particular item, please answer No (**HIGH RISK**) for that particular item.

Risk of bias item	Adjusted criteria for answers (or reason for a criteria being not applicable (N.A.)):	Criteria for answers (please circle one option)	Additional notes and examples
External Validity			
1. Was the study's target population **a close re-pre sentation** of the national population in re lation to relevant variables, e.g. age, sex, occupation?	1 Original tool used.	2 **Yes (LOW RISK):** The study's target population was a close representation of the national population. 3 **No (HIGH RISK):** The study's target population was clearly NOT representative of the national population.	The **target population** refers to the group of people or entities to which the results of the study will be generalised. Examples: **1** The study was a national health survey of people 15 years and over and the sample was drawn from a list that included all individuals in the population aged 15 years and over. The answer is: **Yes (LOW RISK)**. **2** The study was conducted in one province only, and it is not clear if this was representative of the national population. The answer is: **No (HIGH RISK)**. **3** The study was undertaken in one village only and it is clear this was not representative of the national population. The answer is: **No (HIGH RISK)**.
2. Was the sampling frame a **true or close representation** of the target population?	4 **Yes (LOW RISK):** The sampling frame was a true or close representation of the target population described in the guideline. **5 No (HIGH RISK**): The sampling frame was NOT a **true or close** representation of the target population described in the guideline. Limited to no information regarding the demographics of the study population was provided.	6 **Yes (LOW RISK):** The sampling frame was a true or close representation of the target population. 7 **No (HIGH RISK):** The sampling frame was NOT a true or close representation of the target population.	The **sampling frame** is a list of the sampling units in the target population and the study sample is drawn from this list. Examples: 8 The sampling frame was a list of almost every individual within the target population. The answer is: **Yes (LOW RISK).** 9 The cluster sampling method was used and the sample of clusters/villages was drawn from a list of all villages in the target population. The answer is: **Yes (LOW RISK).** **10** The sampling frame was a list of just one particular ethnic group within the overall target population, which comprised many groups. The answer is: **No (HIGH RISK).**

Risk of bias item	Adjusted criteria for answers (or reason for a criteria being not applicable (N.A.)):	Criteria for answers (please circle one option)	Additional notes and examples
3 Was some form of **random selection** used to select the sample, OR, was a census under taken?	11 Original tool used.	12 **Yes (LOW RISK):** A census was undertaken, OR, some form of random selection was used to select the sample (e.g. simple random sampling, stratified random sampling, cluster sampling, systematic sampling). 13 **No (HIGH RISK):** A census was NOT undertaken, AND some form of random selection was NOT used to select the sample.	A census collects information from every unit in the sampling frame. In a survey, only part of the sampling frame is sampled. In these instances, random selection of the sample helps minimise study bias. Examples: 14 The sample was selected using simple random sampling. The answer is: **Yes (LOW RISK)**. 15 The target population was the village and every person in the village was sampled. The answer is: **Yes (LOW RISK)**. **16** The nearest villages to the capital city were selected in order to save on the cost of fuel. The answer is: **No (HIGH RISK).**
4 Was the likelihood of **non-response bias minimal?**	• N.A. No survey studies have been included. All data is obtained from electronic databases or file studies; in which we have made the assumption that all described data was collected from the aforementioned databases.	17 **Yes (LOW RISK):** The response rate for the study was >/=75%, OR, an analysis was performed that showed no significant difference in relevant demographic characteristics between responders and non- responders 18 **No (HIGH RISK):** The response rate was <75%, and if any analysis comparing responders and non-responders was done, it showed a significant difference in relevant demographic characteristics between responders and non-responders.	Examples: 19 The response rate was 68%; however, the researchers did an analysis and found no significant difference between responders and non-responders in terms of age, sex, occupation and socio- economic status. The answer is: **Yes (LOW RISK)**. 20 The response rate was 65% and the researchers did NOT carry out an analysis to compare relevant demographic characteristics between responders and non-responders. The answer is: **No (HIGH RISK)**. 21 The response rate was 69% and the researchers did an analysis and found a significant difference in age, sex and socio-economic status between responders and non-responders. The answer is: **No (HIGH RISK)**.

Risk of bias item	Adjusted criteria for answers (or reason for a criteria being not applicable (N.A.)):	Criteria for answers (please circle one option)	Additional notes and examples
Internal Validity			
5 Were data collected **directly from the subjects** (as opposed to a proxy)?	22 Original tool used.	23 **Yes (LOW RISK):** All data were collected directly from the subjects. 24 **No (HIGH RISK):** In some instances, data were collected from a proxy.	25 A proxy is a representative of the subject. Examples:All eligible subjects in the household were interviewed separately. The answer is: **Yes (LOW RISK)**. 26 A representative of the household was interviewed and questioned about the presence of low back pain in each household member. The answer is: **No (HIGH RISK).**
6 Was an acceptable case definition used in the study?	• **Yes (LOW RISK):** An acceptable case definition was used. A clear definition or reference to an appropriate guideline is presented. • **No (HIGH RISK):** An acceptable case definition was NOT used. It is unclear which definition or guideline has been used.	• **Yes (LOW RISK):** An acceptable case definition was used. • **No (HIGH RISK):** An acceptable case definition was NOT used.	• For a study on low back pain, the following case definition was used: "Low back pain is defined as activity-limiting pain lasting more than one day in the area on the posterior aspect of the body from the bottom of the 12th rib to the lower gluteal folds." The answer is: **Yes (LOW RISK).** • For a study on back pain, there was no description of the specific anatomical location „back" referred to. The answer is: **No (HIGH RISK).** • For a study on osteoarthritis, the following case definition was used: "Symptomatic osteoarthritis of the hip or knee, radiologically confirmed as Kellgren-Lawrence grade 2-4". The answer is: LOW RISK.
7 Was the study instrument that measured the parameter of interest (e.g. prevalence of low back pain) shown to have **reliability and validity (if necessary)?**	• N.A. none of the studies included used an instrument to measure the parameter of interest, all studies used data directly derived from databases.	• **Yes (LOW RISK):** The study instrument had been shown to have reliability and validity (if this was necessary), e.g. test-re- test, piloting, validation in a previous study, etc. • **No (HIGH RISK)**: The study instrument had NOT been shown to have reliability or validity (if this was necessary).	• The authors used the COPCORD questionnaire, which had previously been validated. They also tested the inter-rater reliability of the questionnaire. The answer is: **Yes (LOW RISK).** • The authors developed their own questionnaire and did not test this for validity or reliability. The answer is: **No (HIGH RISK).**

Risk of bias item	Adjusted criteria for answers (or reason for a criteria being not applicable (N.A.)):	Criteria for answers (please circle one option)	Additional notes and examples
8 Was the **same mode of data collection** used for all subjects?	• **Yes (LOW RISK):** The same mode of data collection was used for all subjects. All data was derived from the same database. • **No (HIGH RISK):** The same mode of data collection was NOT used for all subjects. Data collected for the different participants was obtained from different databases.	• **Yes (LOW RISK):** The same mode of data collection was used for all subjects. • **No (HIGH RISK):** The same mode of data collection was NOT used for all subjects.	The mode of data collection is the method used for collecting information from the subjects. The most common modes are face-to- face interviews, telephone interviews and self-administered questionnaires. Examples: • All eligible subjects had a face-to-face interview. The answer is: **Yes (LOW RISK)**. • Some subjects were interviewed over the telephone and some filled in postal questionnaires. The answer is: **No (HIGH RISK)**.
9 Was the **length of the shortest prevalence period** for the parameter of interest appropriate?	• N.A. no prevalence period for the parameter of interest was used in our situation. We selected studies which use database data and thus are recorded at the time of treatment/consultation.	• **Yes (LOW RISK):** The shortest prevalence period for the parameter of interest was appropriate (e.g. point prevalence, one-week prevalence, one-year prevalence). • **No (HIGH RISK):** The shortest prevalence period for the parameter of interest was not appropriate (e.g. lifetime prevalence)	The prevalence period is the period that the subject is asked about e.g. "Have you experienced low back pain over the previous year?" In this example, the prevalence period is one year. The longer the prevalence period, the greater the likelihood of the subject forgetting if they experienced the symptom of interest (e.g. low back pain). Examples: • Subjects were asked about pain over the past week. The answer is: **Yes (LOW RISK)**. • Subjects were only asked about pain over the past three years. The answer is: **No (HIGH RISK)**.
10 Were the **numerator (s) and denominato r(s)** for the parameter of interest appropriate?	**27** Original tool used.	• **Yes (LOW RISK):** The paper presented appropriate numerator(s) AND denominator(s) for the parameter of interest (e.g. the prevalence of low back pain). • **No (HIGH RISK):** The paper did present numerator(s) AND denominator(s) for the parameter of interest but one or more of these were inappropriate.	There may be errors in the calculation and/or reporting of the numerator and/or denominator. Examples: • There were no errors in the reporting of the numerator(s) AND denominator(s) for the prevalence of low back pain. The answer is: **Yes (LOW RISK)**. • In reporting the overall prevalence of low back pain (in both men and women), the authors accidentally used the population of women as the denominator rather than the combined population. The answer is: **No (HIGH RISK).**
11 Summary item on the overall risk of study bias			

- **LOW RISK OF BIAS:** Further research is very unlikely to change our confidence in the estimate.
- **HIGH RISK OF BIAS:** Further research is very likely to have an important impact on our confidence in the estimate and is likely to change the estimate.

Author	1. Was the study's target population **a close representation** of the national population in relation to relevant variables, e.g. age, sex, occupation?	2. Was the sampling frame **a true or close representation** of the target population?	3. Was some form of **random selection** used to select the sample, OR, was a census undertaken?	4. Was the likelihood of **non-response bias minimal?**	5. Were data collected **directly from the subjects** (as opposed to a proxy)?	6. Was an acceptable case definition used in the study?	7. Was the study instrument that measured the parameter of interest (e.g. prevalence of low back pain) shown to have **reliability and validity (if necessary)?**	8. Was the **same mode of data collection** used for all subjects?	9. Was the **length of the shortest prevalence period** for the parameter of interest appropriate?	10. Were the numerator(s) and denominator(s) for the parameter of interest appropriate?	***11. Summary item on the overall risk of study bias:*** ***- 2x high: High*** ***- 1x high + 1x unclear: High*** ***- Rest of combinations: low***
Badgery-Parker et al., 2019 [15]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Bouck et al., 2019 [39]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Chalmers et al., 2019 [16]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Charlesworth et al., 2016 [35]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Chmiel et al., 2015 [19]	Low	Low	Low	N.A.	High	Unclear	N.A.	Low	N.A.	Low	Low
Choi et al., 2011 [40]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Colla et al., 2014 [27]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Colla et al., 2018 [36]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Doukky et al., 2016 [41]	High	High	Low	N.A.	High	Unclear	N.A.	Low	N.A.	Low	High
Drangsholt et al., 2019 [42]	High	High	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	High
Farghaly et al., 2006 [43]	Low	Unclear	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low

Author	1. Was the study's target population **a close representation** of the national population in relation to relevant variables, e.g. age, sex, occupation?	2. Was the sampling frame **a true or close representation** of the target population?	3. Was some form of **random selection** used to select the sample, OR, was a census undertaken?	4. Was the likelihood of **non-response bias minimal?**	5. Were data collected **directly from the subjects** (as opposed to a proxy)?	6. Was an acceptable case definition used in the study?	7. Was the study instrument that measured the parameter of interest (e.g. prevalence of low back pain) shown to have **reliability and validity (if necessary)?**	8. Was the **same mode of data collection** used for all subjects?	9. Was the **length of the shortest prevalence period** for the parameter of interest appropriate?	10. Were the numerator(s) and denominator(s) for the parameter of interest appropriate?	***11. Summary item on the overall risk of study bias:*** ***- 2x high: High*** ***- 1x high + 1x unclear: High*** ***- Rest of combinations: low***
Feng, et al., 2016 [44]	High	Low	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High
Flaherty et al., 2018 [38]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Ganguli et al., 2019 [45]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Gidwani et al., 2016 [46]	High	High	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	High
Gill et al., 2017 [47]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Gold et al., 2016 [37]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Hajati et al., 2018 [48]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Kool et al., 2020 [18]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Kovacs et al., 2013 [20]	Low	Unclear	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Lalude et al., 2014 [49]	Unclear	Low	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High
Lehnert et al., 2010 [50]	High	Unclear	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High

Mafi et al., 2017 [17]	Low	Unclear	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Martin et al., 2012 [51]	High	Low	Low	N.A.	High	Unclear	N.A.	Low	N.A.	Low	High
McAlister et al. 2018 [52]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Morden et al., 2014 [53]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Mou et al., 2017 [54]	High	Low	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High
Pendrith et al., 2017 [55]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Petruzziello et al., 2012 [21]	Unclear/high	Low	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High
Schwartz et al., 2014 [13]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Scott et al., 2014 [56]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Sharp et al., 2015 [57]	High	Low	Low	N.A.	High	High	N.A.	Low	N.A.	Low	High
Sheffield et al., 2013 [58]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Sprenger et al., 2016 [59]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low
Xu et al., 2013 [8]	Low	Low	Low	N.A.	High	Low	N.A.	Low	N.A.	Low	Low

Supplementary file 6: Details of low-value care assessments extracted from the included studies

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
Badgery-Parker et al., 2019 [15]	Australia	- Endoscopy in adults <55 (for dyspepsia)	Service	CW Australia, RACP EVOLVE and NICE 'do not do' guidelines/recommendations	14 813	2 360	15.93 [15.35 ; 16.53]	Imaging
		- Arthroscopic lavage and debridement of knee for osteoarthritis or degenerative meniscal tears			4 218	3 002	71.17 [69.78 ; 72.53]	Imaging
		- Colonoscopy in adults < 50 (for constipation)			11 790	608	5.16 [4.76 ; 5.57]	Imaging
		- ERCP (endoscopic retrograde cholangiopancreatography) for acute gallstone pancreatitis without cholangitis			420	79	18.80 [15.18 ; 22.88]	Imaging
Bouck et al., 2019 [39]"-Joris"=>" Included"}</custom1><language>eng</language></record></Cite></EndNote>	Canada	- Spinal X-ray, CT, MRI following visit for low-back pain.	Patient-indication	CW Canada guidelines/recommendations from the Canadian Anesthesiologists' Society (CAS), Canadian Cardiovascular Society (CCS) & Canadian Society of Internal Medicine (CSIM)	97 740	30 006	30.70 [30.41 ; 30.99]	Imaging
		- Cardiac tests (electrocardiogram, chest x-ray, stress test, or transthoracic echocardiogram) prior to low-risk procedures			527 691	167 278	31.70 [31.60 ; 31.80]	Imaging
Chalmers et al., 2019 [16]	Australia	- Knee arthroscopy in case of osteoarthritis or meniscal derangements	Service	RACP EVOLVE, CW Australia, CW USA, CW Canada, CW UK guidelines/recommendations	3 620	2 958	81.70 [80.41 ; 82.96]	Imaging
		- Endoscopy in case of dyspepsia below 50 years			5 021	501	9.97 [9.16 ; 10.84]	Imaging
		- Colonoscopy in case of constipation below 50 years			4 017	133	3.31 [2.78 ; 3.91]	Imaging

Charlesworth et al., 2016 [35] a	United States	- Imaging for nonspecific low-back pain	Patient-indication	CW USA , NICE guidelines/recommendations and QualityNet measures from the Center for Medicaid & Medicare Services (CMS)	82 659	12 977	15.70 [15.45 ; 15.95]	Imaging
		- Head imaging for uncomplicated headache			65 931	7 252	11.0 [10.76 ; 11.24]	Imaging
		- Head imaging for syncope			7 466	672	9.0 [8.36 ; 9.67]	Imaging
		- Imaging for plantar fasciitis			10 986	1 999	18.20 [17.48 ; 18.93]	Imaging
		- T3 test for hypothyroidism			31 228	5 090	16.30 [15.89 ; 16.71]	Laboratory test
		- Pre-operative chest radiography			39 169	1 684	4.30 [4.10 ; 4.5]	Imaging
		- Abdomen CT combined studies			28 075	2 106	7.50 [7.20 ; 7.82]	Imaging
		- Simultaneous brain and sinus CT			9 742	360	3.70 [3.33 ; 4.09]	Imaging
		- CT for uncomplicated acute rhinosinusitis			59 961	1 799	3.0 [2.87 ; 3.14]	Imaging
		- Arthroscopic surgery for knee osteoarthritis			23 190	1 136	4.90 [4.62 ; 5.18]	Imaging
		- Thorax CT combined studies			9 287	111	1.20 [0.98 ; 1.44]	Imaging
		- Preoperative echocardiography			39 169	235	0.60 [0.53 ; 0.68]	Imaging
		- Preoperative stress testing			39 169	118	0.30 [0.25 ; 0.36]	Imaging
		- Electroencephalogram for headache			59 936	60	0.10 [0.08 ; 0.13]	Other test

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
Charlesworth et al., 2016 [35] b	United States	- Imaging for nonspecific low-back pain	Patient-indication	CW USA , NICE guidelines/recommendations and QualityNet measures from the Center for Medicaid & Medicare Services (CMS)	18 871	4 625	22.60 [22.01 ; 23.2]	Imaging
		- Head imaging for uncomplicated headache			23 211	4 317	18.60 [18.10 ; 19.11]	Imaging
		- Head imaging for syncope			3 174	448	14.10 [12.92 ; 15.37]	Imaging
		- Imaging for plantar fasciitis			1 450	180	12.40 [10.76 ; 14.22]	Imaging
		- T3 test for hypothyroidism			5 891	619	10.50 [9.74 ; 11.32]	Laboratory test
		- Pre-operative chest radiography			7 848	753	9.60 [8.95 ; 10.27]	Imaging
		- Abdomen CT combined studies			13 416	657	4.90 [4.54 ; 5.28]	Imaging
		- Simultaneous brain and sinus CT			7 761	357	4.60 [4.14 ; 5.09]	Imaging
		- CT for uncomplicated acute rhinosinusitis			11 992	348	2.90 [2.61 ; 3.22]	Imaging
		- Arthroscopic surgery for knee osteoarthritis			6 143	166	2.70 [2.31 ; 3.14]	Imaging
		- Thorax CT combined studies			3 822	76	2.0 [1.57 ; 2.48]	Imaging
		- Preoperative echocardiography			7 848	86	1.10 [0.88 ; 1.35]	Imaging
		- Preoperative stress testing			7 848	47	0.60 [0.44 ; 0.8]	Imaging
		- Electroencephalogram for headache			20 091	60	0.30 [0.23 ; 0.38]	Other test
Chmiel et al., 2015 [19]	Switzerland	- Performance of a diagnostic coronary angiography without previous non-invasive ischemia testing	Patient-indication	The American heart association (AHA), NICE, the European Society of Cardiology (ESC) & Swiss Society of Cardiology (SGK) guidelines/recommendations	2 714	1 018	37.50 [35.68 ; 39.36]	Imaging

Choi et al., 2011 [40]	United States	- Routine cross-sectional imaging (CT, MRI, endorectal coil MRI) for staging low-risk prostate cancer	Patient-indication	American College of Radiology (ACR) & National Comprehensive Cancer Network (NCCN) guidelines/recommendations	2 330	548	23.50 [21.81 ; 25.3]	Imaging
		- Routine bone scan for staging low-risk prostate cancer			2 330	617	26.50 [24.70 ; 28.32]	Imaging
		- Routine abdominal ultrasound for staging low-risk prostate cancer			2 330	42	1.80 [1.30 ; 2.43]	Imaging
Colla et al., 2014 [27]	United States	- Low back X-ray, CT or MRI within six weeks of incident low-back pain diagnosis	Patient-indication	American Academy of Family Physicians (AAFP), American College of Physicians (ACP) & North American Spine Society (NASS) guidelines/recommendations	8 440 000	1 899 000	22.50 [22.48 ; 22.52]	Imaging
		- Intravenous pyelogram or an abdominal CT, MRI, or ultrasound within 60 days of the index diagnosis benign prostatic hyperplasia (BPH)			75 000 000	900 000	1.20 [1.20 ; 1.20]	Imaging
		- DXA scans performed on female beneficiaries at low risk for fracture within 23 months of a previous scan			147 420 000	14 299 740	9.70 [9.70 ; 9.70]	Imaging
		- Non-indicated cardiac test, including stress tests, echocardiograms, electrocardiograms and advanced cardiac imaging in the 30 days before cataract surgery			6 490 000	999 460	15.40 [15.37 ; 15.43]	Imaging
		- Non-indicated cardiac test, including stress tests, echocardiograms, electrocardiograms, CTs, MRIs or PETs within 30 days before low-risk surgery			3 550 000	1 650 750	46.50 [46.46 ; 46.54]	Imaging
Colla et al., 2018 [36] a	United States	- Early imaging inf patients with uncomplicated, incident low-back pain who received nonindicated low-back-pain imaging in the 6 weeks following diagnosis	Patient-indication	American Academy of Family Physicians (AAFP)& American College of Physicians (ACP) guidelines/recommendations	3 110 000	895 680	28.80 [28.76 ; 28.84]	Imaging
		- Short-interval repeat bone densitometry (DXA)			8 000 000	600 000	7.50 [7.48 ; 7.52]	Imaging

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
Colla et al., 2018 [36] c	United States	- Early imaging inf patients with uncomplicated, incident low-back pain who received nonindicated low-back-pain imaging in the 6 weeks following diagnosis	Patient-indication	American Academy of Family Physicians (AAFP)& American College of Physicians (ACP) guidelines/recommendations	8 400 000	1 898 400	22.60 [22.58 ; 22.62]	Imaging
		- Short-interval repeat bone densitometry (DXA)			143 000 000	12 155 000	8.50 [8.50 ; 8.5]	Imaging
Doukky et al., 2016 [41]	United States	- SPECT-MPI (myocardial perfusion imaging)	Patient-indication	The American College of Cardiology (ACC)& American Society of Nuclear Cardiology (ASNC) revised appropriate use criteria (AUC) for SPECT MPI of 2009	1 511	688	45.53 [22.58 ; 22.62]	Imaging
Drangsholt et al., 2019 [42]	United States	- Routine imaging for staging low-risk prostate cancer (Whole body scan)	Patient-indication	National Comprehensive Cancer Network (NCCN) guidelines/recommendations	414	381	92.0 [88.99 ; 94.45]	Imaging
		- Routine imaging for staging low-risk prostate cancer (CT)			414	273	66.0 [61.15 ; 70.5]	Imaging
		- Routine imaging for staging low-risk prostate cancer (abdominal/pelvic MRI)			414	8	2.0 [0.84 ; 3.77]	Imaging
Farghaly et al., 2006 [43]	United States	- V-Pap after T-Hyst	Service	US Preventive Services Task Force (USPTF) & American Society for Colposcopy and Cervical Pathology (ASCCP) guidelines/recommendations	1 303	581	44.59 [41.87 ; 47.34]	Laboratory test

Feng et al., 2016 [44]	United States	- Preoperative ECGs in patients undergoing sling surgery	Service	Summary guidelines from the American Academy of Family Physicians (AAFP)	63	13	20.60 [11.47 ; 32.7]	Imaging
		- Preoperative chest X-ray (CXR) in patients undergoing sling surgery			23	9	39.10 [19.71 ; 61.46]	Imaging
		- Preoperative basic metabolic panel in patients undergoing sling surgery			59	28	47.50 [34.30 ; 60.88]	Laboratory test
		- Preoperative complete blood count determination in patients undergoing sling surgery			63	46	73.0 [60.35 ; 83.43]	Laboratory test
		- Preoperative coagulation studies in patients undergoing sling surgery			41	31	75.60 [59.70 ; 87.64]	Laboratory test
Flaherty et al., 2018 [38]	United States	- MRI use in case of uncomplicated LBP	Patient-indication	American College of Radiology (ACR) guidelines	5 103	1 838	36.01 [59.70 ; 87.64]	Imaging
		- MRI use in case of non-traumatic Knee pain			6 935	3 384	48.79 [47.61 ; 49.98]	Imaging
		- MRI use in case of non-traumatic shoulder pain			9 388	4 011	42.72 [41.72 ; 43.73]	Imaging
		- X-ray use in case of uncomplicated LBP			10 540	6 650	63.09 [62.16 ; 64.02]	Imaging
		- X-ray use in case non-traumatic Knee pain			37 543	25 668	68.37 [67.90 ; 68.84]	Imaging
		- X-ray use in case of non-traumatic shoulder pain			36 453	26 100	71.60 [71.13 ; 72.06]	Imaging
Ganguli et al.,2019 [45]	United States	- EKGs for cataract surgeries	Patient-indication	Choosing Wisely recommendations from the American Academy of Ophthalmology & American Society for Clinical Pathology	110 183	12 451	11.30 [11.11 ; 11.49]	Imaging

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
Gidwani et al., 2016 [46]	United States	- Lumbar Spine MRI's for non-specific or nonpersistent low-back pain	Service	The American College of Physicians (ACP) and the American Association of Neurological Surgeons (AANS) guidelines/recommendations and the National Quality forum (NQF)-endorsed CMS criteria	110 661	33 973	30.70 [30.43 ; 30.97]	Imaging
Gill et al., 2017 [47]	Canada	- Thyroid testing	Service	Clinical practice guidelines and Choosing Wisely guidelines/recommendations from the American Society for Clinical Pathology (ASCP), American Association for Clinical Chemistry (AACC), American Association of Clinical Endocrinologists (AACE), Canadian Society of Endocrinology and Metabolism (CSEM), Endocrine society.	752 217	75 974	10.10 [10.04 ; 10.16]	Laboratory test
Gold et al., 2016 [37] d	United States	- Imaging in patients with non-specific low-back pain (X-ray, CT, MRI)	Patient-indication	American Academy of Family Physicians National Quality Measures Clearinghouse (NQMC) guidelines/recommendations	19 503	3 288	16.86 [16.34 ; 17.39]	Imaging
Gold et al., 2016 [37] e	United States	- Imaging in patients with non-specific low-back pain (X-ray, CT, MRI)	Patient-indication	American Academy of Family Physicians National Quality Measures Clearinghouse (NQMC) guidelines/recommendations	2 694	449	16.67 [15.28 ; 18.13]	Imaging

Hajati et al., 2018 [48]	Australia	- High density lipoprotein cholesterol testing more often than once every 12 months for high-risk groups	Patient-indication	Royal Australian College of General Practitioners (RACGP) guidelines	1 628 477	309 411	19.0 [18.95 ; 19.05]	Laboratory test
Kool et al., 2020 [18]	The Netherlands	- Doppler or plethysmography for diagnosis of varices	Patient-indication	Dutch general practitioner, CW USA & CW Canada guidelines/recommendations	15 990	1 279	8.0 [7.58 ; 8.43]	Imaging
Kovacs et al., 2013 [20]	Spain	- Lumbar Spine MRI's for low-back pain without red flags.	Service	National Institute for Clinical Excellence (NICE), the American College of Physicians and the American College of Radiologists guidelines/ recommendations	602	64	10.60 [8.28 ; 13.37]	Imaging
Lalude et al., 2014 [49]	United States	- Inappropriate Single photon emission computed tomography myocardial perfusion imaging (SPECT MPI) utilisation	Service	The American College of Cardiology (ACC)& American Society of Nuclear Cardiology (ASNC) revised appropriate use criteria (AUC) for SPECT MPI of 2009	54	8	14.0 [6.62 ; 27.12]	Imaging
Lehnert et al., 2010 [50]	United States	- CT use	Service	American College of Radiology (ACR) Appropriateness Criteria and other evidence-based guidelines/recommendations	284	77	27.11 [17.37 ; 30.41]	Imaging
		- MRI use			175	41	23.43 [22.03 ; 32.68]	Imaging

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
Mafi et al., 2017 [17]	United States	- Baseline lab tests for low risk patients having low-risk surgery	Service	CW USA, US Preventive Services Task Force (USPTF), Medicare's Healthcare Effectiveness Data and Information Set (HEDIS) criteria and other clinical guidelines/ recommendations	595 552	468 104	78.60 [78.51 ; 78.69]	Laboratory test
		- Stress cardiac or other cardiac imaging in low-risk, asymptomatic patients			244 487	27 383	11.20 [11.08 ; 11.32]	Imaging
		- Routine head CT scans for Emergency Department visits for severe dizziness			29 816	15 713	52.70 [52.13 ; 53.27]	Imaging
		- EKGs, chest x-rays, or pulmonary function tests in low-risk patients having low-risk surgery			33 754	32 910	97.50 [97.33 ; 97.66]	Imaging
		- Routine imaging for uncomplicated acute rhinosinusitis			14 196	7 226	50.90 [50.08 ; 51.73]	Imaging
		- Imaging for low-back pain within the first six weeks of symptom onset, in absence of red flags			48 857	48 857	86.20 [85.89 ; 86.51]	Imaging
Martin et al., 2012 [51]	United States	- MRI use in musculoskeletal oncology	Patient-indication	Referral guidelines for patients with cancer	320	20	6.25 [3.86 ; 9.49]	Imaging
McAlister et al. 2018 [52]	Canada	- PSA testing for men 75 or older with no history of prostate cancer	Patient-population	CW and NICE guidelines/ recommendations	100 227	55 596	55.47 [55.16 ; 55.78]	Laboratory test
		- Bone mineral density testing within 2 years of prior scan			271 854	31 617	11.63 [11.52 ; 11.74]	Imaging
		- Hypercoagulability testing in patients with first Deep Vein Trombosis/ Pulmonary Embolism			21 311	744	3.49 [3.25 ; 3.75]	Laboratory test
		- Preoperative coronary CT scan or cardiac stress test before non-cardiac surgery			698 683	7 266	1.04 [1.02 ; 1.06]	Imaging
		- Homocysteine testing without B12/folate deficiency			2 585 832	10 602	0.41 [0.40 ; 0.42]	Laboratory test
		- Carotid artery imaging but without history of stroke / TIA			3 162 394	2 846	0.09 [0.09 ; 0.09]	Imaging
		- Carotid artery imaging for patients with syncope but no history of stroke / TIA			74 060	355	0.48 [0.43 ; 0.53]	Imaging

Morden et al., 2014 [53]	United States	- Short-interval (repeated in under 2 years) dual-energy X-ray absorptiometry tests (DXAs) for bone density	Service	The American College of Rheumatology guidelines/recommendations	13 800 000	1 393 800	10.10 [10.08 ; 10.12]	Imaging
Mou et al., 2017 [54]	United States	- Thrombophilia in adult patients with venous thromboembolism occurring in the setting of major transient risk factors (surgery, trauma, or prolonged immobility)	Service	The American Society of Hematology (ASH) and Clinical guidelines/recommendations	1 817	777	42.76 [40.47 ; 45.08]	Laboratory test
Pendrith et al., 2017 [55]	Canada	- Imaging for low back pain in absence of red flags - DEXA scans repeated under 2 years	Patient-indication	CW Canada, CW USA	271 588 2 229 113	12 222 468 114	4.5 [4.42 ; 4.58] 21.0 [20.95 ; 21.05]	
Petruzziello et al., 2012 [21]	Italy	- Colonoscopy use	Patient-indication	The American Society for Gastrointestinal Endoscopy (ASGE) guidelines/ recommendations	432	125	29.0 [24.70 ; 33.46]	Imaging
Schwartz et al., 2014 [13]	United States	- Bone mineral density test <2 y after prior bone mineral density test	Patient-population	CW USA, the US Preventive Services Task Force "D", NICE "do not do", Canadian Agency for Drugs and Technologies in Health (CADTH) guidelines/ recommendations and guidelines from peer-reviewed medical literature	1 360 908	127 925	9.40 [9.35 ; 9.45]	Imaging
		- Homocysteine testing for cardiovascular disease			1 360 908	20 414	1.50 [1.48 ; 1.52]	Laboratory test
		- Hypercoagulability testing for patients with deep vein thrombosis			1 360 908	1 361	0.10 [0.09 ; 0.11]	Laboratory test
		- PTH (parathyroid hormone) measurement for patients with stage 1-3 CKD (chronic kidney disease)			1 360 908	34 023	2.50 [2.47 ; 2.53]	Laboratory test
		- Preoperative chest radiograph specified as a preoperative assessment or occurring within 30 days before a low- or intermediate-risk noncardiothoracic surgical procedure			1 360 908	69 406	5.10 [5.06 ; 5.14]	Imaging

Study	Country	Type of low-value care examined	Lens	Guideline(s)/recommendations used	Cohort size	Amount LVC	Prevalence estimate (95%CI)	Type
		- Preoperative echocardiography before a low- or intermediate-risk noncardiothoracical surgery			1 360 908	10 887	0.80 [0.78 ; 0.82]	Imaging
		- CT of the sinuses for uncomplicated acute rhinosinusitis			1 360 908	8 165	0.60 [0.59 ; 0.61]	Imaging
		- Head imaging in the evaluation of syncope			1 360 908	17 692	1.30 [1.28 ; 1.32]	Imaging
		- Head imaging for uncomplicated headache			1 360 908	42 188	3.10 [3.07 ; 3.13]	Imaging
		- EEG for headaches			1 360 908	1 361	0.10 [0.09 ; 0.11]	Other test
		- Back imaging for patients with nonspecific low back pain			1 360 908	127 925	9.40 [9.35 ; 9.45]	Imaging
		- Arthroscopic surgery for knee osteoarthritis			1 360 908	2 722	0.2 [0.19 ; 0.21]	Imaging
Scott et al., 2014 [56]	United States	- Carotid duplex ultrasound (CDUS)for simple syncope in the absence of focal neurological signs or symptoms suggestive of stroke	Patient-indication	European Society of Cardiology's Task Force on Syncope & American College of Physicians' Clinical (ACP) Efficacy Assessment Project guidelines/recommendations	137 424	8 933	6.50 [6.37 ; 6.63]	Imaging
Sharp et al., 2015 [57]	United States	- CT scan for acute sinusitis	Patient-indication	American Academy of Family Physicians (AAFP), American Academy of Allergy, Asthma and Immunology (AAAAI), American Academy of Otolaryngology-Head and Neck surgery (AAO-HNS) CW guidelines/recommendations	152 774	1681	[1.05 ; 1.15]	Imaging

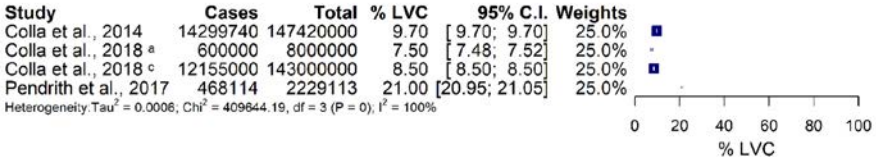
Sheffield et al., 2013 [58]	United States	- Cardiac Stress tests prior to elective non-cardiac, non-vascular surgery: Echocardiograms, Myocardial nuclear imaging, Exercise treadmill or pharmacological stress tests	Patient-indication	American College of Cardiology (ACC)& American Heart Association guidelines (AHA) guidelines/recommendations	74 785	2 804	3.75 [3.61 ; 3.89]	Imaging
Sprenger et al., 2016 [59]	Austria	- DEXA scans more often than every 2 years	Patient-population	CW USA, CW Canada, NICE recommendations/ guidelines and contradicted medical practices and potentially low-value health care practices identified by the studies Prasad et al., 2013 and Elshaug et al., 2012	246 131	12 282	4.99 [4.91 ; 5.07]	Imaging
		- Total or free T3 level when assessing levothyroxine (T4) dose in hypothyroid patients			246 131	34 434	13.99 [13.86 ; 14.12]	Laboratory test
		- Men 75 to 80 years old with a PSA less than 3ng/ml are unlikely to die or experience aggressive prostate cancer during their remaining life, suggesting that PSA testing may be safely discontinued in these men.			246 131	13 390	5.44 [5.35 ; 5.53]	Laboratory test
		- Pap smears on women younger than 21 years who have had a hysterectomy for non-cancer disease.			246 131	10 387	4.22 [4.14 ; 4.3]	Laboratory test
		- Routine monitoring of bone mineral density after starting bisphosphonate treatment			246 131	271	0.11 [0.10 ; 0.12]	Imaging
Xu et al., 2013 [8]	United States	- X-ray in outpatient management of uncomplicated Upper Respiratory Infections (URI's)	Patient-indication	Centres for Disease Control and Prevention (CDC) guidelines	2 200 000	1 012 000	46.0 [45.95 ; 46.05]	Imaging
		- CT in outpatient management of uncomplicated Upper Respiratory Infections (URI's)			2 200 000	44 000	2.0 [1.98 ; 2.02]	Imaging

Among the included studies, some studies contained multiple assessments undertaken in different cohorts. These assessments are distinguished by the following:
[a] assessment performed among a commercially insured population, [b] assessment performed among Medicaid beneficiaries, [c] assessment performed among Medicare beneficiaries, [d] assessment performed using Kaiser Permanente Epic EHR data, e assessment performed using data
derived from the Oregon Community Health Information Network. Abbreviations: CT, Computed Tomography; CW, Choosing Wisely; DEXA or DXA, Dual Energy X-ray absorptiometry; EKG, Electrocardiogram; LBP, low-back pain; MRI, Magnetic Resonance Imaging; NICE, National Institute for Health and Care Excellence; PSA, Prostate-Specific Antigen; RACP, Royal Australasian College of Physicians; SPECT MPI, Single-Photon Emission computed tomography; TIA, Transient Ischemic Attack; T-hyst, Total Hysterectomy

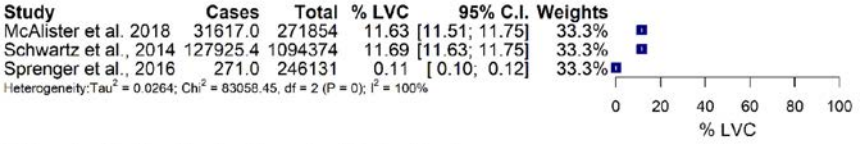
Supplementary file 7: Forest plots of assessment outcomes from studies using the same lens to assess overuse of similar low-value care (LVC) diagnostic tests

Among the included studies, some studies contained multiple assessments undertaken in different cohorts. These assessments are distinguished by the following: a assessment performed among [a] commercially insured population, [b] assessment performed among Medicaid beneficiaries, [c] assessment performed among Medicare beneficiaries, [d] assessment performed using Kaiser Permanente (KP) Epic EHR data, [e] assessment performed using data derived from the Oregon Community Health Information Network (OCHIN).

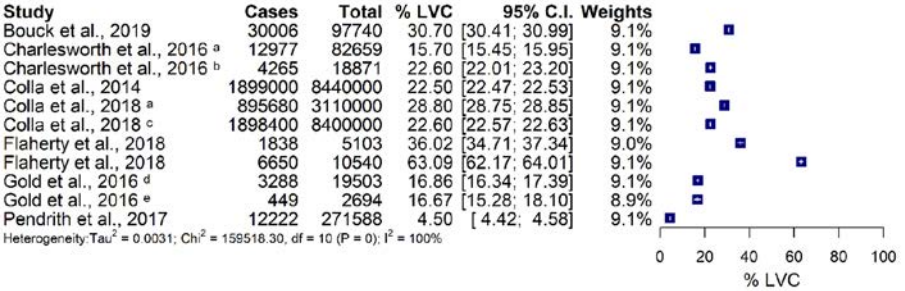
A. Short-interval repeat bone densitometry: patient-indication lens



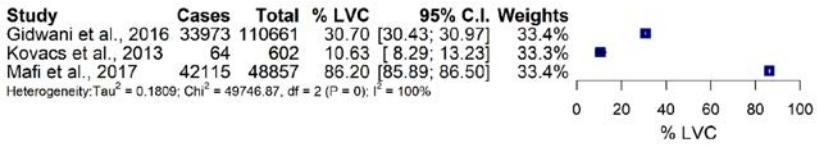
B. Short-interval repeat bone densitometry: patient-population lens



C. Imaging for low-back pain: patient-indication lens



D. Imaging for low-back pain: service lens



CHAPTER 3

Low-value pharmaceutical care among Dutch general practitioners: a retrospective cohort study

Joris L.J.M. Müskens
Simone A. van Dulmen
Tjerk Wiersma
Jako S. Burgers
Karin Hek
Gert P. Westert
Rudolf B. Kool

Published in The British Journal of General Practice, Volume 72, Issue 718, May 2022

Abstract

Background: Low-value pharmaceutical care exists in general practice. However, the extent among Dutch general practitioners (GPs) remains unknown.

Aim: To assess the prevalence of low-value pharmaceutical care among Dutch GPs.

Design and Setting: Retrospective cohort study using patient record data.

Method: We examined the prevalence of three types of pharmaceutical care, topical antibiotics for conjunctivitis, benzodiazepines for non-specific lower back pain (LBP), and chronic acid reducing medication (ARM) prescriptions, prescribed by GPs between 2016-2019. Multilevel logistic regression analysis was performed to assess prescribing variation and the influence of patient characteristics on receiving a low-value prescription.

Results: Large variation in prevalence and practice variation was observed among the examined types of low-value pharmaceutical GP care. Between 53%-61% of patients received an inappropriate antibiotics prescription for conjunctivitis, around 3% of LBP patients received an inappropriate benzodiazepine prescription and 88% received an inappropriate chronic ARM prescription during the years examined. The odds of receiving an inappropriate antibiotic or benzodiazepine prescription increased with age ($p<0.001$), but decreased for chronic inappropriate ARM prescriptions ($p<0.001$). Sex affected only the odds of receiving a non-indicated chronic ARM, with males being at higher risk ($p<0.05$). The odds for receiving an inappropriate ARM also increased with increasing neighbourhood socio-economic status ($p<0.05$). Increasing practice size also decreased the odds of inappropriate antibiotic and benzodiazepine prescriptions ($p<0.001$).

Conclusion: Our results show that the prevalence of low-value pharmaceutical GP care varies among the clinical problems examined. Significant variation in inappropriate prescribing exists between different types of pharmaceutical care - and GP practices.

Introduction

Low-value care, defined as care that is unlikely to benefit the patient given the potential harm, cost, available alternatives and patient preferences, is considered one of the most complex problems in modern healthcare. [1, 2] In an effort to support clinicians in their daily practice, professional bodies, such as the Dutch College of General Practitioners (NHG), have published more than 120 evidence-based clinical practice guidelines. [3] However, despite the wide distribution and promotion of these guidelines, studies show that adherence among Dutch general practitioners (GPs) could be improved. [4-10] Up to one third of Dutch GP pharmaceutical prescriptions could be of low-value [6]. International studies show that low-value GP prescribing is also common outside of the Netherlands. [2, 11-17]

Obtaining insight into the prevalence of low-value prescribing is an essential first step in improving practice. [18] Although some assessments of low-value prescribing among Dutch GPs exist, these are outdated and conducted using data that are not nationally representative. [4, 19-21] We therefore aimed to quantify the prevalence and variation in low-value pharmaceutical treatments among GPs by using national medical records data. We selected three recommendations from the set of GP guidelines that clearly emphasise that physicians should refrain from prescribing medication except when specific indications are met (box 1 contains a detailed description of the rationale behind the selection, and supplementary box 1 an description of the operationalization of each recommendation):

1 The prescription of local antibiotics for an infectious conjunctivitis. [22]
2 The prescription of benzodiazepines in the treatment of non-specific lower back pain (LBP). [23]
3 The chronic prescription or continuation of acid-reducing medication (ARM). [24]

Through quantification of these prescribing practices, we hope to obtain a clearer view of low-value prescribing among GPs in the Netherlands. This is a first step in addressing the specific issue of low-value GP prescribing. In addition to both the prevalence and the variation in prescribing behaviour, we were interested in the characteristics of patients associated with low-value prescribing.

Box 1: Rationale behind the recommendations selected.

1 *The prescription of local antibiotics for an infectious conjunctivitis.* [22] A Dutch study from2007 showed that up to 80% of conjunctivitis episodes were inappropriately treated with a topical antibiotic. [20] The indications for prescribing an antibiotic for conjunctivitis have not changed since then. One study from the USA showed that the number of conjunctivitis diagnoses has increased between 2005 and 2014 and the percentage of low-value antibiotic prescriptions slightly decreased

from 18.3% to 17.2%. [46] It would be interesting to see whether the prevalence of antibiotic description in the Netherlands has also changed over time.

2 *The prescription of benzodiazepines in the treatment of non-specific lower back pain.* [23] Lower back pain is one of the most prevalent conditions in general practice. [58] Its treatment, however, is complex. [59] Studies indicate that both inappropriate imaging [11, 12, 60, 61] and prescribing of opioids [59, 62, 63] are highly prevalent. However, there is a lack of information about the prevalence of inappropriate prescribing for LBP in the Netherlands.

3 *The chronic prescription or continuation of acid-reducing medication (ARM) without indication.* [24] Inappropriate prescription of ARM, predominantly Proton Pump Inhibitors (PPIs), has been shown to be an international problem. [49-54] However, the extent of this problem in the Netherlands is unknown.

Supplementary box 1 contains a detailed description of how each recommendations was operationalized and the used definitions.

Methods

Design and database

We conducted a retrospective cohort study with data derived from the Nivel Primary Care Database (Nivel-PCD). The Nivel-PCD contains care data routinely collected from electronic medical records from GPs throughout the Netherlands. The data were obtained from 529 GP practices, representing approximately 2 million registered patients. [25] The sample was shown to be representative for the total population of Dutch primary care practices. [26-28] The database contains longitudinal information regarding patient characteristics such as age, sex, GP consultations, diagnoses, and drug prescriptions. Diagnoses are recorded using the International Classification of Primary Care version 1 (ICPC-1). Prescriptions are recorded using the Anatomical Therapeutic Chemical classification system (ATC). This study was approved by the relevant governance bodies of the Nivel-PCD (nr. NZR00320.001).

Cohort selection

For each low-value care examined, patients with relevant episodes were extracted from the Nivel-PCD. Next, the prescription files for each type of pharmaceutical care were filtered for prescriptions associated with relevant ICPC codes. The resulting selection was then used in our analysis. Supplementary box 1 contains an overview of the ATC and ICPC codes used to define the patient population for each of the recommendations. We only included GP practices in our analysis for which sufficient prescription data of high quality

were available between 2016 and 2019. GP practices had to meet the following criteria to be included in our analysis: 1) At least 85% of the prescriptions were encoded with a valid ATC code; 2) A minimum of 46 weeks of prescription data had to be present and; 3) A minimum of 500 patients per practice should be present in the data.

Data analyses

Our assessments were performed using a patient-indication lens, as described by Chalmers et al. [29] Meaning we only included patients with a specific indication in our denominators. Our primary outcome is the percentages of patients, with an indication, who received a low-value prescription at least once. Analysis was performed using STATA 16 [30]. Data visualisation was carried out using R-V3.6.3. [31] and the R-package ggplot2. [32]

Practice variation

Variation among GP practices was assessed through multilevel logistic regression analysis over 2019, with random effects on the practice level. Before performance of multilevel logistic regression, variance inflation factors (VIF) were calculated to test for collinearity among the included variables. In order to prevent the standard errors of the (multilevel) regression coefficients becoming too large, GP practices with fewer than five cases of low-value prescribing, or fewer than 30 patients with a relevant indication for the type of care examined, were excluded from the analysis. Intraclass correlation coefficients (ICC) were calculated to assess variation in low-value prescribing between GP practices. [33] C-statistics were calculated for models with, and without, a random effect for the level of the practice. The presence of higher c-statistics associated with the models with a random intercept for the level of practice suggest that these models have more predictive accuracy compared to the models without the random intercept. The difference between both C-statistics was used as a measure for variation among GP practices. Additionally, we have also assessed whether or not a significant correlation exists between the prescription rates of the three types of low-value pharmaceutical care over 2019. Correlations were assessed using the Pearson correlation coefficient for normally distributed variables and the Spearman correlation coefficient for non-normally distributed variables. Normality was assessed using both density plots and the Shapiro-Wilk test (supplementary table 2).

Case mix variables

The following patient variables were included in our models in order to assess their influence on the odds of receiving low-value care: age, sex, and neighbourhood socio-economic status (SES). These case mix variables were selected based on previous research indicating that these factors affect the amount of care patients require, receive, and have access to. [34-38] SES scores from 2017 were derived from the Dutch Institute for Social Research (SCP). [39] Patients were assigned to one of five categories (lowest, below

average, average, above average, highest) based on quintiles. In addition to these patient characteristics, the number of patients registered at each GP practice was also included in our analysis and categorised as a small, medium or large practice, based on the division of each population into tertiles (supplementary table 5 contains the tertile boundaries).

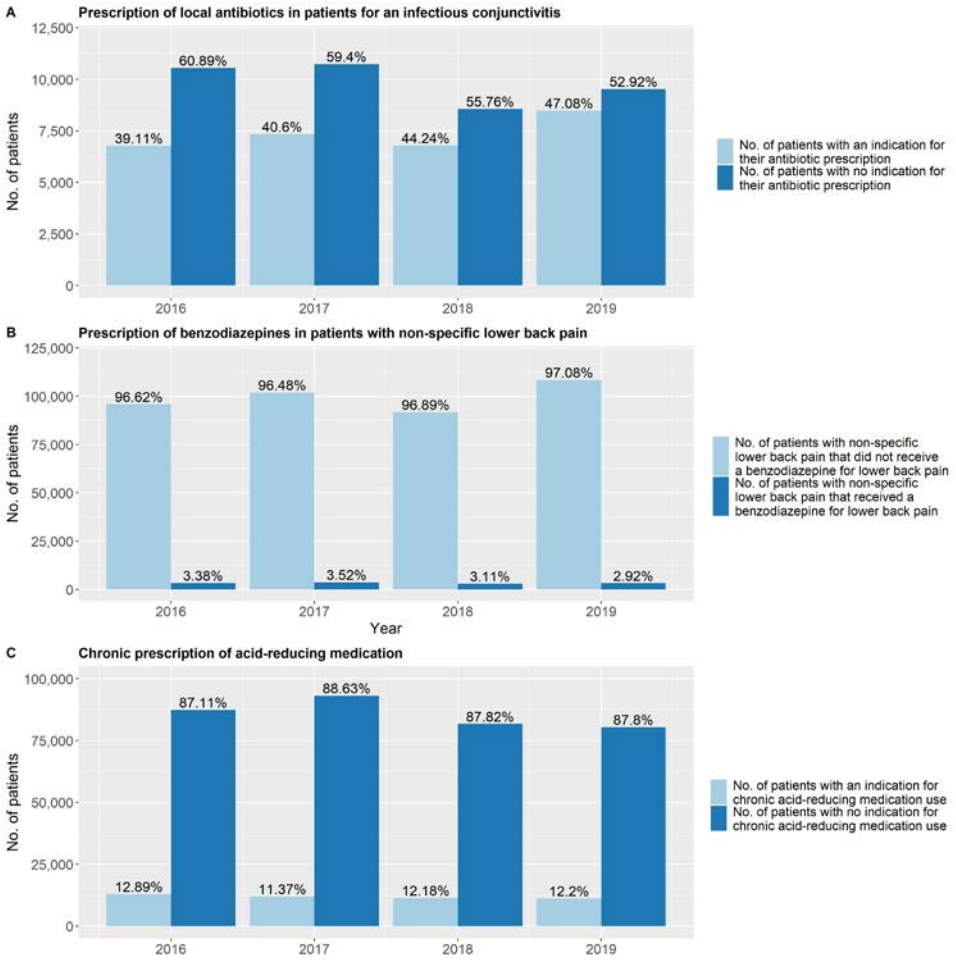


Figure 1 | Estimates of the prevalence of patients receiving one of three types of low-value care.

Results

Figure 1 and table 1 provide a summary of the study results. Patients with an episode of infectious conjunctivitis were regularly prescribed local antibiotics without appropriate indication. The proportion of patients inappropriately prescribed antibiotics decreased

Table 1 | Overview of assessment outcomes.

1. Do not prescribe a local antibiotic for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient.								
	2016		**2017**		**2018**		**2019**	
No. of practices included	316		329		296		346	
No. of patients with (at least) one episode of conjunctivitis (denominator)	17,332		18,076		15,345		17,994	
	Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
No. of patients whom received appropriate treatment	6,778	39.11	7,339	40.6	6,789	44.24	8,472	47.08
No. of patients with no clear indication for their antibiotic prescription (numerator)	10,554	60.89	10,737	59.4	8,556	55.76	9,522	52.92
2. Do not prescribe benzodiazepines in patients with non-specific lower back pain								
	2016		**2017**		**2018**		**2019**	
No. of practices included	313		328		296		346	
No. of patients with (at least) one episode of lower back pain (denominator)	99,262		105,641		94,685		111,703	
	Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
No. of patients whom did not receive a benzodiazepine or received one for a different indication	95,909	96.62	101,922	96.48	91,742	96.89	108,441	97.08
No. of patients with no clear indication for benzodiazepine use (numerator)	3,353	3.38	3,719	3.52	2,943	3.11	3,262	2.92
3. Do not chronically prescribe or continue acid-reducing medication, without proper indication								
	2016		**2017**		**2018**		**2019**	
No. of practices included	284		276		250		245	
No. of patients with chronic prescription of ARM (denominator)	100,319		105,043		93,053		91,563	
	Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
No. of patients with an indication for chronic ARM use	12,931	12.89	11,941	11.37	11,334	12.18	11,174	12.20
No. of patients with no clear indication for chronic ARM use (numerator)	87,388	87.11	93,102	88.63	81,719	87.82	80,389	87.80

from 61% to 53% between 2016 and 2019. The chronic use of ARMs without an appropriate indication was highly prevalent. Between 2016 and 2019, around 88% of patients with a chronic ARM prescription lacked an appropriate indication. The prescription of benzodiazepines for LBP remained around 3% over the four years. A more detailed table is presented in supplementary table 1.

Variation on the practice level

Figure 2 shows the prevalence of low-value prescribing among GP practices during 2019. We observed large variation in the proportion of patients receiving at least one non-indicated prescription for antibiotics for conjunctivitis (figure 2A). This varied between 0 and 90.3% (median: 52.8%). Benzodiazepines were prescribed largely in line with the guidelines, showing limited variation. Between 0 and 11% (median: 3.0%) of the LBP patients at each of the included GP practices received an inappropriate prescription. ARMs were prescribed chronically without an appropriate indication, in between 79% and 97% of the GP practices included (median: 88%). Comparison of the rates of non-indicated prescription of the three low-value pharmaceutical GP care across practices only revealed a significant weak positive correlation (correlation coefficient: 0.17) between the rate of low-value antibiotic and low-value benzodiazepine prescriptions (supplementary table 2). No other significant correlations were identified. Supplementary table 3 contains an overview of the rates of low-value care for each of the examined types of care across practices in 2019. After adjusting for case mix variables, the ICCs on the practice level for each of the prescriptions for low-value care ranged from 5.55% to 10.24% (Table 2). Analysis of the VIF factors revealed that little or no collinearity exists among the variables included in our analysis (supplementary table 4). The C-statistics of the models with a random effect on the practice level were significantly higher for all three types of low-value GP care examined, compared to the models without random effect. Supplementary table 5 presents an overview of the contribution of each case mix variable to the final model.

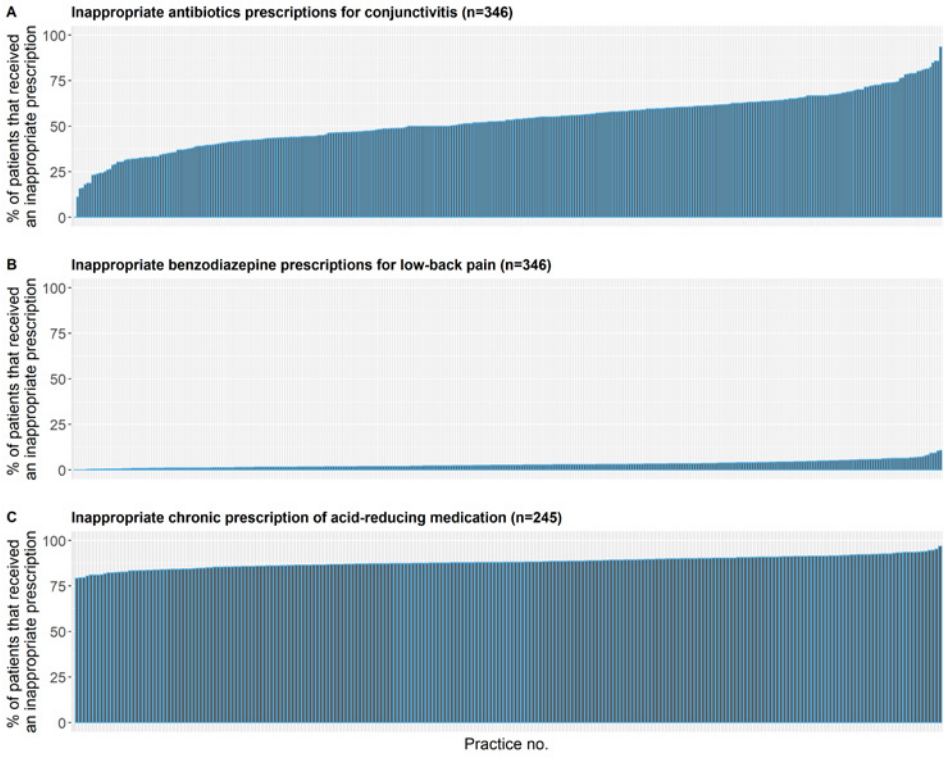


Figure 2 | Proportion of patients of each practice who received each of the low-value prescriptions for care at least once during 2019. The practice numbers do not directly correlate to the practice numbers as provided in supplementary table 3.

Table 2 | Overview of model characteristics; Intraclass correlation coefficient (ICC); Confidence Intervals (CIs).

Recommendation	ICC practice (and 95% CIs)	C-statistic model with random effect practice (and 95% CIs)	C-statistic model without random effect practice (and 95% CIs)
1. Do not prescribe a local antibiotic for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient.	0.08 (0.06 - 0.11)	0.65 (0.64 - 0.66)	0.54 (0.53 - 0.55)
2. Do not prescribe benzodiazepines in patients with non-specific lower back pain	0.10 (0.07 - 0.15)	0.67 (0.66 - 0.68)	0.59 (0.58 - 0.60)
3. Do not prescribe or continue acid-reducing medication, without proper indication	0.06 (0.03 - 0.09)	0.72 (0.71 - 0.72)	0.70 (0.69 - 0.70)

Patient characteristics associated with receiving low-value prescriptions

The inappropriate prescription of antibiotics for conjunctivitis, and benzodiazepine for non-specific LBP, showed a significant increase in odds with increasing age ($p<0.001$, supplementary table 5). Conversely, patients were less likely to receive an inappropriate chronic prescription of ARMs with increasing age ($p<0.001$). Sex ($p<0.001$) and SES ($p<0.05$) significantly affected the odds of receiving a non-indicated chronic ARM prescription. Females were slightly less prone to receiving a non-indicated chronic ARM prescription. Patients showed significantly increased odds of receiving an inappropriate ARM with increasing SES. Furthermore, SES had only a small significant increase in odds of receiving an inappropriate antibiotic for conjunctivitis, when comparing the average SES with the lowest category ($P<0.05$). The size of the GP practice significantly affected the odds of receiving any of the three types of low-value care examined. In general, larger GP practices are less prone to providing any of the three types of low-value care compared to smaller ones ($P<0.05$). Only in cases of the chronic ARM use, it appeared that medium-sized practices did not significantly differ in odds from the smaller ones (Figure 3).

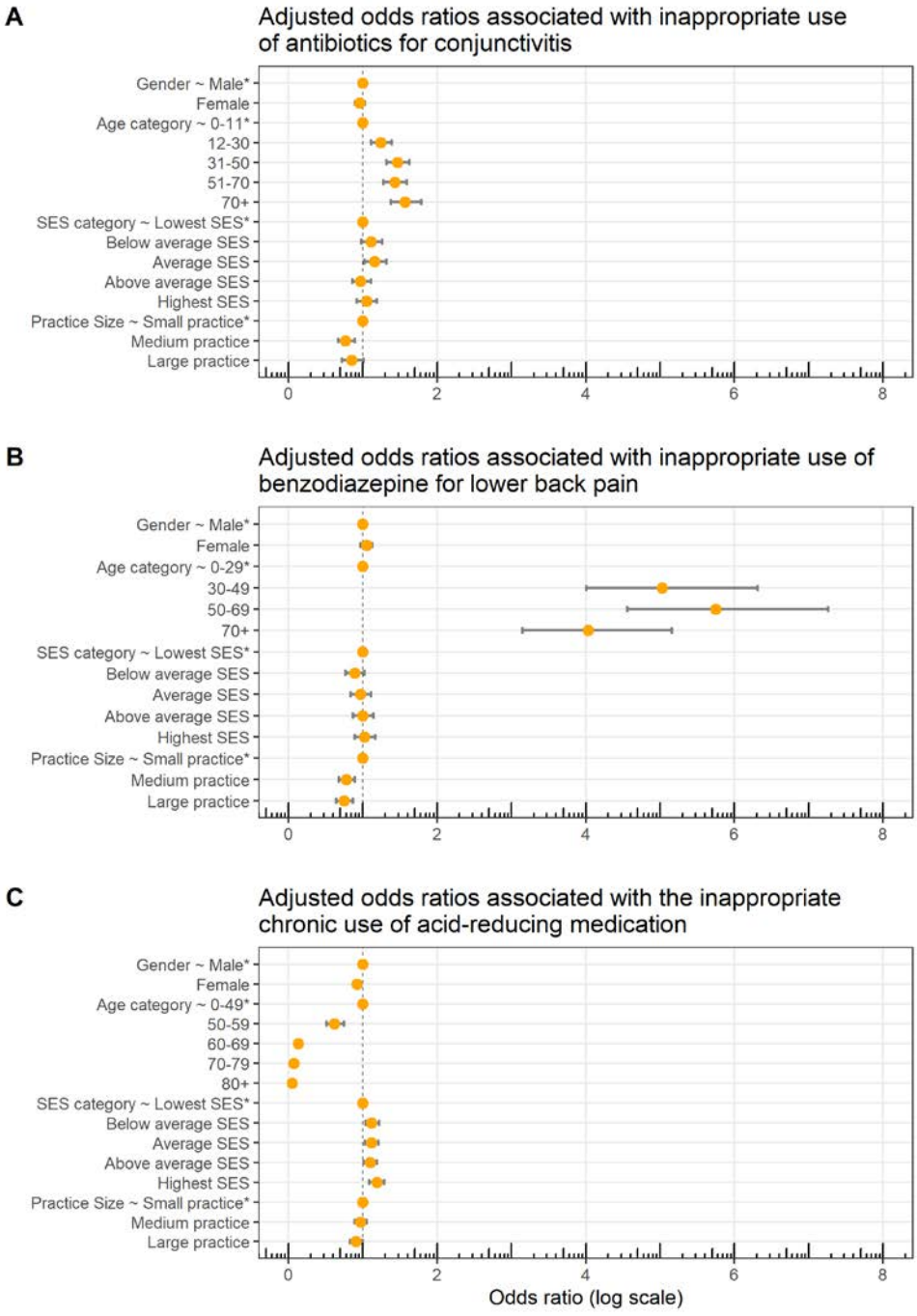


Figure 3 | Adjusted odds ratios and confidence intervals associated with patient characteristics for all three types of low-value GP care. The asterisk indicates the reference categories.

Discussion

Our study shows that the prescription of low-value pharmaceutical GP care varies depending upon the clinical problem. Inappropriate prescriptions of both antibiotics for conjunctivitis and ARMs are highly prevalent, while the proportion of patients with LBP receiving a benzodiazepine is small. Large variation in pharmaceutical treatment was found for the prescription of a non-indicated antibiotic for conjunctivitis, whereas limited variation was found in the inappropriate prescription of benzodiazepines for lower back pain or non-indicated chronic ARM prescriptions. Our analysis of correlation among the practices over 2019 only revealed a significant weak positive correlation between the rate of low-value antibiotic and benzodiazepine prescriptions. The odds of a patient receiving either of the three low-value treatments are significantly affected by age. Men were found to have significantly higher odds of receiving a non-indicated chronic ARM prescription compared to females, while the odds of patients receiving an inappropriate ARM significantly increased with increasing neighbourhood SES. Apart from ARMs, SES only showed a small significant increase in the odds of receiving an inappropriate antibiotic for conjunctivitis, when comparing the average with the lowest category. The odds of receiving an inappropriate antibiotic or benzodiazepine significantly decreased as the size of the GP practice increased. ARMs also showed a similar decrease as the size of the GP practice increased. However, this was only found to be significant in cases of the largest practice sizes.

Strengths and limitations

A strength of this study is that low-value care among Dutch GPs was assessed of patients receiving an inappropriate prescription among using routinely collected, nationally representative, data over four consecutive years. Furthermore, the use of high quality and complete clinical data made it possible to distinguish appropriate from inappropriate care.

Our study has some limitations. First, there is an inherent uncertainty in identifying whether a prescription is of low-value. Recommendations contain terms that do not map directly to data variables; also, diagnosis and procedure codes may not precisely identify patients for whom care is of low value. For example, the recommendations regarding conjunctivitis and ARMs were not described with enough detail or required variables which are absent in the data to distinguish appropriate from inappropriate prescribing. The recommendation regarding conjunctivitis requires us to identify patients with conjunctivitis caused by a banal pathogen, and who are at high-risk. The information required to distinguish the cause of an episode of conjunctivitis is not recorded within the Nivel-PCD, and therefore not available in the data used. We also could not identify patients whom were at high risk, since the recommendation and the guideline did not provide sufficient detail on the definition of the high-risk population, to be able to distinguish them (if the information was available). Our assessment therefore could be

an overestimate. However, we do not expect these factors to have a major effect on our assessment outcome, since conjunctivitis is most commonly caused by a banal pathogen. [40] Furthermore, in our analysis of chronic ARM prescriptions, the guideline states that gastro-protection using an non-selective NSAID is justified if a patient is using a high dosage of a NSAID. However, information regarding the dosage of the prescribed NSAIDs was not present within the data used. We therefore could not include these requirements in our assessment. Our assessments might therefore under, or over, estimate the magnitude of the problem. Secondly, we could not identify patients suffering from chronic heartburn, while we only had access to diagnosis established within the years examined. Patients diagnosed with heartburn outside of this period could therefore not be identified. Also, heartburn is often only present for a short period of time, until ARMs are prescribed. The prescription of ARMs often resolve the patients' complaints resulting in removal of the diagnosis from the patients' medical records, making it difficult to define chronic heartburn. For similar reasons, we did not have access to practice and physician characteristics, which could explain variation in prescribing behaviour between GP practices, for example, the number of physicians and their age or sex. [41-43] Thirdly, both the antibiotic and benzodiazepine recommendations are directly linked to specific diagnoses, thereby making their assessment relatively straightforward. However this is not the case for the recommendation of chronic ARMs use, which makes its assessment difficult and more uncertain in comparison to the other recommendations. Finally, our final logistic models reported moderate C-statistics. This suggests that these models are unsuitable for predicting, reliably, the risk of patients receiving either of the low-value GP prescriptions. Receiving low-value GP care could have been influenced by other patient or GP characteristics that are not available in the data. [44]

Comparison with existing literature

Previous studies from Australia, the US, and the Netherlands reported between 60% and 80% of patients with infectious conjunctivitis received a non-indicated antibiotics prescription, which is higher compared to our assessment outcome. [20, 45, 46] The differences between the assessments could be explained by differences in the data sources, study designs and populations included. For example, Shekhawat et al. used a patient-indication lens and included all GPs in their analysis, while Cherry et al. included only patients who visited GP registrars in their sample resulting in distinct assessment denominators and outcomes. [45, 46]

Our assessment of low-value prescribing of benzodiazepines for non-specific LBP shows that Dutch GPs mostly adhere to professional guidelines. Only 3% of patients received an inappropriate prescription, which is lower compared to the findings of recent studies from the US. [47, 48] Agarwal et al. reported that 8.5% of patients with back or chronic pain received a benzodiazepine prescription. [47] Furthermore, Azad et al., reported that 11.5% of US patients new to opioids with LBP, received a benzodiazepine within 12 months of their diagnosis. [48] However, it is difficult to compare these assessments with our study as

patient-population lenses were used in the US studies, while our study applied a patient-indication lens resulting in different denominators being used. [15, 29]

We did not identify any assessments regarding non-indicated chronic ARM use within the literature as most studies focus solely on chronic Proton Pump Inhibitor (PPI) use. However, since only approximately 3.5% of all ARM prescriptions in our assessment did not concern PPIs, we feel our assessment closely resembles those that solely focus on PPIs. Our assessment shows that the non-indicated chronic use of ARMs is highly prevalent in the Netherlands. However, this is only slightly higher compared to what is reported in international literature. According to recent literature between 30% to 80% of PPI prescriptions have no appropriate indication. [49-55] Again, these high levels of variation could be explained by differences in population characteristics, study design – such as the inclusion of all ARMs – and setting. Furthermore, unlike other studies, we did not limit our assessment to a specified population; such as the elderly [51, 56]. According to Dutch and international guidelines, the elderly have more indications justifying the use of a PPI which might have affected the prevalences reported.

The high amount of inappropriate antibiotic prescriptions could be explained by GPs experiencing patient pressure to provide low-value care. Previous research, shows that patient pressure and the GPs need to maintain a good patient-physician relationship, could induce inappropriate prescribing. [57] Furthermore, the low-level of inappropriate benzodiazepine prescriptions could be explained by a long-term policy, promoting cautious prescribing of benzodiazepines due to their addictive properties. Finally, the high prevalence of inappropriate chronic ARM prescriptions could be due to their reputation, at least, for being harmless. ARMs, in contrast to antibiotics and benzodiazepines, are commonly sold over-the-counter at most drugstores in the Netherlands. We therefore expect that our assessment is most likely an underestimate, since we were not able to capture all chronic ARM users in this study, thereby missing non-prescription ARM in our assessment.

The implications for research and/or practice

Our assessment of three low-value pharmaceutical GP prescriptions demonstrates that both antibiotics for conjunctivitis, and the chronic use of ARM, are prescribed inappropriately with no indication that this will greatly decline over time. This suggests a joint national effort is required to change prescribing behaviour. For such an effort, detailed insight into the views of patients and prescribers, and the barriers and facilitators for the withdrawal of inappropriate medication – deprescribing – is required in order to design a tailor-made deprescribing strategy. Furthermore, our observation that little to no correlation exists between the low-value prescription rates of the three types of low-value pharmaceutical care within practices, suggests that the problem of low-value pharmaceutical GP care cannot be addressed through a single deprescribing strategy. But rather requires deprescribing strategies that are tailor-made to the type

of pharmaceutical care that one aims to address. The assessment methods from this study could be used to monitor changes resulting from any following interventions. Furthermore, the knowledge of potential patient characteristics that are associated with the increased odds of receiving any of the low-value prescriptions examined could provide some focus for policy interventions.

Conclusion

Our research shows that low-value pharmaceutical care is prevalent among Dutch GPs, but its prevalence varies depending on the clinical problem. Between 2016 and 2019, many patients received an inappropriate antibiotic or chronic ARM prescription. Benzodiazepines for LBP are generally prescribed in line with the guidelines. Among the three types of low-value pharmaceutical care, large variation between GP practices and general variation in prescribing were observed. These insights could help in designing a national campaign in order to change this behaviour.

References

1. Colla CH, Mainor AJ, Hargreaves C, et al. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. Med Care Res Rev. 2017;74(5):507-50.
2. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. Lancet. 2017;390(10090):156-68.
3. Dutch College of General Physicians (NHG). NHG-standaarden 2021 [last accessed 29-11-2021]. Available from: https://richtlijnen.nhg.org/.
4. Kool RB, Verkerk EW, Meijs J, et al. Assessing volume and variation of low-value care practices in the Netherlands. Eur J Public Health. 2020.
5. Linden MVD, Westert G, Bakker D, Schellevis F. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk: klachten en aandoeningen in de bevolking en in de huisartspraktijk2004 [last accessed 29-11-2021]. Available from: https://www.nivel.nl/nl/publicatie/tweede-nationale-studie-naar-ziekten-en-verrichtingen-de-huisartspraktijk-klachten-en.
6. Volkers A, van Dijk L, Wolters I, Bakker D. Formularia in de huisartspraktijk: gebruik en het opvolgen van adviezen2009 [last accessed 09-09-2021]. Available from: https://www.nivel.nl/nl/publicatie/formularia-de-huisartspraktijk-gebruik-en-het-opvolgen-van-adviezen.
7. Bruin-Huisman L, Abu-Hanna A, van Weert HCPM, Beers E. Potentially inappropriate prescribing to older patients in primary care in the Netherlands: a retrospective longitudinal study. Age Ageing. 2017;46(4):614-9.
8. Teichert M, Schermer T, van den Nieuwenhof L, et al. Prevalence of inappropriate prescribing of inhaled corticosteroids for respiratory tract infections in the Netherlands: a retrospective cohort study. npj Prim Care Respir Med. 2014;24(1):14086.
9. Lugtenberg M, Zegers-van Schaick JM, Westert GP, Burgers JS. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. Implement Sci. 2009;4:54.
10. Simon E, Miake-Lye IM, Smith SW, et al. An Evaluation of Guideline-Discordant Ordering Behavior for CT Pulmonary Angiography in the Emergency Department. J Am Coll Radiol. 2019;16(8):1064-72.
11. Schwartz AL, Landon BE, Elshaug AG, et al. Measuring low-value care in Medicare. JAMA Intern Med. 2014;174(7):1067-76.
12. Bouck Z, Pendrith C, Chen XK, et al. Measuring the frequency and variation of unnecessary care across Canada. BMC Health Serv Res. 2019;19(1):446.
13. Colla CH, Morden NE, Sequist TD, et al. Choosing wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2015;30(2):221-8.
14. Charlesworth CJ, Meath TH, Schwartz AL, McConnell KJ. Comparison of Low-Value Care in Medicaid vs Commercially Insured Populations. JAMA Intern Med. 2016;176(7):998-1004.
15. Müskens JLJM, Kool RB, van Dulmen SA, Westert GP. Overuse of diagnostic testing in healthcare: a systematic review. BMJ Qual Saf. 2021:bmjqs-2020-012576.
16. Pendrith C, Bhatia M, Ivers NM, et al. Frequency of and variation in low-value care in primary care: a retrospective cohort study. CMAJ Open. 2017;5(1):E45-e51.

17. McAlister FA, Lin M, Bakal J, Dean S. Frequency of low-value care in Alberta, Canada: A retrospective cohort study. BMJ Qual Saf. 2018;27(5):340-6.
18. Grimshaw JM, Patey AM, Kirkham KR, et al. De-implementing wisely: developing the evidence base to reduce low-value care. BMJ Qual Saf. 2020;29(5):409-17.
19. Schers H, Braspenning J, Drijver R, et al. Low back pain in general practice: reported management and reasons for not adhering to the guidelines in The Netherlands. Br J Gen Pract. 2000;50(457):640-4.
20. Rietveld RP, ter Riet G, Bindels PJ, et al. Do general practitioners adhere to the guideline on infectious conjunctivitis? Results of the Second Dutch National Survey of General Practice. BMC Fam Pract. 2007;8:54.
21. Hofstede H, van der Burg HAM, Mulder BC, et al. Reducing unnecessary vitamin testing in general practice: barriers and facilitators according to general practitioners and patients. BMJ Open. 2019;9(10):e029760.
22. Nederlands Huisartsen Genootschap. NHG-Standaard: Rood oog en oogtrauma: NHG; 2017 [last accessed 29-11-2021]. Available from: https://richtlijnen.nhg.org/standaarden/rood-oog-en-oogtrauma.
23. Nederlands Huisartsen Genootschap. NHG-Standaard: Aspecifieke lagerugpijn: NHG; 2017 [last accessed 29-11-2021]. Available from: https://richtlijnen.nhg.org/standaarden/aspecifieke-lagerugpijn#volledige-tekst.
24. Nederlands Huisartsen Genootschap. NHG-Standaard: Maagklachten: NHG; 2013 [last accessed 29-11-2021]. Available from: https://richtlijnen.nhg.org/standaarden/maagklachten.
25. Hendriksen J, Hooiveld, M., Nielen M., Hek, K Methoden vaststellen cijfers Nivel Zorgregistraties Eerste Lijn 2021 [Available from: https://www.nivel.nl/nl/zorgregistraties-eerste-lijn/methoden.
26. Biermans MC, Verheij RA, de Bakker DH, et al. Estimating morbidity rates from electronic medical records in general practice. Evaluation of a grouping system. Methods Inf Med. 2008;47(2):98-106.
27. Gijsen R, Poos MJ. Using registries in general practice to estimate countrywide morbidity in The Netherlands. Public Health. 2006;120(10):923-36.
28. Nivel zorgregistraties. Nivel Zorgregistraties. 2019.
29. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
30. LLC S. Stata Statistical Software: Release 16. In: LLC S, editor. 16 ed. College Station, TX: StataCorp LLC.: StataCorp LLC; 2019. p. StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.
31. R Core Team. R: A language and environment for statistical computing. v3.6.3 ed. Vienna, Austria: R foundation for Statistical Computing; 2019.
32. Wickham H. ggplot2: elegant graphics for data analysis. In: Springer New York, editor. 2009.
33. Snijders TAB, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling: SAGE Publications; 2011.

34. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. Front Public Health. 2020;8:231.
35. Arpey NC, Gaglioti AH, Rosenbaum ME. How Socioeconomic Status Affects Patient Perceptions of Health Care: A Qualitative Study. J Prim Care Community Health. 2017;8(3):169-75.
36. Asch SM, Kerr EA, Keesey J, et al. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
37. David JL, Kaplan HB. Sex, social roles and health care utilization. Appl Behav Sci Rev. 1995;3(1):39-64.
38. Bertakis KD, Azari R, Helms LJ, et al. Sex differences in the utilization of health care services. J Fam Pract. 2000;49(2):147-52.
39. The Netherlands Institute for Social Research. Statusscores 2017. Netherlands: The Netherlands Institute for Social Research (SCP); 2017.
40. Azari AA, Barney NP. Conjunctivitis: a systematic review of diagnosis and treatment. Jama. 2013;310(16):1721-9.
41. Haastrup PF, Rasmussen S, Hansen JM, et al. General practice variation when initiating long-term prescribing of proton pump inhibitors: a nationwide cohort study. BMC Fam Pract. 2016;17:57.
42. Orzella L, Chini F, Giorgi Rossi P, Borgia P. Physician and patient characteristics associated with prescriptions and costs of drugs in the Lazio region of Italy. Health Policy. 2010;95(2-3):236-44.
43. Jayaweera A, Chung Y, Jabbarpour Y. Primary Care Physician Characteristics Associated with Prescribing Potentially Inappropriate Medication for Elderly Patients: Medicare Part D Data. J Am Board Fam Med. 2020;33(4):561-8.
44. Barreto TW, Chung Y, Wingrove P, et al. Primary Care Physician Characteristics Associated with Low Value Care Spending. J Am Board Fam Med. 2019;32(2):218-25.
45. Cherry MD, Tapley A, Quain D, et al. Antibiotic prescribing patterns of general practice registrars for infective conjunctivitis: a cross-sectional analysis. J Prim Health Care. 2021;13(1):5-14.
46. Shekhawat NS, Shtein RM, Blachley TS, Stein JD. Antibiotic Prescription Fills for Acute Conjunctivitis among Enrollees in a Large United States Managed Care Network. Ophthalmology. 2017;124(8):1099-107.
47. Agarwal SD, Landon BE. Patterns in Outpatient Benzodiazepine Prescribing in the United States. JAMA Netw Open. 2019;2(1):e187399.
48. Azad TD, Zhang Y, Stienen MN, et al. Patterns of Opioid and Benzodiazepine Use in Opioid-Naïve Patients with Newly Diagnosed Low Back and Lower Extremity Pain. J Gen Intern Med. 2020;35(1):291-7.
49. Blackett JW, Chen L, Li J, et al. Inappropriateness of Proton Pump Inhibitors After Hospital Discharge Is Associated with Thirty-Day Hospital Readmission. Dig Dis Sci. 2021.
50. Leri F, Ayzenberg M, Voyce SJ, et al. Four-year trends of inappropriate proton pump inhibitor use after hospital discharge. South Med J. 2013;106(4):270-3.
51. Mafi JN, May FP, Kahn KL, et al. Low-Value Proton Pump Inhibitor Prescriptions Among Older Adults at a Large Academic Health System. J Am Geriatr Soc. 2019;67(12):2600-4.

52. Pasina L, Nobili A, Tettamanti M, et al. Prevalence and appropriateness of drug prescriptions for peptic ulcer and gastro-esophageal reflux disease in a cohort of hospitalized elderly. Eur J Intern Med. 2011;22(2):205-10.
53. McDonald EG, Jones J, Green L, et al. Reduction of inappropriate exit prescriptions for proton pump inhibitors: A before-after study using education paired with a web-based quality-improvement tool. J Hosp Med. 2015;10(5):281-6.
54. Giannini EG, Crespi M, Djahandideh A, et al. Appropriateness of proton pump inhibitors treatment in clinical practice: Prospective evaluation in outpatients and perspective assessment of drug optimisation. Dig Liver Dis. 2020;52(8):862-8.
55. Fleming-Dutra KE, Hersh AL, Shapiro DJ, et al. Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011. JAMA. 2016;315(17):1864-73.
56. Voukelatou P, Vrettos I, Emmanouilidou G, et al. Predictors of Inappropriate Proton Pump Inhibitors Use in Elderly Patients. Curr Gerontol Geriatr Res. 2019;2019:7591045.
57. Kool RB, Verkerk EW, Winnemuller LJ, et al. Identifying and de-implementing low-value care in primary care: the GPs perspective-a cross-sectional survey. BMJ Open. 2020;10(6):e037019.
58. Vos T, Abajobir AA, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.
59. Foster NE, Anema JR, Cherkin D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet. 2018;391(10137):2368-83.
60. Mafi JN, Reid RO, Baseman LH, et al. Trends in Low-Value Health Service Use and Spending in the US Medicare Fee-for-Service Program, 2014-2018. JAMA Netw Open. 2021;4(2):e2037328.
61. Gold R, Esterberg E, Hollombe C, et al. Low Back Imaging When Not Indicated: A Descriptive Cross-System Analysis. Perm J. 2016;20(2):25-33.
62. Jeffrey Kao MC, Minh LC, Huang GY, et al. Trends in ambulatory physician opioid prescription in the United States, 1997-2009. Pm r. 2014;6(7):575-82.e4.
63. Deyo RA, Von Korff M, Duhrkoop D. Opioids for low back pain. BMJ. 2015;350:g6380.

Supplementary box 1 | Operationalization of the recommendations: ATC and ICPC codes used to define the study populations for each of the examined recommendations.

Recommendation	Additional information	ICPC codes relevant for the analysis	ATC codes relevant for the analysis
1. Do not prescribe local antibiotics for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient. • Numerator: no. of patients with (at least) one episode of conjunctivitis. • Denominator: no. of patients that received an antibiotic prescription without having a clear indication during their episode of conjunctivitis.	- Based on the feedback received from the representatives of the NHG we made the assumption that all identified episodes of conjunctivitis were caused by a banal pathogen. - We were unable to distinguish high-risk patients based on the information that was provided in either the guideline and database. We therefore did not make this distinction in our assessment.	F70: Conjunctivitis Infectious	**Antibiotics:** **Excluded:** J01: Antibacterials for systemic use **Included:** S01: Ophthalmologicals S03: Ophthalmologicals and Otological preparations
2. Do not prescribe benzodiazepines in patients with non-specific lower back pain. • Numerator: no. of patients with (at least) one episode of lower back pain. • Denominator: no. of patients that received a benzodiazepine without having a clear indication during their episode of lower back pain.		L02: back symptoms / complaints L03: low back symptoms / complaints L86: back syndrome with radiating pain	**Benzodiazepines:** N05BA: Benzodiazepine derivatives N05CD: Benzodiazepine derivatives N05CF: Benzodiazepine related drugs

3. Do not chronically prescribe or continue acid-reducing medication (ARM), without proper indication. • Numerator: no. of patients with a chronic prescription of ARM. • Denominator: no. of patients with no clear indication for their chronic ARM prescription. A more elaborate description of the different conditions in which ARMs are indicated is presented below this table.	- We defined chronic acid-reducing medication users as individuals that received an acid-reducing medication for at least 180 days in the previous year. - We were unable to identify chronic ARM users suffering from chronic heartburn. - We were unable to identify patients using high-dose high dose non-selective NSAID, while no clear definition of high-dose was present in the guideline nor did we have information regarding doses of prescriptions.	D85: Duodenal ulcer D86: Peptic ulcer other D03: Heartburn L88: Rheumatoid / seropositive arthritis K77: Hearth failure T90: Diabetes non-insulin dependent	**Co-medications associated with increased risk of stomach complications:** • Coumarin derivative (B01AA) • Direct oral anticoagulants (B01AF02, B01AE07, B01AF03, B01AF01) • P2Y12-inhibitors (B01AC04, B01AC22, B01AC24) • Acetylsalicylic acid derivatives: Acetylsalicylic acid (Aspirin): A01AD05, B01AC06, B01AC56, C10BX01, C10BX02, C10BX04, C10BX05, M01BA03, N02BA01, N02BA51, N02BA71. • Systemic glucocorticoids (H02AB) • Selective serotonin reuptake inhibitors (SSRI) (N06AB) • Venlafaxine (N06AX16) • Duloxetine (N06AX21) • Trazodone (N06AX05) • Spironolactone (C03DA01) **Acid-reducing medication:** A02BA: H_2-receptor antagonists A02BB: Prostaglandins A02BC: Proton pump inhibitors M01AB: Acetic acid derivatives and related substances M01AE52: Naproxen and misoprostol

More elaborate description of the ARM recommendation

Do not chronically prescribe or continue acid-reducing medication (ARM), without proper indication. ARM prescriptions are indicated in the following cases according to the guideline:

- Gastro-protection with an proton pump inhibitor (PPI) in case of a non-selective nonsteroidal anti-inflammatory drug (NSAID)
  i. Age of 70 years or older;
  ii. Presence of an Ulcus Duodeni (D85) or Ulcus pepticum (D86) in their medical history, irrespective of their age.
  iii. When two or more of the following factors are applicable (the risk of complications increases with increasing number of factors present):
    - Age between 60 and 70 years.
    - Severe disabling rheumatoid arteritis (L88), Hearth failure (K77) or diabetes (T90).
    - Use of high dose non-selective NSAID
    - Use of comedication which increase the risk of stomach complications.

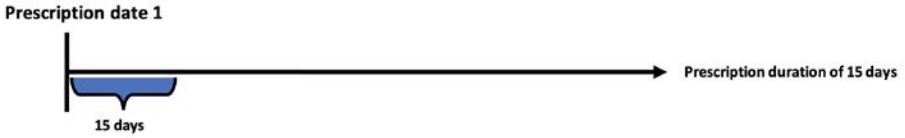
- Gastro-protection with an proton pump inhibitor (PPI) in case of a Acetylsalicylic acid derivative as platelet aggregation inhibitor and in absence of a non-selective NSAID is indicated in case of:
  i. Age 80 years or older.
  ii. Age 70 or older combined with use of comedication which increases the risk of stomach complications (except Acetylsalicylic acid derivatives).
  iii. Age of 60 or older combined with the presence of an Ulcus Duodeni (D85) or Ulcus pepticum (D86) in their medical history.

- Patients suffering from chronic heartburn, which do not sufficiently benefit from alternative acid-reducing medication (or in which these have not been tried).
  i. Heartburn (D03); we could not identify patients with chronic heartburn because we only received data from the requested data period and therefore only could identify patients which received a diagnosis of chronic heartburn within the 4 years examined.

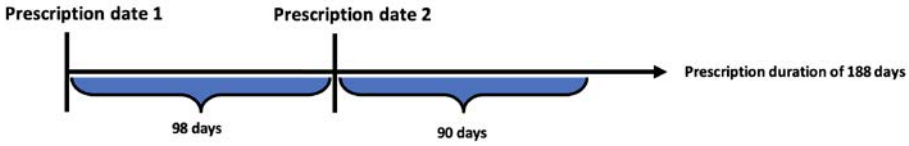
Calculation of prescription duration

Within Nivel-PCD only the dates on which the general practitioner prescribed the medication in question is recorded. The database does not contain the end date of a certain prescription. Therefore, in order to be able to define chronic use, we used the following method to calculate prescription duration:

When patients only have had a single prescription, we assumed these patients only received a 'start prescription'. In the Netherlands, a start prescription has a duration of 15 days and since there is only one prescription registered, we assumed that these patients only received this single prescription and therefore only received acid-reducing medication for the duration of 15 days.

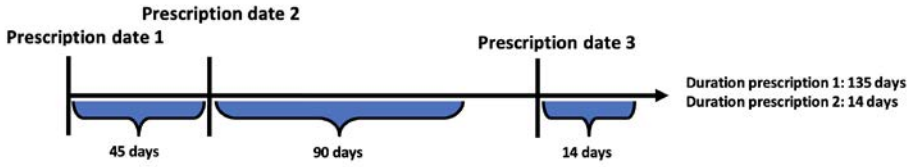


However, when a patient has received more than one prescription we have made the assumption that in between both dates the patient used the prescription continuously when the in between time is less or equal to 180 days. We chose to use 180 days since most prescriptions have a duration of 90 days, but medication is not always picked up after exactly 90 days. Furthermore, to assure that we also included patients that use their medication every other day instead of daily or patients which use their medication as needed we chose to double the prescription duration (e.g. 2 x 90 days) in order for those patients to be included. An additional 90 days are added to the latest prescription date of the two prescriptions, while a regular refill prescription has a duration of 90 days and the latest date of the two prescription dates us also the start date of the second prescription.



The same principle applies to a patient whom has 3 prescriptions, but who's prescriptions are less than 180 days apart. The duration between prescription 1 and 2 and the duration between prescription 2 and 3 are added up, and an additional 90 days are added for the duration of the third prescription.

In case a patient has more than 180 days between consecutive prescriptions, the calculate will be performed in the following manner:



When two prescriptions are more than 180 days apart, then a new prescription will commence which in turn will be treated according to the same rules as described above.

Supplementary table 1 | Extended table 1, number of patients separated by gender, and subdivided by age and SES categories for each of the three research questions.

1. Do not prescribe a local antibiotic for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient.

		2016		2017		2018		2019	
	No. of practices included	316		329		296		346	
	No. of patients with (at least) one episode of conjunctivitis	17,332		18,076		15,345		17,994	
		Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
	No. of patients with an indication(s) for their antibiotic prescription	6,778	39.11	7,339	40.6	6,789	44.24	8,472	47.08
	No. of patients with no indication for their antibiotic prescription	10,554	60.89	10,737	59.4	8,556	55.76	9,522	52.92
		Total	**Without (clear) indication**	**Total**	**Without (clear) indication**	**Total**	**Without (clear) indication**	**Total**	**Without (clear) indication**
Male	Overall no. of patients with at least one prescription included	7,715	4,719	7,871	4,674	6,789	3,829	7,924	4,225
Age cat.	0-9	1,654	932	1,849	1,022	1,420	692	1,690	787
	10-29	1,110	702	1,102	659	1,010	567	1,106	571
	30-49	1,883	1,151	1,816	1,117	1,600	934	1,857	1,017
	50-69	2,077	1,327	2,113	1,280	1,786	1,079	2,114	1,183
	70 +	991	607	991	596	973	557	1,157	667
Neighbourhood	*Lowest socio-economic status**	1,526	976	1,600	978	1,303	740	1,566	810
SES cat.	*Socio-economic status below average*	1,561	960	1,565	959	1,354	780	1,565	904
	Average socio-economic status	1,493	911	1,547	916	1,329	732	1,609	885
	Socio-economic status above average	1,475	891	1,524	879	1,362	774	1,543	795
	Highest socio-economic status	1,543	916	1,607	925	1,340	739	1,596	807
	Missing	117	65	28	17	101	64	45	24

Female	Overall no of patients with at least one prescription included	9,617	5,835	10,205	6,063	8,556	4,665	10,070	5,297
Age cat.	0-9	1,427	793	1,614	865	1,277	595	1,485	668
	10-29	1,632	986	1,615	958	1,409	732	1,661	850
	30-49	2,396	1,504	2,460	1,497	2,031	1,165	2,432	1,337
	50-69	2,723	1,702	2,921	1,788	2,540	1,445	2,878	1,543
	70 +	1,439	850	1,595	955	1,299	728	1,614	899
Neighbourhood	Lowest socio-economic status*	1,906	1,155	2,023	1,234	1,722	929	2,049	1,059
SES cat.	Socio-economic status below average	1,846	1,146	2,017	1,233	1,677	936	1,984	1,061
	Average socio-economic status	1,945	1,169	2,067	1,215	1,707	925	2,013	1,100
	Socio-economic status above average	1,904	1,134	2,082	1,234	1,693	949	2,005	1,030
	Highest socio-economic status	1,869	1,148	1,980	1,128	1,628	853	1,979	1,025
	Missing	147	83	36	19	129	73	40	22

2. Benzodiazepines for lower back pain

		2016		**2017**		**2018**		**2019**	
	No. of practices included	313		328		296		346	
	No. of patients with at least one episode of low-back pain	99,262		105,641		94,685		111,703	
		Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
	No. of patients with an indication for benzodiazepine use	95,909	96.62	101,922	96.48	91,742	96.89	108,441	97.08
	No. of patients with no indication for benzodiazepine use	3,353	3.38	3,719	3.52	2,943	3.11	3,262	2.92
	Missing	290	3	165	2	562	5	249	8

		Total	Without (clear) indication	Total	Without (clear) indication	Total	Without (clear) indication	Total	Without (clear) indication
Male	Overall no of patients with at least one prescription included	43,269	1,416	46,434	1,562	41,501	1,266	49,359	1,411
Age cat.	0-29	4,969	46	5,414	54	4,849	48	5,653	41
	30-49	12,719	511	13,561	506	11,860	436	13,623	451
	50-69	17,940	690	18,853	787	16,852	588	20,337	717
	70 +	7,641	169	8,606	215	7,940	194	9,746	202
Neighbourhood	*Lowest socio-economic status**	8,611	273	9,093	299	7,937	238	9,604	269
SES cat.	*Socio-economic status below average*	8,510	285	9,402	321	8,169	241	9,821	254
	Average socio-economic status	8,747	287	9,475	321	8,386	297	9,992	293
	Socio-economic status above average	8,504	267	9,233	292	8,330	234	10,171	292
	Highest socio-economic status	8,607	301	9,066	327	8,117	251	9,522	295
	Missing	290	3	165	2	562	5	249	8
Female	Overall no of patients with at least one prescription included	55,993	1,937	59,207	2,157	53,184	1,677	62,344	1,851
Age cat.	0-29	7,017	80	7,675	103	6,637	65	7,629	54
	30-49	15,409	566	16,452	603	14,651	482	16,564	546
	50-69	20,474	840	21,296	913	19,107	768	22,774	805
	70 +	13,093	451	13,784	538	12,789	362	15,377	446
Neighbourhood	*1*	11,428	383	12,003	442	10,742	336	12,652	404
SES cat.	*2*	10,891	376	11,623	432	10,504	326	12,441	336
	3	11,128	368	12,078	452	10,692	369	12,282	355
	4	11,070	420	11,612	400	10,129	317	12,733	385
	5	11,099	386	11,705	422	10,359	323	11,982	362
	Missing	377	4	186	9	758	6	254	9

3. Chronic anti-acids prescriptions

2016 2017 2018 2019

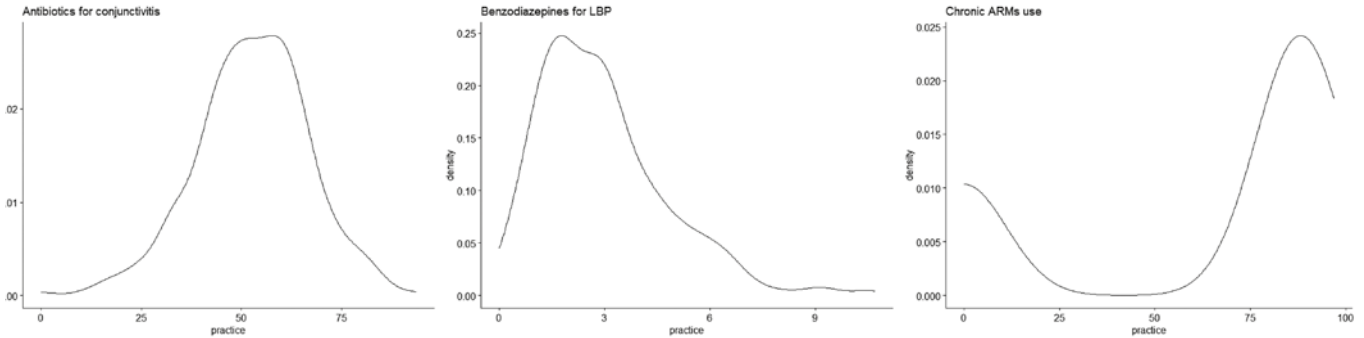
3. Chronic anti-acids prescriptions

		2016		2017		2018		2019	
	No. of practices included	284		276		250		245	
	No. of patients with chronic prescription of PPIs	100,319		105,043		93,053		91,563	
		Total no.	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**	**Total no.**	**% of total**
	No. of patients with an indication for (chronic) PPI use	12,931	12.89	11,941	11.37	11,334	12.18	11,174	12.2
	No. of patients with no indication for (chronic) PPI use	87,388	87.11	93,102	88.63	81,719	87.82	80,389	87.8
		Total	**Without (clear) indication**	**Total**	**Without (clear) indication**	**Total**	**Without (clear) indication**	**Total**	**Without (clear) indication**
Male	Overall no of patients with a chronic prescription included	44,566	39,164	46,714	41,864	41,879	37,163	41,333	36,579
Age cat.	0-49	5,692	5,637	5,942	5,877	5,051	5,002	4,643	4,593
	50-59	7,789	7,613	7,899	7,733	6,985	6,826	6,619	6,476
	60 - 69	11,884	10,718	12,145	10,973	10,725	9,755	10,370	9,410
	70 - 79	11,550	9,536	12,585	10,507	11,690	9,871	12,075	10,198
	80+	7,651	5,660	8,143	6,774	7,428	5,709	7,626	5,902
Neighbourhood	*Lowest socio-economic status**	8,567	7,555	9,060	8,128	8,090	7,218	8,029	7,062
SES cat.	*Socio-economic status below average*	8,786	7,662	9,253	8,295	8,166	7,250	8,205	7,284
	Average socio-economic status	9,020	7,932	9,282	8,299	8,336	7,346	8,243	7,258
	Socio-economic status above average	8,894	7,847	8,353	8,353	8,200	7,271	8,333	7,401
	Highest socio-economic status	9,109	7,997	8,611	8,611	8,511	7,563	8,285	7,363
	Missing	190	171	178	178	576	515	238	211

Female	Overall no of patients with a chronic prescription included	55,753	48,224	58,329	51,238	51,174	44,556	50,230	43,810
Age cat.	0-49	6,767	6,646	7,135	6,993	6,039	5,908	5,630	5,520
	50-59	9,194	8,982	9,566	9,327	8,156	7,938	7,730	7,533
	60 - 69	13,406	11,974	13,662	12,171	12,072	10,840	11,674	10,480
	70 - 79	13,950	11,129	15,119	12,271	13,717	11,145	14,102	11,567
	80+	12,436	9,493	12,847	10,476	11,190	8,725	11,094	8,710
Neighbourhood	*Lowest socio-economic status**	11,430	9,789	12,111	10,528	17,065	14,817	10,315	8,927
SES cat.	*Socio-economic status below average*	11,225	9,716	11,654	10,283	10,798	9,499	10,095	8,807
	Average socio-economic status	11,224	9,748	11,491	10,103	8,551	7,322	9,797	8,574
	Socio-economic status above average	10,884	9,435	11,598	10,175	8,254	7,210	9,956	8,661
	Highest socio-economic status	10,829	9,402	11,313	10,009	5,794	5,089	9,864	8,660
	Missing	161	134	162	140	712	619	203	181

Supplementary table 2 | Density plots and Shapiro-Wilk normality test outcomes and the correlation test outcomes.

Variable	Shapiro-Wilk normality test	P-value (significance p<0.05)	Outcome
Rate of non-indicated antibiotic prescriptions for conjunctivitis	0.99369	0.159	Normally distributed
Rate of non-indicated benzodiazepine prescriptions for LBP	0.92634	4.996e-12	Not normally distributed
Rate of non-indicated chronic ARMs prescriptions	0.62566	2.2e-16	Not normally distributed



Variable	Variable	Test	Correlation coefficient	P value (significance p<0.05)
Rate of non-indicated antibiotic prescriptions for conjunctivitis	Rate of non-indicated benzodiaze-pine prescriptions for LBP	Pearson correlation	0.16770910	0.002
Rate of non-indicated antibiotic	Rate of non-indicated chronic ARMs prescriptions for conjunctivitis	Pearson correlation	-0.02785586	0.61
Rate of non-indicated benzodiaze-pine prescriptions for LBP	Rate of non-indicated chronic ARMs prescriptions	Spearman correlation	0.04286825	0.43

Supplementary table 3 | Rates of patients over 2019 receiving a low-value prescription for each of the types of pharmaceutical GP care examined for each of the practices included. N.A. values indicate that no episodes correlating to the type of low-value pharmaceutical GP care were identified at that particular practice. However, a zero indicates that we did find episodes of the type of pharmaceutical GP care examined, but none of these were considered of low-value.

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
1	70	1	1.77	1	86.63
2	60.47	2	1.88	2	84.24
3	46.48	3	0.82	3	N.A.
4	63.96	4	1.39	4	N.A.
5	56.25	5	2	5	92.4
6	52.63	6	1.95	6	83.62
7	74.07	7	5.79	7	92.16
8	60.27	8	1.08	8	N.A.
9	73.81	9	1.08	9	83.55
10	47.37	10	2.96	10	N.A.
11	78.46	11	2.94	11	N.A.
12	72.5	12	0.61	12	N.A.
13	60	13	1.94	13	85
14	61.02	14	3.06	14	N.A.
15	66.67	15	5.56	15	93.87
16	30.43	16	5.1	16	83.33
17	68.75	17	3.44	17	82.61
18	61.36	18	2.5	18	84.38
19	56.14	19	1.21	19	N.A.
20	39.02	20	1.68	20	N.A.
21	64.29	21	2.74	21	91.45
22	24.32	22	1.75	22	N.A.
23	50	23	2.88	23	87.78
24	57.38	24	1.89	24	86.36
25	51.52	25	1.79	25	89.83
26	50	26	3.08	26	N.A.
27	32.56	27	0.74	27	N.A.
28	36.84	28	5.43	28	N.A.
29	32.56	29	4.29	29	90.28
30	31.71	30	1.54	30	88.27

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
31	66.67	31	1.21	31	N.A.
32	54.55	32	5.74	32	84.77
33	55.07	33	1.55	33	91.34
34	32.86	34	2.31	34	91.14
35	16	35	1.54	35	87.37
36	46.81	36	6.43	36	N.A.
37	54.17	37	1.12	37	90.28
38	37.04	38	2.72	38	93.2
39	58.33	39	3.43	39	96.84
40	65.62	40	0.92	40	83.17
41	32	41	0.66	41	92.31
42	41.94	42	6.34	42	88.89
43	52	43	5.7	43	90.88
44	60.47	44	4.06	44	N.A.
45	58.7	45	9.12	45	N.A.
46	61.04	46	2.87	46	92.23
47	50	47	5.65	47	N.A.
48	70	48	1.09	48	84.66
49	63.64	49	1.47	49	N.A.
50	61.22	50	1.36	50	87.15
51	72.73	51	3.33	51	81
52	63.79	52	1.15	52	88.26
53	58.93	53	4.22	53	N.A.
54	51.52	54	1.09	54	92.47
55	50.27	55	2.39	55	88.12
56	54.88	56	4.09	56	85.78
57	53.33	57	2.86	57	N.A.
58	47.06	58	1.1	58	N.A.
59	65.62	59	0.88	59	89.76
60	46.88	60	1.59	60	85.33
61	48.84	61	2.92	61	N.A.
62	23.68	62	0.82	62	N.A.
63	55.81	63	1.28	63	N.A.
64	60.61	64	4.73	64	87.86
65	43.4	65	1.64	65	87.19

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
66	15.62	66	0	66	N.A.
67	52.63	67	1.23	67	N.A.
68	57.81	68	2.88	68	83.66
69	33.33	69	7.74	69	N.A.
70	55.26	70	3.03	70	N.A.
71	34.09	71	2.26	71	88.1
72	37.04	72	0	72	88.35
73	69.23	73	2.18	73	85.54
74	56	74	2.9	74	88.57
75	40.26	75	1.64	75	N.A.
76	73.77	76	1.96	76	N.A.
77	53.33	77	1.38	77	86.07
78	50	78	2.56	78	84.44
79	54.35	79	6.23	79	90.14
80	59.68	80	4.18	80	86.48
81	57.14	81	3.88	81	86.26
82	66.67	82	4.02	82	86.82
83	60.53	83	6.43	83	93.81
84	61.7	84	1.52	84	89.65
85	62.71	85	9.13	85	87.5
86	60.87	86	2.78	86	87.17
87	38.95	87	3.45	87	89.89
88	18.75	88	0.71	88	93.26
89	62.79	89	2.96	89	90.34
90	61.84	90	5.41	90	90.81
91	57.95	91	6.36	91	N.A.
92	31.37	92	4.26	92	85.71
93	57.81	93	6.25	93	N.A.
94	66.67	94	0.85	94	87.05
95	78.95	95	2.16	95	89.25
96	46.38	96	3.05	96	87.89
97	39.62	97	3.84	97	N.A.
98	29.03	98	1.65	98	87.34
99	48.57	99	4.6	99	86.57
100	52.38	100	1.93	100	92.95

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
101	66.67	101	2.17	101	N.A.
102	52.17	102	1.08	102	87.22
103	43.75	103	1.56	103	91.42
104	76.09	104	1.89	104	84.23
105	63.22	105	3.39	105	88.67
106	46.91	106	3.25	106	82.66
107	68.29	107	0.32	107	N.A.
108	55.93	108	1.41	108	87.63
109	48.48	109	2.29	109	85.38
110	61.54	110	0.55	110	88.68
111	78.95	111	6.43	111	N.A.
112	32.86	112	0.6	112	90.64
113	48.78	113	3.36	113	88.78
114	17.86	114	1.37	114	N.A.
115	48.75	115	2.39	115	90.54
116	70	116	5.45	116	92.25
117	35.29	117	4.03	117	88.11
118	61.74	118	3.53	118	84.99
119	60.23	119	4.52	119	88.63
120	43.86	120	3.59	120	87.38
121	43.64	121	2.97	121	N.A.
122	57.35	122	3.64	122	90.33
123	57.61	123	1.95	123	90.36
124	57.69	124	0	124	N.A.
125	42.11	125	1.74	125	N.A.
126	28.57	126	2.87	126	N.A.
127	31.71	127	3.37	127	88.69
128	62.16	128	1.92	128	88.14
129	44	129	1.45	129	N.A.
130	41.67	130	4.43	130	N.A.
131	76.6	131	2.52	131	91.32
132	61.54	132	1.65	132	93.6
133	63.08	133	4.49	133	N.A.
134	74.47	134	9.26	134	85.5
135	39.39	135	3.45	135	92.67

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
136	30.28	136	0.96	136	91.59
137	59.72	137	5.79	137	86.54
138	56.52	138	3.49	138	N.A.
139	39.47	139	2	139	84.38
140	44.44	140	0.59	140	84.08
141	56.41	141	3.29	141	88
142	71.43	142	3.81	142	91.2
143	42.62	143	2.82	143	81.03
144	41.38	144	1.69	144	83.33
145	59.55	145	3.43	145	86.96
146	55	146	3.3	146	N.A.
147	44.44	147	4.05	147	N.A.
148	50	148	3.03	148	93.53
149	47.62	149	4.13	149	87.97
150	37.5	150	3.64	150	86.76
151	32	151	1.13	151	N.A.
152	35	152	3.35	152	90.69
153	60.59	153	2.75	153	N.A.
154	50	154	1.26	154	N.A.
155	40.54	155	0	155	91.86
156	32.14	156	1.97	156	86.04
157	58.54	157	5.09	157	89.29
158	42.42	158	2.25	158	89.56
159	47.22	159	3.33	159	89.97
160	58.93	160	2.74	160	79.53
161	30.3	161	1.29	161	92.43
162	74	162	2.93	162	N.A.
163	56.72	163	2.68	163	91.22
164	51.52	164	1.15	164	N.A.
165	66.67	165	4.87	165	92.6
166	55.56	166	3.1	166	91.82
167	51.35	167	3.3	167	92.06
168	40	168	2.02	168	87.31
169	46.67	169	2.42	169	82.26
170	11.11	170	7.25	170	83.76

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
171	52.38	171	10.34	171	82.27
172	65	172	4.42	172	85.35
173	47.06	173	3	173	88.27
174	64.71	174	5.85	174	88.24
175	39.53	175	1.48	175	N.A.
176	53.33	176	2.16	176	N.A.
177	41.67	177	1.65	177	N.A.
178	42.22	178	1.52	178	86.41
179	24.32	179	2.18	179	82.52
180	61.33	180	2.61	180	89.37
181	84.62	181	1.94	181	85.89
182	81.48	182	3.08	182	89.24
183	59.62	183	6.13	183	90.89
184	50.39	184	6.33	184	N.A.
185	44.26	185	3.13	185	88.15
186	50	186	1.12	186	90.57
187	23.08	187	2.38	187	89.47
188	44.44	188	5.88	188	93.43
189	51.85	189	4.94	189	88.49
190	51.85	190	1.13	190	87
191	44	191	0	191	92.31
192	47.44	192	1.87	192	86.1
193	68.97	193	5.22	193	N.A.
194	50.94	194	2.78	194	87.51
195	73.33	195	0.37	195	N.A.
196	45	196	6.72	196	88.12
197	60	197	1.25	197	87.69
198	48.48	198	2.87	198	92.65
199	46.67	199	1.96	199	N.A.
200	44.07	200	6.95	200	81.13
201	51.06	201	3.19	201	83.45
202	47.83	202	1.9	202	89.11
203	43.55	203	8.25	203	N.A.
204	55.56	204	1.86	204	90.63
205	46.15	205	2.91	205	N.A.

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
206	58.62	206	5.83	206	87.76
207	53.57	207	3.54	207	87.83
208	34.62	208	1.08	208	87.99
209	51.92	209	2.51	209	87.44
210	54.76	210	3.98	210	89.46
211	25	211	2.98	211	93.44
212	61.9	212	1.54	212	94.48
213	54.39	213	2.95	213	N.A.
214	66.67	214	4.78	214	86.92
215	78.38	215	5	215	90.79
216	23.33	216	1.82	216	95.21
217	67.53	217	2.17	217	85.62
218	53.85	218	0.85	218	91.4
219	67.44	219	1.98	219	86.39
220	35.48	220	1.51	220	88.93
221	50	221	1.2	221	N.A.
222	60.98	222	0.93	222	N.A.
223	82.35	223	2.7	223	90.24
224	71.43	224	3.41	224	86.22
225	38	225	0.99	225	81.66
226	46.58	226	2.11	226	87.7
227	55	227	5.65	227	88.55
228	48.98	228	4.48	228	N.A.
229	48.28	229	3.42	229	N.A.
230	62.5	230	0.4	230	89.71
231	40.91	231	0.9	231	83.94
232	66.07	232	5.16	232	91.08
233	55.95	233	5.3	233	88.95
234	62.5	234	1.12	234	87.9
235	49.35	235	2.66	235	N.A.
236	67.68	236	2.67	236	89.97
237	37.7	237	0.48	237	85.68
238	45.33	238	3.95	238	N.A.
239	79.03	239	4.02	239	N.A.
240	65.12	240	4.49	240	N.A.

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
241	34.55	241	2.73	241	85.87
242	80.33	242	3.07	242	N.A.
243	53.85	243	2.83	243	86.16
244	46.43	244	3.08	244	87.6
245	55.24	245	6.04	245	79.22
246	43.24	246	1.83	246	N.A.
247	26	247	1.28	247	N.A.
248	59.52	248	1.9	248	87.5
249	44	249	2.21	249	91.31
250	39.13	250	3.83	250	82.34
251	42.06	251	1.85	251	90
252	52.5	252	3.01	252	87.91
253	73.53	253	0.47	253	N.A.
254	26.32	254	4.42	254	90.72
255	64.15	255	2.13	255	N.A.
256	54.08	256	1.18	256	86.38
257	0	257	1.11	257	89.22
258	56	258	2.51	258	87.43
259	40.74	259	4.78	259	N.A.
260	63.64	260	6.83	260	N.A.
261	80.82	261	3.06	261	86
262	35.71	262	0.27	262	N.A.
263	44	263	1.52	263	83.75
264	42.86	264	5.26	264	88.51
265	72	265	7.14	265	87.65
266	58.14	266	2.09	266	N.A.
267	44.44	267	0.62	267	87.85
268	55.1	268	0.55	268	89.62
269	65.31	269	1.76	269	85.92
270	50.43	270	5.15	270	87.47
271	42.5	271	1.95	271	89.02
272	52.63	272	2.4	272	90.08
273	42.11	273	1.6	273	87.06
274	58.06	274	2.62	274	N.A.
275	62.5	275	2.3	275	89.27

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
276	59.52	276	2.64	276	86.6
277	43.48	277	1.58	277	89.36
278	50	278	2.17	278	N.A.
279	81.25	279	4.32	279	87.72
280	32.84	280	1.38	280	80.39
281	52	281	1.12	281	88.51
282	59.62	282	4.58	282	N.A.
283	48.78	283	1.55	283	N.A.
284	61.29	284	1.75	284	91.67
285	63.64	285	0.96	285	84.33
286	63.16	286	1.51	286	N.A.
287	72.55	287	2.56	287	87.25
288	64.1	288	3.33	288	88.46
289	62.5	289	1.8	289	94.58
290	67.19	290	0.43	290	N.A.
291	65	291	2.65	291	85.67
292	50	292	5	292	91.55
293	63.33	293	2.08	293	93.45
294	48.7	294	2.81	294	89.94
295	50	295	1.65	295	84.8
296	55.56	296	2.86	296	87.18
297	59.46	297	3.23	297	83.96
298	33.33	298	2.62	298	N.A.
299	50	299	3.52	299	86.06
300	58.62	300	3.65	300	91.09
301	50	301	1.14	301	91.61
302	37.21	302	3.9	302	80.97
303	50	303	3.54	303	N.A.
304	50	304	0.68	304	87.8
305	18.75	305	2.75	305	86.89
306	64.29	306	0.5	306	N.A.
307	85.71	307	10.68	307	86.49
308	50	308	5.3	308	N.A.
309	45	309	3.24	309	87.5
310	41.18	310	1.65	310	90.85

Practice ID	Rate of non-indicated antibiotics for conjunctivitis (%)	Practice ID	Rate of non-indicated benzodiazepine prescriptions for LBP (%)	Practice ID	Rate of non-indicated chronic ARMs use (%)
311	47.37	311	1.57	311	89.16
312	68.42	312	2.66	312	88.14
313	43.1	313	3.16	313	87.38
314	60	314	1.35	314	87.31
315	48.15	315	4.97	315	91.41
316	44.58	316	3.63	316	N.A.
317	46.27	317	2.65	317	88.89
318	46.51	318	3.07	318	87.95
319	38.75	319	1.8	319	79.57
320	80	320	1.66	320	N.A.
321	52.38	321	4.09	321	89.95
322	62.96	322	1.9	322	90.32
323	50	323	4.52	323	85.26
324	72.09	324	4.62	324	90.82
325	63.16	325	3.47	325	N.A.
326	93.55	326	6.71	326	89.69
327	41.38	327	4.49	327	90.93
328	67.24	328	1.99	328	87.77
329	64.71	329	0.84	329	N.A.
330	57.14	330	0.92	330	90.21
331	53.78	331	2.02	331	N.A.
332	55.13	332	2.66	332	88.01
333	45	333	4.1	333	89.86
334	33.33	334	1.09	334	86.89
335	85.71	335	2.19	335	N.A.
336	43.86	336	1.6	336	N.A.
337	57.89	337	2.68	337	N.A.
338	60.32	338	4.79	338	85.43
339	42.59	339	2.54	339	90.2
340	55	340	4.22	340	90.21
341	66.67	341	3.65	341	N.A.
342	58.14	342	3.93	342	N.A.
343	66.67	343	6.42	343	88.32
344	69.57	344	3.88	344	N.A.
345	43.24	345	2.32	345	91.35
346	56.67	346	2.83	346	84.25

Supplementary table 4 | Overview calculated variance inflation factors.

Recommendation	Variable	VIF	1/VIF
Do not prescribe local antibiotics for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient.	*Practice size*	1.01	0.994160
	Socio-economic status category	1.01	0.994656
	Age category	1.00	0.995460
	Gender	1.00	0.996029
Do not prescribe benzodiazepines in patients with non-specific lower back pain.	Practice size	1.00	0.996771
	Socio-economic status category	1.00	0.997390
	Age category	1.00	0.998751
	Gender	1.00	0.999346
Do not chronically prescribe or continue acid-reducing medication (ARM), without proper indication.	Practice size	1.01	0.994908
	Socio-economic status category	1.01	0.994779
	Age category	1.00	0.999174
	Gender	1.00	0.999042

Supplementary table 5 | overview of found odds ratios. The Asterisk indicates which category was used as reference category for the respective variables.

1. Do not prescribe a local antibiotic for an infectious conjunctivitis due to a banal pathogen, unless for a high-risk patient				
	Odds ratio (95% CI)	**Model without random intercept for practice**	**Model with random intercept for practice**	**P-value**
	Patient characteristic			
Gender	*Male**	1.0	1.0	-
	Female	0.96 (0.90 – 1.03)	0.96 (0.90 – 1.03)	0.230
Age-cat	*0-11**	1.0	1.0	-
	12-30	1.22 (1.09 – 1.36)	1.24 (1.11 – 1.39)	0.000
	31-50	1.44 (1.31 – 1.59)	1.47 (1.32 – 1.63)	0.000
	51-70	1.41 (1.28 – 1.55)	1.43 (1.28 – 1.59)	0.000
	71+	1.54 (1.38 – 1.72)	1.57 (1.38 – 1.79)	0.000
Neighbourhood SES-cat.	*Lowest socio-economic status**	1.0	1.0	-
	Socio-economic status below average	1.17 (1.06 – 1.29)	1.11 (0.98 – 1.26)	0.103
	Average socio-economic status	1.17 (1.06 – 1.29)	1.16 (1.02 – 1.32)	0.025
	Socio-economic status above average	0.97 (0.88 – 1.07)	0.97 (0.86 – 1.11)	0.673
	Highest socio-economic status	0.98 (0.89 – 1.09)	1.05 (0.92 – 1.19)	0.504
Practice size (no. of patients)	*Small practice: 0 – 3.646 patients**	1.0	1.0	-
	Medium practice: 3.647 – 6.774 patients	0.78 (0.72 – 0.84)	0.77 (0.67 – 0.89)	0.000
	Large practice: 6.775 – 17.645 patients	0.85 (0.79 – 0.92)	0.85 (0.72 – 1.01)	0.000
2. Do not prescribe benzodiazepines in patients with non-specific low back pain				
	Odds ratio (95% CI)	**Model without random intercept for practice**	**Model with random intercept for practice**	**P-value**
	Patient characteristic			
Gender	*Male**	1.0	1.0	-
	Female	1.05 (0.97 – 1.13)	1.05 (0.97 – 1.13)	0.199
Age-cat	*0-29**	1.0	1.0	-
	30-49	4.63 (3.72 – 5.76))	5.03 (4.01 – 6.31)	0.000
	50-69	5.11 (4.12 – 6.33)	5.75 (4.56 – 7.26)	0.000
	70+	3.51 (2.81 – 4.40)	4.03 (3.15 – 5.16)	0.000

Neighbourhood SES-cat.	*Lowest socio-economic status**	1.0	1.0	-
	Socio-economic status below average	0.87 (0.77 – 0.98)	0.89 (0.77 – 1.02)	0.098
	Average socio-economic status	0.96 (0.85 – 1.07)	0.97 (0.84 – 1.11)	0.656
	Socio-economic status above average	0.98 (0.88 – 1.10)	1.00 (0.87 – 1.14)	0.943
	Highest socio-economic status	0.98 (0.87 – 1.09)	1.02 (0.89 – 1.17)	0.802
Practice size (no. of patients)	*Small practice: 0 – 3.906 patients**	1.0	1.0	-
	Medium practice: 3.907 – 6.876 patients	0.79 (0.72 – 0.86)	0.78 (0.68 – 0.89)	0.000
	Large Practice: 6.877 – 17.645 patients	0.78 (0.71 – 0.85)	0.75 (0.64 – 0.87)	0.000

3. Do not prescribe or continue acid-reducing medication, without proper indication

	Odds ratio (95% CI)	**Model without random intercept for practice**	**Model with random intercept for practice**	**P-value**
	Patient characteristic			
Gender	*Male**	1.0	1.0	-
	Female	0.92 (0.88 – 0.96)	0.92 (0.88 – 0.96)	0.000
Age-cat	*0-49**	1.0	1.0	-
	50-59	0.65 (0.54 – 0.79)	0.62 (0.51 – 0.75)	0.000
	60-69	0.15 (0.12 – 0.17)	0.13 (0.11 – 0.16)	0.000
	70-79	0.08 (0.07 – 0.09)	0.07 (0.06 – 0.08)	0.000
	80+	0.06 (0.05 – 0.07)	0.05 (0.04 – 0.06)	0.000
Neighbourhood SES-cat.	*Lowest socio-economic status**	1.0	1.0	-
	Socio-economic status below average	1.13 (1.06 – 1.21)	1.12 (1.04 – 1.22)	0.004
	Average socio-economic status	1.11 (1.04 – 1.18)	1.12 (1.03 – 1.21)	0.008
	Socio-economic status above average	1.10 (1.03 – 1.17)	1.10 (1.01 – 1.19)	0.021
	Highest socio-economic status	1.15 (1.08 – 1.23)	1.19 (1.09 – 1.29)	0.000
Practice size (no. of patients)	*Small practice: 0 – 3.132 patients**	1.0	1.0	-
	Medium practice: 3.133 – 5.962 patients	0.94 (0.89 – 0.99)	0.97 (0.89 – 1.05)	0.499
	Large practice: 5.963 – 17.645 patients	0.92 (0.87 – 0.97)	0.91 (0.83 – 1.00)	0.043

CHAPTER 4

Trends in number of vitamin B12- and D-determinations in the Netherlands

Joris L.J.M. Müskens
Rudolf B. Kool
Simone A. van Dulmen
Martijn Maessen
Femke Atsma
Gert P. Westert

Published in The Dutch Journal of Medicine, 2023

Abstract

Aim: To gain insight into the volume of vitamin B12- and D-tests between 2015-2019.

Design: A retrospective cohort study.

Method: Using claims data from between 3.5 and 3.8 million insured individuals, the volume of vitamin B12 and D diagnostic testing was examined between 2015-2019. Both trends in the number of vitamin B12- and D-tests as well as the percentage of patients that received a test over the examined period were assessed.

Results: Between 2015 and 2019, the number of vitamin B12 tests increased by 98.1%, and the number of vitamin D tests increased by 112.0%. The percentage of patients per practice with a vitamin B12 test increased from 4.8% to 8.4%, and the percentage with a vitamin D test increased from 4.7% to 9.1% over the examined period.

Conclusion: Both the number of vitamin B12- and D-tests conducted by general practitioners significantly increased over the examined time period. Additionally, the percentage of patients with a vitamin test also show to have significantly risen. The findings of the present study emphasise the need for the provision of clear information regarding the indications for vitamin testing for both general practitioners and patients, in order to limit overdiagnosis.

Introduction

In the *Integraal Zorg Akkoord (IZA)*, appropriate care is described as the standard for the Dutch healthcare system. Appropriate care includes, but is not limited to, providing care that has been proven to be effective and reducing inappropriate, or ineffective, care. [1] An example of such inappropriate care, also referred to as low-value care, would be vitamin B12- or D-testing in patients experiencing non-specific complaints such as fatigue. The guidelines of the Dutch College of General Practitioners only describe a few indications which would justify vitamin B12- or D-testing. [2-4] For example, a vitamin B12 test is indicated in patients suffering from non-microcytic anaemia or neurological symptoms (such as paraesthesia or ataxia), while a vitamin D-test, or 25-OH vitamin D-test, is indicated for patients with osteoporosis or osteomalacia. However, although only a limited number of indications justifying a vitamin B12- or D-test are described, several valid indications for supplementing vitamin B12 or D exist. Vitamin B12 suppletion is recommended for vegetarians or vegans who do not consume meat substitutes. While vitamin D-suppletion is recommended to pregnant women, patients with a dark skin complexion, women over 50 years old and patients aged 70 years and older are all recommended to supplement. Box 1 provides and overview of the indications for vitamin B12- and D-testing and suppletion as described in the guidelines from the Dutch College of General practitioners.

Despite there only being a limited number of indications for vitamin B12- or D- testing, studies show that the number of vitamin B12- and D-tests has significantly increased over the past few years [5-8]. In addition, some studies show that a large part of these tests are likely of low-value and lack an appropriate indication [9-11]. Low-value vitamin B12- and D-testing promotes medicalization and overdiagnosis [12] and contribute to the increasing healthcare costs; in 2022, respectively, €6.69 and €8.75 are charged per vitamin B12- or D-test [13].

However, while relatively much is known about the extent of vitamin B12- and D-testing in other countries, its extent is not well-known for the Netherlands. The most recent insights into vitamin B12- and D-testing in the Netherlands come from a study conducted in 2015, which showed that over the previous decade, the number of vitamin B12-tests in the Nijmegen region had almost sextupled [14]. However, it is unknown whether this trend has continued over the last few years and whether this increase occurred throughout the entirety of the Netherlands. However, recent evidence suggest that sufficient room for improvement in vitamin B12- and D-testing among general practitioners (GPs) is present at the regional level. Hence, a regional study conducted in Utrecht and Rotterdam showed that through the provision of GP education, benchmarking information, and patient information, the number of requested vitamin tests was reduced by 20-25% [12, 15].

The aim of this study is to examine the trends in the number of vitamin B12- and D-tests ordered by Dutch general practitioners over the examined time period. The results can

contribute to raising awareness and identifying the extent of potentially inappropriate vitamin testing by general practitioners.

Box 1: Indications for vitamin B12 or D testing and supplementation extracted from the most recent guidelines of the Dutch College of General Practitioners (NHG). [2, 3]

From the current NHG-guidelines the following main recommendations are made regarding vitamin B12- and D-testing:

- **Vitamine B12-testing**: The general practitioner should consider requesting a vitamin B12-test in case of:
  - Non-macrocytic anemia
  - Neurological symptoms (in particular paraesthesia and ataxia)
  - Nutrient deficient diet and diseases that lead to reduced absorption of vitamin B12.

  Routinely requesting vitamin B12 levels tests in case for long-term use of metformin, proton pump inhibitors, cognitive disorders, and general complaints such as fatigue or muscle weakness without the presence of any other indications of vitamin B12 deficiency is not recommended.

  Suppletion: Vitamin B12 supplementation is recommended for patients who are vegetarian or vegan and do not consume meat substitutes.

- **Vitamin D-testing:** The general practitioner should consider requesting a vitamin B12-test in case of:
  - Individuals for whom it is unclear whether they are being sufficiently exposed to sunlight;
  - Individuals with osteoporosis or a (moderately) increased risk of falls who have sufficient dietary intake of calcium;
  - Individuals with symptoms that may indicate osteomalacia: diffuse bone and muscle pain, and proximal muscle weakness;

  Suppletion: the *Gezondheidsraad* recommends vitamin D-suppletion for large population groups, regardless of their vitamin D levels.
  - The NHG guideline for Fracture Prevention recommends vitamin D-suppletion for patients with osteoporosis and patients at increased risk of fractures.
  - Patient groups at risk for vitamin D deficiency: The *Gezondheidsraad* vitamin D-suppletion, regardless of vitamin D-levels, for all children under 4 years of age, residents of nursing homes, women over 50 years of age, men over 70 years of age, and individuals with dark skin complexion or who dress in clothing that covers most of their skin.

- Pregnant women: The NHG guideline for Pregnancy and Puerperium advises against testing for vitamin D-levels and suppletion with vitamin D (unless the pregnant woman already has an indication for supplementation due to another reason).

Methods

Design and database

To assess the extent of vitamin B12 and D diagnostics by Dutch general practitioners (GPs), we conducted a retrospective cohort study using claims data from Coöperatie VGZ from the period 2015-2019. The data used encompassed all VGZ claims for vitamin B12- and D-tests during the investigated period. In addition, we used pseudonymized information regarding the age and postcode-4 of the patients who received a vitamin B12- or D-test, the requesting GP, and the GP practice where the vitamin test was conducted. Patients were divided into four age categories based on indications described in the relevant Dutch College of General Practitioners (NHG) guidelines: 0-29, 30-49, 50-69, and 70. [2-4] Information on the patient's socioeconomic status (SES) was obtained from the Social and Cultural Planning Office (SCP) and dated from 2017. [16] The SES information was linked to the claim data based on postcode-4, and patients were classified into categories based on SES quintiles calculated over the entire population. Both age and SES were included in our analysis because previous research has shown that these patient characteristics could affect the amount of care that patients require, use or need. [17-19] VGZ has a 24% market share of the entire Dutch insured population, spread across the Netherlands, and is considered representative of the total Dutch insured population. [20] Analyses and data visualization were performed in R V3.6.3. [21] The Research Ethics Committee (CMO) of the Radboud University Medical Center evaluated the research and assessed it as being non-WMO research (file number; 2020-6767).

Trends in vitamin B12- and D-testing over time

To assess trends in vitamin B12- and D-testing, all distinct vitamin B12 and D tests between 2015 and 2019 among VGZ-insured individuals in the database were identified. Subsequently, the absolute numbers of vitamin B12 and D tests were aggregated by year (also for SES and age categories) to make trends in absolute numbers for each year visible. Additionally, we also assessed the proportion of patients that received a vitamin B12 or D test over the years included in our analysis.

Results

Between 2015-2019 both the number of vitamin B12- and D-tests, as well as the proportion of patients receiving a vitamin B12- and D-tests, show to have increased substantially. Specifically, the number of vitamin B12-tests increased by 98.1% and vitamin D-tests by 112.0%, while the percentage of patients who received a vitamin B12 or D test increased from 4.8% to 8.4% and 4.7% to 9.1%, respectively. Furthermore, the range in the number of tests per patient within the study population also increased during this period (as is shown in supplementary file 1). The study also found that the majority of vitamin B12 and D tests were ordered for patients in lower socioeconomic status (SES) categories. In addition, the number of vitamin B12 and D tests ordered increased with increasing age until the 50-69 age group, after which the numbers decreased (supplementary file 2 shows the distribution over the different age and SES categories). Figure and table 1 provide an overview of the number included GP practices, GPs, patients and number of vitamin B12- and D-tests over the examined period. Supplementary table 1 contains a more elaborate version of table 1.

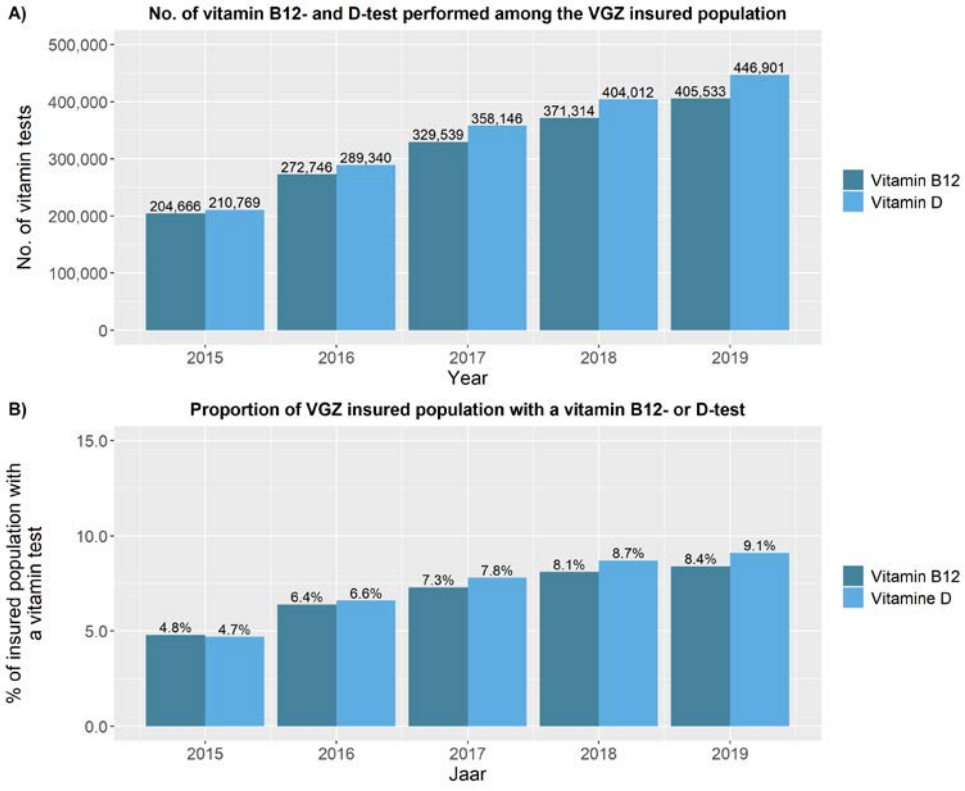


Figure 1 | Trends in number of vitamin tests and the proportion of patients with a vitamin test between 2015-2019.

Table 1 | Overview of vitamin B12- and D-tests ordered among Dutch general practitioners.

Year	2015	2016	2017	2018	2019
No. of practices	5,132	5,096	5,123	5,016	5,022
No. of general practitioners	6,849	6,794	6,776	6,686	6,630
No. of patients	3,458,853	3,446,677	3,571,435	3,624.809	3,828,135
No. Of patients whom received at least one vitamin B12-test.	166.919	219.,246	259,100	294,340	320,374
% of the total patient population that received at least one vitamin B12-test	4.8	6.4	7.3	8.1	8.4
Total no. of vitamin B12-tests	204,666	272,746	329,539	371,314	405,533
% tests in women	68.36	68.45	68.39	68.46	68.74
Average number of B12-test per patient	0.06	0.08	0.09	0.10	0.11
Average number of B12-test per general practitioner	30	40	49	56	61
Average number of B12-test per practice	40	54	64	74	81
No. Of patients whom received at least one vitamin D-test	161,817	228,207	277,949	316,276	348,917
% of the total patient population that received at least one vitamin D-test	4.7	6.6	7.8	8.7	9.1
Total no. of vitamin D-tests	210,769	289,340	358,146	404,012	446,901
% tests in women	70.22	69.60	69.54	69.18	69.60
Average number of D-test per patient	0.06	0.08	0.10	0.11	0.12
Average number of D-test per general practitioner	31	43	53	60	67
Average number of D-test per practice	41	57	70	81	89

Discussion

This study shows that the number of vitamin B12- and D-tests ordered by general practitioners has increased significantly over recent years. Between 2015 and 2019, the total number of vitamin B12- and D-test increased by 98.1% and 112.0%, respectively. The percentage of patients for whom the general practitioner requested a vitamin B12- or D-test increased from 4.8% to 8.4% for vitamin B12 and from 4.7% to 9.1% for vitamin D. The number of vitamin assays shows to increase over time, as well as the number of patients with an vitamin test (regardless of age and SES categories). The range of number of repeated tests per patient per year also shows to increase. These observations indicate that vitamin B12 and D assays are still being increasingly requested by general practitioners.

Our findings show that the observed increase in vitamin B12- and D-testing in the first half of the 2010s has continued in the following years in the Netherlands. (14, 22) This upward trend in the number of vitamin B12- and D-tests is not unique to the Netherlands. Studies from the United States (23, 24), Canada (9, 11), Australia (6), England (8, 25), and Italy (7) show that the number of vitamin B12- and D-tests has increased over the past years, with assessment outcomes ranging from 10% to 600%. The increase observed in our study is relatively low compared to increases reported in international studies, for which there might be several possible explanations. First, some of these studies were limited to, for example, only patients under 65 or performed their assessment in only one hospital, region, or province. In contrast, we included a large population in our measurement, which is considered to be representative of the entire Dutch population (20). Which makes it difficult to compare the results of these studies to our outcomes.

The increase in both the number and proportion of patients receiving a vitamin test may be related to the growing pressure patients exert on physicians to perform such tests. This is partly due to the large amount of (often) misleading information about the benefits of vitamin B12 and D testing that is spread through the media and among patients. (26,27) Providing good, reliable patient information about the indications for vitamin testing (as shown in Box 1) and updating physicians' knowledge and the provision of feedback regarding their ordering behaviour can help reduce the number of vitamin tests performed. (12, 27,28) For example, in the REVERT study, offering feedback and physician education reduced the total number of vitamin B12 and D tests by 20% and 23%, respectively. Moreover, the same study showed that the provision of patient information resulted in an additional reduction of 4% and 10% of the number of vitamin B12- and D-tests, respectively.

The available and used data unfortunately was prone to some limitations. First, the database contained insufficient clinical information to assess the appropriateness of the included vitamin tests. Data with information regarding the diagnosis codes associated to the tests or test results are required in order to be able to determine appropriateness.

This study is therefore limited to describing the number of tests performed from 2015 to 2019, without being able to draw any conclusions regarding their appropriateness. Second, we cannot say with certainty that our findings at the practice level are caused by the ordering behaviour of general practitioners. Because, although the entire database can be considered representative of the Netherlands, the level of coverage by VGZ may vary among the patient populations of the GP practices included. Both the total population size of each practice and the proportion of each practice population included in our dataset are unknown. It is therefore also unknown whether or not the proportion of the practice population that is present in the data could be considered as representative for the total practice population. For these reasons, we are unable to reliably assess the proportion of patients receiving a vitamin test on the practice level, and compare the proportions of patients receiving one between practices. However, it is unlikely that the observed increases in testing can be attributed to changes in the insured population over the course of the examined period. While recent analyses of market shares of Dutch health insurers show that the market share of VGZ remained relatively stable during the study period. (20, 29) Future research conducted with national data (such as VEKTIS data) or more complete data per practice (NIVEL) is needed to identify differences in vitamin testing among practices and identify target areas for improvement. Additionally, follow-up research using data containing information regarding test results or associated diagnoses (such as the PHARMO database or the database of the Julius General Practitioner Network) could also be used to explore the appropriateness of vitamin diagnostics.

Follow-up – National Campaign

To encourage the appropriate use of vitamin B12- and D-testing among general practitioners a nationwide initiative coordinated by the program 'To do or not to do?' has been started. Various materials for both patients and general practitioners are available which encourage appropriate use of vitamin diagnostics (https://doenoflaten.nl/vitamine/). In addition, a free e-learning is available for all general practice staff, which is also accredited for general practitioners. The e-learning provides information on the causes and diagnosis of vitamin B12 and D deficiencies, among other topics.

Conclusion

The number of vitamin B12- and D-tests as well as the proportion of patients receiving either test have increased significantly over the course of recent years. Further research should focus on generating insights into the volume and appropriateness of vitamin B12 and D diagnostics on a national level, differences between practices, and opportunities for further de-implementation. Additionally, more focus should be applied to the evaluation of de-implementation strategies, and the scale-up of successful (local) de-implementation strategies to the rest of the country.

References

1. ActiZ DNg, Federatie Medisch Specialisten, InEen, Nederlandse Federatie van Universitair Medische Centra, Nederlandse Vereniging van Ziekenhuizen, Nederlandse Zorgautoriteit, Patiëntenfederatie Nederland, Vereniging van Nederlandse Gemeenten, Verpleegkundigen & Verzorgenden Nederland, Zelfstandige Klinieken Nederland, Zorginstituut Nederland, Zorgthuisnl, Zorgverzekeraars Nederland, Ministerie van Volksgezondheid, Welzijn en Sport. Integraal Zorg Akkoord: Samen werken aan gezonde zorg. 2022.
2. Nederlands Huisartsen Genootschap. Laboratoriumdiagnostiek Vitamine B12-deficiëntie (LESA): NHG; 2018 [Available from: https://www.nhg.org/themas/publicaties/laboratoriumdiagnostiek-vitamine-b12-volledige-tekst.
3. Nederlands Huisartsen Genootschap. Vitamine D-deficiëntie (LESA Laboratoriumdiagnostiek): NHG; 2018 [Available from: https://www.nhg.org/themas/publicaties/vitamine-d-deficientie-samenvatting?tmp-no-mobile=1].
4. Wiersma T, Woutersen-Koch H. NHG-standpunt Diagnostiek van vitamine B12-deficiëntie. Huisarts Wet. 2014;57:472-5.
5. Essig S, Merlo C, Reich O, Trottmann M. Potentially inappropriate testing for vitamin D deficiency: a cross-sectional study in Switzerland. BMC Health Services Research. 2020;20(1):1097.
6. Bilinski K, Boyages S. Evidence of overtesting for vitamin D in Australia: an analysis of 4.5 years of Medicare Benefits Schedule (MBS) data. BMJ Open. 2013;3(6):e002955.
7. Lanzoni M, Fornili M, Felicetta I, Maiavacca R, Biganzoli E, Castaldi S. Three-year analysis of repeated laboratory tests for the markers total cholesterol, ferritin, vitamin D, vitamin B(12) , and folate, in a large research and teaching hospital in Italy. J Eval Clin Pract. 2017;23(3):654-61.
8. O'Sullivan JW, Stevens S, Hobbs FDR, Salisbury C, Little P, Goldacre B, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. Bmj. 2018;363:k4666.
9. Chami N, Simons JE, Sweetman A, Don-Wauchope AC. Rates of inappropriate laboratory test utilization in Ontario. Clin Biochem. 2017;50(15):822-7.
10. Morgen EK, Naugler C. Inappropriate repeats of six common tests in a Canadian city: a population cohort study within a laboratory informatics framework. Am J Clin Pathol. 2015;144(5):704-12.
11. Gill P, Guo M, Lau CK, Naugler C. Implementation of an educational province-wide intervention to reduce redundant vitamin B12 testing: A cross-sectional study. Clinical Biochemistry. 2020;76:1-4.
12. Van Vugt S, de Schepper E, De Wit N, Bindels P. Intervention to reduce vitamin testing in primary care: a randomized controlled trial. British Journal of General Practice. 2019;69(suppl 1):bjgp19X703061.
13. Diagnostiek voor u. Fares 2022 - 2023 2023 [Available from: https://diagnostiekvooru.nl/tarieven].

14. Franken P, Geutjes P, van den Ouweland J, Zeeman M, de Kok J. Diagnose van vitamine-B12-tekort. Huisarts en wetenschap. 2015;58(10):530-1.
15. Vugt Sv, de Schepper E, van Delft S, Zuithoff N, de Wit N, Bindels P. Reducing vitamin test ordering in primary care; the effectiveness of professional and patient oriented strategies in a Cluster randomized Intervention Study. BJGP Open. 2021:BJGPO.2021.0113.
16. Sociaal Cultureel Plan Bureau (SCP). Statusscores 2017. Netherlands: Sociaal Cultureel Plan Bureau (SCP); 2017.
17. Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
18. David JL, Kaplan HB. Gender, social roles and health care utilization. Appl Behav Sci Rev. 1995;3(1):39-64.
19. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. Front Public Health. 2020;8:231.
20. Vektis. Verzekerden in beeld. 2022.
21. R Core Team. R: A language and environment for statistical computing. v3.6.3 ed. Vienna, Austria: R foundation for Statistical Computing; 2019.

Supplementary file 1 | Elaborate table 1.

Year	2015	2016	2017	2018	2019
No. of unique practices included	5,132	5,096	5,123	5,016	5,022
No. of unique general practitioners included	6,849	6,794	6,776	6,686	6,630
No. of unique patients included	3,458,853	3,446,677	3,571,435	3,624,809	3,828,135
No. of unique patients whom received at least one vitamin B12-test,	166,919	219,246	259,100	294,340	320,374
Proportion of patients whom received at least one vitamin B12-test	4.8	6.4	7.3	8.1	8.4
No. of unique patients whom received at least one vitamin D-test,	161,817	228,207	277,949	316,276	348,917
Proportion of patients whom received at least one vitamin D-test	4.7	6.6	7.8	8.7	9.1
Vitamin B12					
Total no. of vitamin B12-tests	204,666	272,746	329,539	371,314	405,533
Male	64,751	86,051	104,155	117,104	126,754
Female	139,915	186,695	225,384	254,210	278,779
Average no. of vitamin B12-tests among patients	0,06	0,08	0,09	0,10	0,11
Range in no. of vitamin tests among patients	0 – 16	0 – 21	0 – 16	0 – 19	0 – 22
Median	0	0	0	0	0
Average no. of B12-tests among general practitioner	30	40	49	56	61
Range in no. of vitamin tests among general practitioners	0 – 734	0 – 1,223	0 – 1,407	0 – 1,594	0 – 1,640
Median	17	22	27	33	37
Average no. of B12-tests among GP practices	40	54	64	74	81
Range in no. of vitamin tests among GP practices	0 – 734	0 – 1,223	0 – 1,407	0 – 1,594	0 – 1,640
Median	22	30	36	43	49
Proportion of patients receiving at least one vitamin B12-test	4.83	6.36	7.25	8.12	8.37
Male	1.54	2.02	2.31	2.59	2.65
Female	3.29	4.34	4.94	5.53	5.72

Year	2015	2016	2017	2018	2019
No of vitamin B12-tests / patients per age category					
0-29 – No. of tests	28,749	39,237	47,854	57,060	63,051
No. of patients	23,534	31,812	38,043	45,866	50,695
30 – 49 – No. of tests	47,431	63,100	78,952	88,789	98,909
No. of patients	38,652	50,846	62,108	70,367	78,536
50 – 69 – No. of tests	65,948	88,250	107,608	119,731	130,678
No. of patients	54,999	72,475	86,500	96,699	104,801
70 + – No. of tests	62,538	82,159	95,125	105,734	112,895
No. of patients	50,058	64,557	73,062	81,959	87,071
No of vitamin B12-tests / patients per socioeconomic status (SES) category					
Lowest SES category ~ SES below average or equal to -0,78 – No. of tests	58,101	82,277	108,430	123,798	135,329
No. of patients	48,302	66,762	85,421	97,788	107,072
Second to last SES category ~ SES between -0,78 and -0,064 – No. of tests	49,649	61,539	72,153	80,521	88,056
No. of patients	40,650	50,909	58,371	66,126	72,209
Average SES category ~ SES between -0.064 and 0.38 – No. of tests	37,401	49,852	57,946	63,637	69,966
No. of patients	31,549	41,433	47,310	53,223	57,436
Second highest SES category ~ SES between 0.38 and 0.89 – No. of tests	30,268	39,596	44,278	51,287	55,011
No. of patients	25,535	32,950	36,828	42,097	45,289
Highest SES category ~ SES greater or equal to 0.89 – No. of tests	26,342	35,580	42,145	48,103	53,561
No. of patients	22,691	29,749	34,653	40,197	44,942
Missing SES – No. of tests	2,905	3,902	4,587	3,968	3,610
No. of patients	2,537	3,304	3,523	3,370	2,977

Year	2015	2016	2017	2018	2019
Vitamin D					
Total no. of vitamin D-tests	210,769	289,340	358,146	404,012	446,901
Male	62,764	87,957	109,104	124,506	135,838
Female	148,005	201,383	249,042	279,506	311,063
Average no. of vitamin D-tests among patients	0.06	0.08	0.10	0.11	0.12
Range in no. of vitamin tests among patients	0 - 10	0 - 11	0 - 20	0 - 14	0 - 15
Median	0	0	0	0	0
Average no. of D-tests among general practitioner	31	43	53	60	67
Range in no. of vitamin tests among general practitioners	0 – 995	0 – 1,404	0 – 1,600	0 – 1,655	0 – 1,681
Median	17	25	31	37	42
Average no. of D-tests among GP practices	41	57	70	81	89
Range in no. of vitamin tests among GP practices	0 – 955	0 – 1,404	0 – 1,600	0 – 1,655	0 – 1,681
Median	23	33	41	49	55
Proportion of patients receiving at least one vitamin D-test	4.68	6.62	7.78	8.73	9.11
Male	1.42	2.04	2.42	2.75	2.84
Female	3.26	4.58	5.36	5.97	6.27
No of vitamin D-tests / patients per age category					
0-29 – No. of tests	32,652	46,610	58,384	69,206	76,850
No. of patients	24,950	36,866	45,162	54,198	60,138
30 – 49 – No. of tests	53,638	73,772	92,611	103,768	117,093
No. of patients	40,222	56,923	70,404	79,381	89,618
50 – 69 – No. of tests	72,976	99,741	123,696	136,896	151,132
No. of patients	56,864	79,245	96,860	107,923	118,724
70 + – No. of tests	51,503	69,217	83,455	94,142	101,826
No. of patients	40,116	55,695	66,249	75,432	81,247

Year	2015	2016	2017	2018	2019
	Vitamin D				
No of vitamin D-tests / patients per socioeconomic status (SES) category					
Lowest SES category ~ SES below average or equal to -0,78 – No. of tests	59,804	88,033	117,155	132,391	147,083
No. of patients	47,027	69,226	90,088	102,662	113,915
Second to last SES category ~ SES between -0,78 and -0,064 – No. of tests	46,609	57,458	69,560	79,168	87,935
No. of patients	35,260	47,447	56,244	64,609	71,758
Average SES category ~ SES between -0.064 and 0.38 – No. of tests	36,168	50,401	61,975	68,682	75,963
No. of patients	29,272	41,661	50,348	57,314	62,305
Second highest SES category ~ SES between 0.38 and 0.89 – No. of tests	31,955	42,841	50,388	57,982	63,602
No. of patients	25,155	34,933	41,127	47,331	52,014
Highest SES category ~ SES greater or equal to 0.89 – No. of tests	33,293	46,588	53,866	61,301	68,378
No. of patients	27,246	38,085	44,201	50,275	56,511
Missing SES – No. of tests	2,940	4,019	5,202	4,488	3,940
No. of patients	2,437	3,310	4,035	3,749	3,318

Supplementary file 2: Distribution of the no. of vitamin B12- and D-tests over the age and SES-categories

Year	2015		2016		2017		2018		2019		
	Vitamin B12										
No. of vitamin B12-tests per age category											**% change**
	#	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	
0-29	28,749	14.0	39,237	14.4	47,854	14.5	57,060	15.4	63,051	15.5	119.3
30 - 49	47,431	23.2	63,100	23.1	78,952	24.0	88,789	23.9	98,909	24.4	108.5
50 - 69	65,948	32.2	88,250	32.4	107,608	32.7	119,731	32.2	130,678	32.2	98.2
70 +	62,538	30.6	82,159	30.1	95,125	28.9	105,734	28.5	112,895	27.8	80.5
No. of vitamin B12-tests per socio-economic status (SES) category											**% change**
Lowest SES ~ SES below average or equal to -0.78	58,101	28.4	82,277	30.2	108,430	32.9	123,798	33.3	135,329	33.4	132.9
Second to last SES ~ SES between -0.78 and -0.064	49,649	24.3	61,539	22.6	72,153	21.9	80,521	21.7	88,056	21.7	77.4
Average SES ~ SES between -0.064 and 0.38	37,401	18.3	49,852	18.3	57,946	17.6	63,637	17.1	69,966	17.3	87.1
Second highest SES ~ SES between 0.38 and 0.89	30,268	14.8	39,596	14.5	44,278	13.4	51,287	13.8	55,011	13.6	81.7
Highest SES ~ SES greater or equal to 0.89	26,342	12.9	35,580	13.0	42,145	12.8	48,103	13.0	53,561	13.2	103.3
Missing SES	2,905	1.4	3,902	1.4	4,587	1.4	3,968	1.1	3,610	0.9	24.3

CHAPTER 5

Non-indicated vitamin B12- and D-testing among Dutch hospital clinicians: a cross-sectional analysis in data registries

Joris L.J.M. Müskens
Rudolf B. Kool
Gert P. Westert
Maarten Zaal
Hein Muller
Femke Atsma
Simone A. van Dulmen
Tjerk Wiersma

Published in BMJ Open, 2024

Abstract

Objectives: to assess the extent of non-indicated vitamin B12- and D-testing among Dutch clinicians and its variation among hospitals.

Design: cross-sectional study using registration data from 2015-2019.

Participants: patients aged 18-70 years between that received a vitamin B12- or D-test.

Primary and secondary outcome measures: the proportion of non-indicated vitamin B12- and D-testing among Dutch clinicians and its variation between hospitals (n=68) over 2015 to 2019.

Results: Between 2015-2019 at least 79.0% of all vitamin B12- and 82.0% of vitamin D-tests lacked a clear indication. The number of vitamin B12-tests increased by 2.0% over the examined period, while the number of D-tests increased by 12.2%. The proportion of the unexplained variation in non-indicated vitamin B12- and D-tests that can be ascribed to differences between hospitals remained low. ICCs ranged between 0.072-0.085 and 0.081-0.096 for non-indicated vitamin-B12 and D tests respectively. The included casemix variables patient age, Socioeconomic Status and hospital size only accounted for a small part of the unexplained variation in non-indicated testing. Additionally, a significant correlation was observed in non-indicated vitamin B12- and D-testing among the included hospitals.

Conclusion: Hospital clinicians order vitamin B12- and D-tests without a clear indication on a large scale. Only a small proportion of the unexplained variation could be attributed to differences between hospitals.

Introduction

Low-value diagnostic testing is a serious problem in most health care systems [1]. Low-value care is defined as care that offers no net benefit for the patient and which can be associated with harmful outcomes and wasteful spending. [2-4] Within diagnostic testing, studies show that both vitamin B12- and D-tests are frequently ordered within medical practice [5-8]. However, there are only a few indications that justify the ordering of a vitamin B12- or D-test which are described in international guidelines. Among healthy adults, almost no indications for vitamin B12- and D-testing exist. [9, 10] International guidelines for hospital clinicians clearly state that vitamin B12- and D-tests should only be ordered in specific scenarios, such as in patients with Coeliac or Crohn disease, and not in patients with only vague complaints. [9, 11-14]. Assessments indicate that as much as 77% of vitamin B12- and 91% of vitamin D-tests could lack an indication. [7, 15-23]

Studies show that between 8% and 28% of vitamin B12-tests, and between 6% and 91% of vitamin D-tests could be non-indicated. Most of these studies have been conducted among general practitioners (GPs). [17, 19, 21, 23] Only a few studies have clearly specified that their assessment of the volume of non-indicated vitamin testing concerned hospital clinicians alone. Two studies from Italy reported that on average 17% and 6% of vitamin B12- and D-tests were potentially non-indicated. [7, 18] However, these assessments used data from a single hospital and therefore do not provide a representative overview. Furthermore, these studies based their assessment of appropriateness on a proxy such as testing interval, and not directly by examining the diagnose for which the test was requested.

In order to address non-indicated vitamin B12- and D-testing, an accurate assessment of the magnitude and variation of non-indicated testing is needed. We therefore aimed to assess both the volume and proportions of patients aged between 18 and 70 years old that potentially received non-indicated vitamin B12- and D-tests among Dutch hospital clinicians between 2015 and 2019 using clinical registration data. We also examined the hospital variation in non-indicated vitamin B12- and D-testing in order to identify opportunities for improvement. The proportion of outpatient visits receiving a non-indicated vitamin B12- or D-test among the hospitals included, and the diagnose codes often associated with vitamin B12- and D-testing over 2019 were also explored. In doing so, we aimed to gain more insight into the volume of non-indicated vitamin testing among hospital clinicians and the diagnoses underlying such testing.

Methods

Design and database

We conducted a cross-sectional study using registration data among Dutch hospital clinicians between 2015 and 2019. Data were obtained from the Dutch National Basic

Hospital Care Registration (Landelijke Basisregistratie Ziekenhuiszorg, LBZ). [24] The LBZ contains medical, financial and administrative information from all patients undergoing treatment in any Dutch hospital. All vitamin B12- and D- tests ordered by clinicians in Dutch hospitals over the examined time period were extracted; including the associated diagnosis codes, patient age, gender and Socioeconomic Status (SES) and hospital size. After consulting paediatricians and geriatricians, we limited our analysis to patients aged 18 to 70 years old. Both paediatricians and geriatricians indicated that for patients below 18 and above 70, there are many regional screenings protocols which also often vary between hospitals. Patients were assigned to one of three age categories; 18-29, 30-49, 50-70. SES scores were derived from a table containing SES-scores on the level of four-digit postal codes as published by the Dutch Institute for Social Research (SCP) in 2017. [25] Patients were also assigned to one of six SES categories, based on quintiles calculated with the SES information from all Dutch neighbourhoods. Hospital size was operationalized by assigning a hospital to one of three categories (small/medium/large) based on tertiles calculated using the amount of outpatient visits encountered over 2019.

Patient and public involvement

No patients or members of the public were directly involved in the study. Owing to the nature of this study and data privacy constraints, no patients or members of the public were involved in the study design, analysis, interpretation of data, or revision of the manuscript.

Analysis of trends in proportions of non-indicated vitamin B12- and D-testing among hospital clinicians

For the assessment of the proportion of justified indication versus non-indicated vitamin B12- and D-testing, we used a service lens, as previously described by Chalmers et al. [26] This entails that all vitamin B12- or D-tests ordered were included in our denominator, and all vitamin B12- and D-tests ordered with no indication in our numerator. For our distinction of non-indicated vitamin testing, we followed several steps. First, all recommendations regarding vitamin B12- or D-testing were extracted from the relevant guidelines. Initially, we reviewed the Dutch guidelines of hospital clinicians published by the federation of medical specialists for indications for vitamin B12- and D-testing. [10] We managed to identify little to no recommendations concerning indications regarding the use of vitamin B12- or D-tests. We therefore chose to supplement these with indications derived from Dutch general practitioners guidelines. [13, 14] Second, the ICD10 codes corresponding to the diagnoses which warrant the ordering of a vitamin B12- or D-test were collected from these recommendations. Third, the resulting list of ICD10 codes was reviewed by the involved experts to prevent missing relevant codes or diagnoses. We consulted two expert clinicians (an internal medicine physician and haematologist) in the process of generating the list of indications justifying a vitamin B12- or D-test. Fourth, after completion the list of ICD10 codes, all Clinical Classification Software (CCS) codes

associated to these ICD10 codes were extracted from the LBZ database. Subsequently, all Diagnosis-Treatment Combination (DTC) codes associated to the list of relevant CCS codes were extracted. The resulting list was, again, checked for completeness before starting with the assessment of indicated vitamin tests. This process was repeated until all researchers and clinicians agreed on the accuracy of the list of indications justifying vitamin B12- and D-testing. Supplementary file S1 lists of ICD10 and DTC codes used to determine the appropriateness of the identified vitamin B12- or D-tests. It also contains a description of how we identified and linked the identified ICD10 codes to each of the patients included. In an effort to provide potential handles for the design of interventions

Assessment of hospital variation in non-indicated vitamin B12- and D-testing

Hospital variation in non-indicated vitamin B12- and D- testing was assessed using a multilevel logistic regression analysis, with a random effect for hospital. Separate models per year were made to assess whether the variation in non-indicated vitamin B12- and D-testing was robust over time. Generalized variance inflation factors (GVIF) were calculated to test for collinearity among the included variables before multilevel analysis was conducted (supplementary file S2). Models were adjusted for the casemix variables: patient age, gender, Socioeconomic Status (SES), and hospital size. We corrected for patient age, gender and SES, as previous research showed that these affect the amount of care that patients require, receive and have access to. [27-31] We included a proxy for hospital size (e.g. the total number of outpatient visits in each year) while recent evidence shows that larger healthcare providers tend to provide more low-value care. [32] Vitamin tests conducted in patients with a missing SES score or DTC code were excluded from the analysis. Intraclass correlation coefficients (ICCs) were calculated to assess which part of the unexplained variation in non-indicated vitamin B12- and D-testing could be ascribed to differences between the included hospitals, using the method of Snijders and Bosker to assess the error variance. [33]

Correlation in non-indicated vitamin testing over 2019

Additionally, we also examined whether a correlation existed between the proportions of outpatient visits that received a non-indicated vitamin B12- or D-test over 2019 among the hospitals included. Correlations were assessed using the Pearson correlation coefficient for normally distributed variables and the Spearman correlation coefficient for non-normally distributed variables. Normality was assessed using both density plots and the Shapiro-Wilk test.

Results

Volume vitamin tests among Dutch hospital clinicians

Table 1 provides a general overview of the population characteristics of the population included in our study. Between 2015 and 2019, the number of vitamin B12- and D-tests ordered by clinicians increased by 2.0% (from 275,032 to 280,522) and 12.2% (from 300,013 to 336,736) respectively. A similar trend was also observed in the proportion of patients that received at least one vitamin B12- or D-test, increasing by 2.5% and 11.3% over the examined period. The amount of vitamin B12- and D-tests ordered among women remained almost twice as high compared to men over the entire period examined. Table 2 provides an overview of the outcomes, supplementary file S3 contains a more detailed breakdown by gender, age- and SES-groups on the patient level. The number of patients with at least one vitamin test increases rapidly with age. The included patients, with at least one vitamin B12 or D-test, showed to be more equally distributed over the SES categories. Only in the highest SES category a slight decrease in the number of patients with a vitamin determination was observed.

Table 1 | Overview of the study population characteristics of the used population over the entire period examined (2015 – 2019).

General info regarding the used population between 2015 - 2019	Number or proportion	Min	Max	Median	Interquartile range
Total number of unique patients	9,214,425	N.A.	N.A.	N.A.	N.A.
Gender (% female)	64.50%	N.A.	N.A.	N.A.	N.A.
Average number of patients among the hospitals (+SD)	126,722 (± 169,110)	37	1,244,526	126,722	119,601.8
Average number of unique patients among the hospitals (+SD)	47,561 (± 54317.5)	32	415,673	36,561	38,590
Average age of the patients included (+SD)	47.91 (± 14.60)	18	70	N.A.	N.A.
Average SES category of the patients included (+SD)	2.86 (± 1.43)	1	6	N.A.	N.A.
Average no. of outpatient visits among the hospitals included	1,721,337 (± 819,148)	470,117	4,659,172	N.A.	N.A.

Abbreviations; Socioeconomic status (SES), Standard deviation (SD).

Non-indicated testing

Between 2015 and 2019 around 78% of the vitamin B12- tests conducted among patients between age 18 and 70, with an registered DTC code, lacked an indication. In case of vitamin D-testing, around 82% of determinations had no clear indication. Although the number of vitamin tests is higher among women, no large differences in proportions of

non-indicated testing were observed between genders. In case of both age and SES, the proportion of patients with a non-indicated vitamin B12- or D-test remains relatively constant across all groups over the study period (supplementary file S3). With the proportion of non-indicated testing remaining around 80.0% for vitamin B12 and 83.5% for vitamin D across the different age and SES. Our analysis of diagnose codes that are most often associated with non-indicated vitamin B12- and D-testing revealed that tests are ordered for various reasons. Similar diagnostic codes were associated with both non-indicated vitamin B12 and D tests, including general malaise, fatigue without diagnosis, and ulcerative colitis. Supplementary file S4 contains an overview of the top 20 diagnose codes for both vitamin B12- and D-tests.

Hospital variation in non-indicated vitamin B12- and D-tests among Dutch hospital clinicians

The ICCs of the models uncorrected for casemix remained around 9% (ranging from 8,3% to 9,5%) and 9,5% (ranging from 8,5% to 10,1%) for the vitamin B12- and D-models over time. The ICCs of the casemix corrected vitamin B12- and D-models remained stable around 8.0% (ranging between 7,2 and 8,5%) and 9% (ranging between 8,1% and 9.6%), respectively, throughout the examined period. Supplementary file S5 contains the ICCs of all models. Casemix correction minimally impacted the calculated ICCs in case of both the vitamin B12- and D-models. The proportion of outpatient visits receiving a non-indicated vitamin B12- or D-test over 2019 varied widely among the hospitals included, ranging from 0.0% to 27.6% for vitamin B12 and 0.02% to 34.8% for vitamin D (see figure 1 and supplementary file S6).

Correlation in proportions of non-indicated vitamin B12- and D-testing

Normality testing (and inspection of density plots) revealed that both the proportions of vitamin B12- and D-testing observed among the hospitals are non-normally distributed. The subsequent correlation analysis revealed the presence of a significant positive correlation (Rho: 0.86, $p < 0.001$) between the proportions of non-indicated vitamin B12- and D-testing among the included hospitals. Supplementary file S7 contains the density plots, normality tests results and correlation analysis outcomes for both non-indicated vitamin B12- and D-tests.

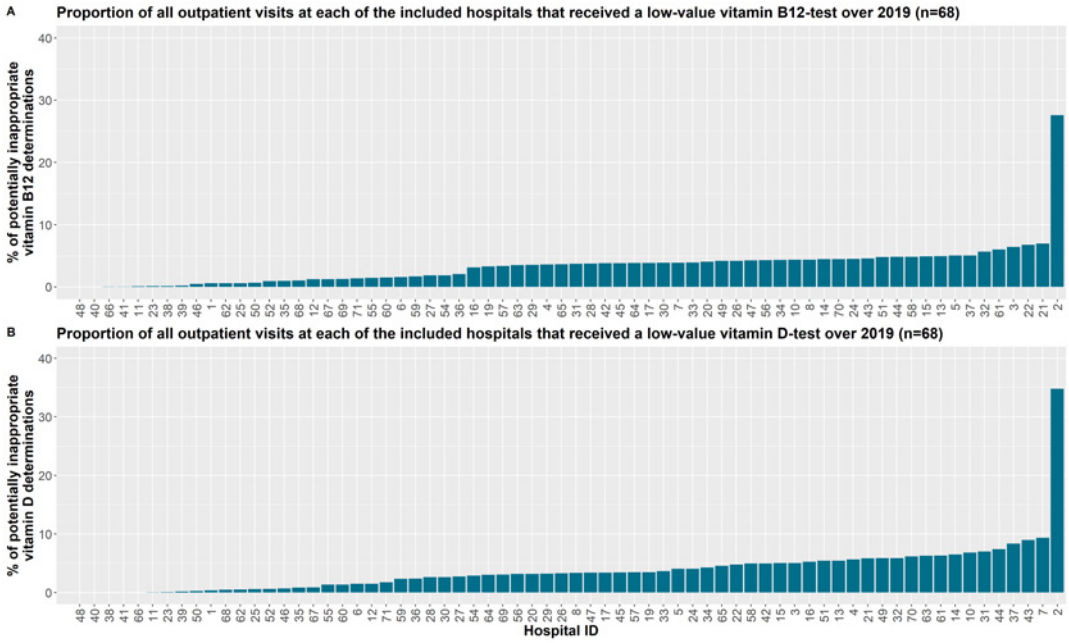


Figure 1 | Proportion of all outpatient visits at each of the included hospitals that received a low-value vitamin (**A**) B12- or (**B**) D- test over 2019 (*n*=68).

Table 2 | Overview of vitamin B12 and D tests performed among hospital clinicians over the period examined (2015 – 2019).

Year	2015	2016	2017	2018	2019
No. of hospitals included	63	64	66	69	68
Total no. B12-tests	275,032	265,546	279,524	279,957	280,522
Total no. B12-tests in hospital with a registered DTC code	118,126	163,244	188,063	208,012	213,308
% of B12-test with a registered DTC without clear indication	79.0%	78.0%	78.2%	78.3%	78.1%
Total no. of patients with (at least) one vitamin B12-test	233,541	226,999	239,033	240.063	239.351
Total no. of patients with (at least) one vitamin B12-test with a registered DTC code	103,540	141,507	162,900	178,933	183,204
Total no. of patients with (at least) one non-indicated vitamin B12-test (with a registered DTC code)	83,549	111,986	128,872	141,914	145,039
% patients that received at least one B12-test that was considered non-indicated	80.7%	79.1%	79.1%	79.3%	79.2%
No. of hospitals included	62	61	63	65	65
Total no. D tests	300,013	300,427	322,703	323,074	336,736
Total no. D tests in hospital with a registered DTC code	121,892	179,271	208,361	230,620	242,040
% of D-test with a registered DTC without clear indication	82.5%	82.9%	82.7%	82.5%	82.0%
Total no. of patients with (at least) one vitamin D-test	244,834	247,186	264,811	265,362	272,380
Total no. of patients with (at least) one vitamin D-test with a registered DTC code	101,932	148,660	172,342	189,423	196,452
Total no. of patients with (at least) one non-indicated vitamin D-test (with a registered DTC code)	84,414	124,162	143,408	157,431	162,345
% patients that received at least one B12-test that was considered non-indicated	82.8%	83.5%	83.2%	83.1%	82.6%

Abbreviations: Diagnosis Treatment Combination (DTC)

Discussion

Between 2015 and 2019, around 78.0% and 82.0% of vitamin B12- and D-tests ordered by Dutch hospital clinicians in patients aged 18-70 lacked a clear indication. The total number of vitamin B12-tests ordered increased by 2.0%, while the total number of vitamin D-tests increased by 12.2%. Although the total number of vitamin determinations increased, the proportion of patients with at least one vitamin test remained relatively constant (around 80% for B12 and 83% for D). Women received approximately twice as many vitamin B12- and D-tests, as well as non-indicated tests, over the examined period compared to men. Our analysis of hospital variation in non-indicated vitamin B12- and D-testing revealed a moderate hospital variation. Furthermore, only a relatively small part of the unexplained variation in non-indicated vitamin B12- and D- testing could be ascribed to differences between the hospitals included. Suggesting that the problem of non-indicated vitamin testing is present among all hospitals. Correlation analysis over 2019 revealed a fairly strong positive (Rho: 0.86) correlation between the rates of non-indicated vitamin B12- and D-testing among the hospitals included.

Comparison to existing literature

The proportion of non-indicated vitamin B12- and D-testing among clinicians found in our study (78.0% and 82.0% respectively) are substantially larger than the proportions reported by other studies. For example, two studies from Italy conducted among hospital clinicians report that on average 17.0% and 6.0% of vitamin B12- and D-tests could potentially be non-indicated. [7, 18] Which is substantially lower than the proportions observed in our study. Our findings are more in line with assessments that are not limited to hospital clinicians. Hence, studies conducted among general practitioners indicate that between 8.0% and 28.0% of vitamin B12- [7, 15, 16, 18] and between 7.0% and 91.0% of vitamin D-tests could be considered non-indicated. [7, 15, 19, 21, 23] Among these studies, only Naugler et al. and Gonalez-Chicas reported proportions of non-indicated vitamin D testing which is similar to ours. Naugler et al., found that following an intervention the prevalence of non-indicated vitamin D-testing decreased by 91.4%. [19] While Gonzalez-Chica reported 76,5% of vitamin D-tests to be non-indicated after introduction of new Medicare criteria for rebates. Besides population differences (e.g. GPs versus hospital clinicians), varying definitions of non-indicated testing and methods used could also account for the large differences in assessment outcomes. For example, both our study and that of Naugler et al. utilized diagnosis codes to distinguish the appropriateness of vitamin testing. [19] Most assessments of the appropriateness of vitamin testing performed to date used testing intervals or laboratory results to discern non-indicated testing, as did Gonzalez-Chica. [23] This methodological difference might explain the large differences in assessment outcomes. Furthermore, the studies that did not limit their assessment to vitamin testing among GPs, were often conducted within a single hospital. [7, 18] Moreover, most studies that assessed vitamin B12- or D-testing do not specify which type of physicians (GP/hospital clinicians) or vitamin B12- and D-tests were included, making comparison to our study challenging.

The absence of clear recommendations regarding the "appropriate" use of vitamin testing in clinician guidelines, does not aid clinicians to appropriately order vitamin testing. The current guidelines for hospital clinicians offer little direction on the appropriate use of vitamin B12 or D tests, leaving clinicians with little guidance when deciding whether to order such tests. These difficulties are magnified by the large knowledge gaps regarding the exact roles of vitamin B12 and D within the human body and its metabolism. [34-37] The combination of these factors may provide an explanation for the high proportion of non-indicated testing. Hence, due to the lack of a clear understanding of the roles of vitamin B12 and D within the body and the fear of missing a diagnosis, clinicians may engage in defensive behaviour resulting in the ordering of vitamin tests.

Strengths and limitations

A strength of this study is its novelty in examining the ordering and appropriateness of vitamin B12- and D-testing among hospital clinicians and its variation over several consecutive years. A second strength is that we used a detailed nationally representative database. This allowed us to generate an accurate and reliable overview of the extent of vitamin B12- and D-testing among solely hospital clinicians. Furthermore, it also enabled us to accurately distinguish appropriate from inappropriate vitamin testing using diagnose codes rather than the proxy of testing interval.

However, our study is also prone to limitations. First, we based our distinction of appropriateness mainly on expert opinion and indications derived from the Dutch GP guidelines, as no universal guideline regarding vitamin testing among hospital clinicians exists. We therefore might have misclassified some of the tests from clinicians as being non-indicated. However, we tried to minimize the risk of misclassification by closely collaborating with the involved experts with respect to the creation of the list of indications. Additionally, some differences exist between international and Dutch guidelines regarding indications for vitamin B12- and D-testing. For example, guidelines from the US indicate that vitamin B12-testing is considered indicated in patients with cognitive impairments or dementia. [38] However, these are not considered as an indication according to the nationwide guidelines published by the Dutch federation of medical specialists. [10] Potentially, such recommendations could exist in local protocols of Dutch hospitals, which unfortunately are not publicly available. This makes it difficult to compare international assessment outcomes to our study, while subtle differences between guidelines might cause large differences in the used criteria of appropriateness (and subsequently the reported outcomes).

Second, the use of DTC codes enabled us to accurately distinguish the appropriateness of vitamin tests, but also has a drawback. DTC diagnosis codes are generally less specific compared to the ICD10 codes described in the used guidelines. We therefore might have misclassified some vitamin tests as being appropriate. Furthermore, although DTC diagnosis codes provide a lot of insight into the diagnosis associated with practices, their

registration is prone to misregistration. [39] Clinicians have a vast amount of (often similar) DTC-codes to choose from when registering a diagnosis code, thereby adding another layer of complexity to the correct registration of diagnoses. Furthermore, DTC codes are updated as the patient passes through the healthcare system. The registered DTC code therefore does not necessarily represent the initial reason (or diagnosis) for which the vitamin determination was ordered, but rather reflects the final diagnosis.

Third, some Dutch hospitals outsource their tests to external commercial laboratories, which are not registered in the LBZ. Our estimate of vitamin testing by clinicians therefore is not complete. However, according to the registration of the national statistical office, Statistics Netherlands (CBS), 82 hospitals were active in 2019. Since we were able to include data from the majority of the hospitals in the Netherlands in our study (68/82, e.g. 83.0% of all hospitals), we do not expect to have missed much in our analysis. [40]

Implications for research and practice

Our assessment reveals that a large proportion of vitamin B12- and D- tests is ordered without a clear indication justifying their use. The total volumes of vitamin B12- and D-tests have increased over the years and show no inclination of declining. Based on publicly available fares, we estimates that roughly 3.8 million euro has been spent on non-indicated vitamin B12- and D-tests in 2019 by Dutch hospital clinicians alone. [41] This estimate of the potential savings, however, is very rough, while it only accounts for the cost price of a vitamin B12- or D-determination. The observed incidence rates of non-indicated vitamin B12- and D-testing, however, suggest that there is ample opportunity to reduce vitamin testing among Dutch clinicians. Especially since non-indicated vitamin B12- and D-tests are often ordered for similar diagnoses, and a positive correlation exists between the proportions of non-indicated testing within the same hospitals. We know that there are effective interventions to reduce inappropriate vitamin testing among GPs. A study among Dutch GPs showed that providing both education and feedback successfully reduced the amount of vitamin tests ordered with 20-25%. [42] Similar interventions might therefore also be effective among hospital clinicians to reduce (non-indicated) vitamin testing. Alternatively, more emphasis could be placed on the institution of fortification and supplementation guidelines to achieve adequate vitamin B12 and D intake among the Dutch population. Especially since, the implementation of such guidelines have shown to positively affect vitamin status among the population rendering vitamin testing obsolete in most cases. [43-45] Future research could focus on further examination of patient and physicians' characteristics associated with non-indicated vitamin B12- and D-testing. Unfortunately, information regarding the requesting physician (age, sex etc.) of non-indicated tests was not available to us in our study. Insight into physician characteristics associated with non-indicated testing could aid in the design of interventions aiming to address the problem of non-indicated vitamin testing.

Conclusion

Our research shows that the number of vitamin B12-tests slightly increased over the examined time period, while the number of vitamin D tests substantially increased among hospital clinicians. Throughout the examined period the proportion of B12- and D-tests without clear indication remained high, and are substantially higher compared to similar (international) assessments. The observed difference in assessment outcome can potentially be explained by differences in methods and definitions used to identify and define non-indicated vitamin B12- and D-tests (e.g. the use of associated diagnose codes instead of test results or proxy's such as testing interval). We also observed the presence of moderate hospital variation, but that this variation could not be explained by the included patient and hospital characteristics age, sex, SES and hospital size. Hospitals hardly differ in the task they have to undertake: bring down the number of non-indicated B12- and D-tests.

References

1. 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.
2. 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.
3. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. Med Care Res Rev. 2017;74(5):507-50.
4. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. Med J Aust. 2015;203(4):179-81.
5. Essig S, Merlo C, Reich O, Trottmann M. Potentially inappropriate testing for vitamin D deficiency: a cross-sectional study in Switzerland. BMC Health Services Research. 2020;20(1):1097.
6. Crowe FL, Jolly K, MacArthur C, Manaseki-Holland S, Gittoes N, Hewison M, et al. Trends in the incidence of testing for vitamin D deficiency in primary care in the UK: a retrospective analysis of The Health Improvement Network (THIN), 2005–2015. BMJ Open. 2019;9(6):e028355.
7. Lanzoni M, Fornili M, Felicetta I, Maiavacca R, Biganzoli E, Castaldi S. Three-year analysis of repeated laboratory tests for the markers total cholesterol, ferritin, vitamin D, vitamin B(12) , and folate, in a large research and teaching hospital in Italy. J Eval Clin Pract. 2017;23(3):654-61.
8. O'Sullivan JW, Stevens S, Hobbs FDR, Salisbury C, Little P, Goldacre B, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. Bmj. 2018;363:k4666.
9. American Society for Clinical Pathology. Thirty Five Things Physicians and Patients Should Question 2021 [Available from: https://www.choosingwisely.org/clinician-lists/ascp-do-not-order-red-blood-cell-folate-levels/].
10. Richtlijnen database [Internet]. Federatie Medisch Specialisten,. 2022 [cited 12-12-2022]. Available from: https://richtlijnendatabase.nl/.
11. Dutch College of Clinical Chemistry and Laboratory Medicine & College of medical immunologists. Verstandige keuzes bij het aanvragen en interpreteren van laboratoriumdiagnostiek. 2015.
12. Canadian Association of Pathologists. Choosingwisely Canada Pathology Recommendations,. 2021.
13. Nederlands Huisartsen Genootschap. Laboratoriumdiagnostiek Vitamine B12-deficiëntie (LESA): NHG; 2018 [Available from: https://www.nhg.org/themas/publicaties/laboratoriumdiagnostiek-vitamine-b12-volledige-tekst.
14. Nederlands Huisartsen Genootschap. Vitamine D-deficiëntie (LESA Laboratoriumdiagnostiek): NHG; 2018 [Available from: https://www.nhg.org/themas/publicaties/vitamine-d-deficientie-samenvatting?tmp-no-mobile=1].
15. Chami N, Simons JE, Sweetman A, Don-Wauchope AC. Rates of inappropriate laboratory test utilization in Ontario. Clin Biochem. 2017;50(15):822-7.

16. Morgen EK, Naugler C. Inappropriate repeats of six common tests in a Canadian city: a population cohort study within a laboratory informatics framework. Am J Clin Pathol. 2015;144(5):704-12.
17. Gill P, Guo M, Lau CK, Naugler C. Vitamin B12 test volume data before and after the implementation of an educational province-wide intervention to reduce redundant testing in Alberta. Data Brief. 2019;27:104785.
18. Tamburrano A, Vallone D, Carrozza C, Urbani A, Sanguinetti M, Nicolotti N, et al. Evaluation and cost estimation of laboratory test overuse in 43 commonly ordered parameters through a Computerized Clinical Decision Support System (CCDSS) in a large university hospital. PLoS One. 2020;15(8):e0237159.
19. Naugler C, Hemmelgarn B, Quan H, Clement F, Sajobi T, Thomas R, et al. Implementation of an intervention to reduce population-based screening for vitamin D deficiency: a cross-sectional study. CMAJ Open. 2017;5(1):E36-e9.
20. Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB. Choosing wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2015;30(2):221-8.
21. Mafi JN, Reid RO, Baseman LH, Hickey S, Totten M, Agniel D, et al. Trends in Low-Value Health Service Use and Spending in the US Medicare Fee-for-Service Program, 2014-2018. JAMA Network Open. 2021;4(2):e2037328-e.
22. Zorginstituut Nederland. Screeningsrapport Endocriene Ziekten; voedings- en stofwisselingstoornissen,. Zorginstituut Nederland; 2018.
23. Gonzalez-Chica D, Stocks N. Changes to the frequency and appropriateness of vitamin D testing after the introduction of new Medicare criteria for rebates in Australian general practice: evidence from 1.5 million patients in the NPS MedicineInsight database. BMJ Open. 2019;9(3):e024797.
24. Dutch Hospital Data. Landelijke Basisregistratie Zorg (LBZ). In: Data DH, editor.
25. The Netherlands Institute for Social Research. Statusscores 2017. Netherlands: The Netherlands Institute for Social Research (SCP); 2017.
26. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
27. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic Status and Access to Healthcare: Interrelated Drivers for Healthy Aging. Front Public Health. 2020;8:231.
28. Arpey NC, Gaglioti AH, Rosenbaum ME. How Socioeconomic Status Affects Patient Perceptions of Health Care: A Qualitative Study. J Prim Care Community Health. 2017;8(3):169-75.
29. Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
30. David JL, Kaplan HB. Gender, social roles and health care utilization. Applied Behavioral Science Review. 1995;3(1):39-64.
31. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract. 2000;49(2):147-52.

32. Segal JB, Sen AP, Glanzberg-Krainin E, Hutfless S. Factors Associated With Overuse of Health Care Within US Health Systems: A Cross-sectional Analysis of Medicare Beneficiaries From 2016 to 2018. JAMA Health Forum. 2022;3(1):e214543-e.
33. Snijders TAB, Bosker RJ. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling: SAGE Publications; 2011.
34. Holick MF. Vitamin D deficiency. N Engl J Med. 2007;357(3):266-81.
35. Green R, Miller JW. Chapter Fifteen - Vitamin B12 deficiency. In: Litwack G, editor. Vitamins and Hormones. 119: Academic Press; 2022. p. 405-39.
36. Rebelos E, Tentolouris N, Jude E. The Role of Vitamin D in Health and Disease: A Narrative Review on the Mechanisms Linking Vitamin D with Disease and the Effects of Supplementation. Drugs. 2023;83(8):665-85.
37. Halczuk K, Kaźmierczak-Barańska J, Karwowski BT, Karmańska A, Cieślak M. Vitamin B12 — Multifaceted In Vivo Functions and In Vitro Applications. Nutrients. 2023;15(12):2734.
38. Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, et al. Practice parameter: diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2001;56(9):1143-53.
39. Marcotte LM, Schuttner L, Liao JM. Measuring low-value care: learning from the US experience measuring quality. BMJ Qual Saf. 2020;29(2):154-6.
40. Statistics Netherlands (CBS). [Zorginstellingen; kerngetallen]. 2019.
41. Diagnostiek voor u. Fares 2022 - 2023 2023 [Available from: https://diagnostiekvooru.nl/tarieven].
42. Vugt SV, de Schepper E, van Delft S, Zuithoff N, de Wit N, Bindels P. Effectiveness of professional and patient-oriented strategies in reducing vitamin D and B12 test ordering in primary care: a cluster randomised intervention study. BJGP Open. 2021;5(6).
43. Itkonen ST, Andersen R, Björk AK, Brugård Konde Å, Eneroth H, Erkkola M, et al. Vitamin D status and current policies to achieve adequate vitamin D intake in the Nordic countries. Scand J Public Health. 2021;49(6):616-27.
44. Jääskeläinen T, Itkonen ST, Lundqvist A, Erkkola M, Koskela T, Lakkala K, et al. The positive impact of general vitamin D food fortification policy on vitamin D status in a representative adult Finnish population: evidence from an 11-y follow-up based on standardized 25-hydroxyvitamin D data. Am J Clin Nutr. 2017;105(6):1512-20.
45. Das JK, Salam RA, Mahmood SB, Moin A, Kumar R, Mukhtar K, et al. Food fortification with multiple micronutrients: impact on health outcomes in general population. Cochrane Database Syst Rev. 2019;12(12):Cd011400.

Supplementary materials

Supplementary file 1: Included International classification of disease (ICD10) codes and subsequent diagnose-treatment combination (DTC) codes

The table below contains the codes which were used to distinguish indicated from non-indicated vitamin B12- and D-tests. In case of vitamin D-testing, tests were considered to be appropriate in patients older than 50 with a fracture. We therefore have also provided a table of the fractures that were included.

Table 1 | Overview of the DTC codes considered an indication for a vitamin B12-determination.

Vitamin B12			
ICD-10 code	**Description**	**DTC code**	**DTC-Diagnosis description**
D51	**Vitamin B_{12} deficiency anaemia** *Excl.: Vitamin B_{12} deficiency (E53.8)*	313 – 701	- Iron-deficiency anemia (not specified)
D52	**Folate deficiency anaemia**	313 – 702	- Pernicious anemia
D53	**Other nutritional anaemias** *Incl.: megaloblastic anaemia unresponsive to vitamin B_{12} or folate therapy*	313 – 709 316 – 6001 316 – 6003 335 – 232	- Other erytrocital deviations (not specified) - Anemia, Iron-deficiency - Anemia, remaining - Anemia
E53.8	**Deficiency of other specified B group vitamins** *Deficiency:* • *Biotin* • *Cyanocobalamin* • *Folate* • *Folic acid* • *Panthothenic acid* • *Vitamin B12*	318 – 207 330 – 301	- (mal)nutrition - itamin-deficiencies
E56.8	**Deficiency of other vitamins**		
E56.9	**Vitamin deficiency, unspecified**		
G63.4	**Polyneuropathy in nutritional deficiency**	327 – 413	- Pheripheral nervedamage, nerve disorder
R20.2	**Paraesthesia of skin** *Excl.: acroparaesthesia (I73.8)*	330 – 312	- Neurological complication systemic disease
R27.0	**Ataxia, unspecified**	330 – 812 335 – 251	- Polyneuropathy, other - Diseases of the nerve system and senses
K29.4	**Chronic atrophic gastritis**	313 – 916 318 – 401	- Erosive gastritis and duodenitis - Gastritis, miscellaneous
K50	**Crohn disease [regional enteritis]** *Incl.: granulomatous enteritis* *Excl.: ulcerative colitis (K51.-)*	303 – 326 313 – 922 - 318 – 601 - 316 - 3314	- Crohn disease (enteritis regionalis) - Crohn disease - Morbus Crohn - Inflammatory bowldisease (colitis ulcerosa/Morbus crohn)
K52	**Other noninfective gastroenteritis and colitis**	313 – 932 313 – 938	Chronic diarrhoea without infection Radiation enteritis
K90	**Intestinal malabsorption** *Excl."following gastrointestinal surgery (K91.2)*	318 - 409	Coeliac disease
Z98.0	**Internal bypass and anastomosis status**	303 – 341 303 – 342	Morbid obesity BMI < 45 Morbid obesity BMI > 45

Table 2 | Overview of the DTC codes considered an indication for a vitamin D-determination.

Vitamin D			
ICD-10 code	**Description**	**DTC code**	**DTC-Diagnosis description**
E20 E21	**Hypoparathyroidism** **Hyperparathyroidism and other disorders of parathyroid gland**	- 307 – E00 - 307 – E11 - 313 – 231 - 313 – 232 - 316 – 7110 - 316 – 7199 - 362 – 304	- Emergency consult Endocrinology - Endocrinology - Hyperparathyroidism - Hypoparathyroidism - Small body length/deviant growth curve - Endocrinology, not specified - Parathyroid glands
E83.5	**Disorders of calcium metabolism**	- 313 – 239 - 313 – 299 - 316 - 7503 - 330 - 399	- Calcium metabolic disease, unspecified - Endocrine and metabolic diseases, unspecified - Metabolic disease - Deficiencies, metabolic, nutritional, unspecified
E84.9	**Cystic fibrosis, unspecified**	- 316 – 3205 - 322 - 1403	- Cystic fibrosis - Cystic fibrosis
E55	**Vitamin D deficiency** *Excl.: adult osteomalacia (M83.-), osteoporosis (M80-M81), sequelae of rickets (E64.3)*	- 305 – 1062 - 335 – 221 - 318 – 207 - 335 – 224	- Rachitis and/or vitamin D resistance osteomalacia - Endocrine and nutrition disorders (mal)nutrition - Calory/protein malnutrition
M83	**Adult osteomalacia** *Excl.: osteomalacia:* • *infantile and juvenile (E55.0)* • *vitamin-D-resistant (E83.3)* *Renal osteodystrophy (N25.0)* *Rickets (active) (E55.0)* *Rickets (active)* • *sequelae (E64.3)* *vitamin-D-resistant (E83.3)*		
E83.3	**Disorders of phosphorus metabolism and phosphates** Acid phosphatase deficiency Familial hypophosphataemia Hypophosphatasia Vitamin-D-resistant: • Osteomalacia • rickets *Excl.: adult osteomalacia (M83.-), osteoporosis (M80-M81)*	- 316 – 3328 - 316 – 7503	- Food related problems/disorders - Metabolic disease
K50	**Crohn disease [regional enteritis]** *Incl.: granulomatous enteritis* *Excl.: ulcerative colitis (K51.-)*	- 303 – 326 - 313 – 922 - 318 – 601	- Crohn disease (enteritis regionalis) - Crohn disease - Morbus Crohn
K90	**Intestinal malabsorption** *Excl."following gastrointestinal surgery (K91.2)*	- 313 – 920 - 318 – 409	- Coeliac disease/malabsorption - Coeliac disease

Vitamin D			
ICD-10 code	**Description**	**DTC code**	**DTC-Diagnosis description**
M80	**Osteoporosis with pathological fracture** *Incl.: osteoporotic vertebral collapse and wedging* *Excl.: collapsed vertebra NOS (M48.5), pathological fracture NOS (M84.4), wedging of vertebra NOS (M48.5)*	- 305 – 1395 - 316 – 5105	- Osteoperotic collapse - Pathological fractures
M81	**Osteoporosis without pathological fracture** *Excl.: osteoporosis with pathological fracture (M80.-)*	- 313 – 233	- Osteoperosis, osteomalacia
M82	**Osteoporosis in diseases classified elsewhere**		
N25.0	*Renal osteodystrophy* *Azotaemic osteodystrophy* *Phosphate-losing tubular disorders* *Renal:* • *rickets* • *short stature*	- 313 – 399 - 313 – 325	- Renal diseases, miscellaneous - Chronic renal disease (eGFR <30 ml/min)

5

Table 3 | Additional DTC codes added after checking the resulting indications.

Vitamin Determination	DTC specialism code	DTC diagnose code	Associated description	No.
D	0316	3304	Celiac disease	6509
D	0316	4006	Chronic renal failure	1245
D	0318	753	Chronic pancreatitis	3415
B12	0316	3304	Celiac disease	6233
B12	0313	772	Polycythemia vera, essential trombocytosis	5434
B12	0313	763	Myelodysplastic syndrome no specified	4843
B12	0313	920	Celiac disease/malabsorption	4206
B12	0313	773	Chronic myelogenous leukemia (CMMoL)	312

Table 4 | Bone fractures above 50 included in our analysis, that justify the ordering of a vitamin D-test.

DTC specialism code	DTC diagnose code	Fracture description.
0305	3011	Radius head
0305	3014	Carpus
0305	3029	Metatarsal bones
0303	207	Humerus proximal and shaft
0305	3024	Tibia (with or without fibula, excluding ankle)
0303	210	Radius head
0304	451	Fracture of distal radius surgical
0303	219	Femur, remaining
0303	221	Tibial plateau
0303	222	Tibia (± fibula, excluding ankle)
0305	3023	Tibial plateau
0303	217	Pelvis / sacrum
0303	214	Metacarpal bones
0305	3008	Humerus proximal and shaft
0303	224	Ankle
0305	3026	Calcaneus
0305	3030	Phalanges of the feet
0308	2301	Trauma vertebral column: conservative treatment
0303	241	Talus
0305	3020	Femur, remaining
0303	237	Tarsus
0303	215	Phalanges of the hand
0305	3027	Talus
0304	420	Fracture / luxation carpalia conservative
0305	3017	Pelvis
0305	3005	Vertebral column with paraplegia
0303	220	Patella
0303	238	Metatarsal bones
0305	3004	Vertebral column
0305	3009	Distal humerus/(epi)condyl(len)
0305	3015	Metacarpal bones
0305	3022	Fibula
0305	3025	Ankle

DTC specialism code	DTC diagnose code	Fracture description.
0303	239	Phalanges of the feet
0303	209	Olecranon
0303	213	Carpus
0305	3013	Wrist
0308	2311	Trauma vertebral column: Surgical intervention including instrumentation c.q. spinal fusion
0304	450	Distal radius facture conservative
0304	8418	Traumatic Vertebral compression fracture
0303	211	Forearm, no further specification
0305	3021	Patella
0304	421	Fracture / luxation carpalia operative
0303	212	Wrist
0303	208	Distal humerus / (epi)condyl(en)
0303	203	Vertebral column
0303	204	Vertebral column with spinal lesion
0308	2305	Treatment of trauma to the vertebral column with external fixation or traction
0305	3010	Olecranon
0305	3028	Tarsus
0303	236	Calcaneus
0305	3012	Forearm
0305	3016	Phalanges of the hand
0303	218	Fracture of neck of femur (disorder)
0305	3019	Femur proximaal (+ collum)

Supplementary file 2: Overview calculated generalized variance inflation factors (GVIF)

Table 1 | Generalized Variation Inflation Factors (GVIF) calculated for the models made using DHD data.

Year	Vitamin test	Variable	GVIF	GVIF^(1/(2*DF))
2015	Vitamin B12	*Gender*	1.038412	1.019025
		Age category	1.039533	1.009740
		Socio-economic status category	1.016783	1.002083
		Hospital size	1.016361	1.004065
	Vitamin D	*Gender*	1.013052	1.006505
		Age category	1.015735	1.003911
		Socio-economic status category	1.028674	1.003540
		Hospital size	1.028064	1.006943
2016	Vitamin B12	*Gender*	1.036084	1.017882
		Age category	1.037416	1.009225
		Socio-economic status category	1.013498	1.001677
		Hospital size	1.013126	1.003266
	Vitamin D	*Gender*	1.011964	1.005964
		Age category	1.013244	1.003295
		Socio-economic status category	1.019246	1.002386
		Hospital size	1.019576	1.004858
2017	Vitamin B12	*Gender*	1.034843	1.017272
		Age category	1.036149	1.008917
		Socio-economic status category	1.017334	1.002150
		Hospital size	1.017386	1.004319
	Vitamin D	*Gender*	1.013020	1.006489
		Age category	1.014772	1.003673
		Socio-economic status category	1.022610	1.002799
		Hospital size	1.022912	1.005679
2018	Vitamin B12	*Gender*	1.037842	1.018745
		Age category	1.039502	1.009733
		Socio-economic status category	1.024565	1.003038
		Hospital size	1.024302	1.006021
	Vitamin D	*Gender*	1.015657	1.007798
		Age category	1.017710	1.004398
		Socio-economic status category	1.029201	1.003604
		Hospital size	1.029685	1.007340

Year	Vitamin test	Variable	GVIF	GVIF^(1/(2*DF))
2019	Vitamin B12	*Gender*	1.035305	1.017499
		Age category	1.036884	1.009096
		Socio-economic status category	1.034798	1.004285
		Hospital size	1.034633	1.008548
	Vitamin D	*Gender*	1.014550	1.007249
		Age category	1.016820	1.004179
		Socio-economic status category	1.039715	1.004880
		Hospital size	1.039941	1.009839

Supplementary file 3: Extended table 1; distribution of vitamin B12- and D-testing across the sex, age and SES categories

Table 1 | Extended table 1 for vitamin B12.

1. Vitamin B12							
		2015			**2016**		
No. of hospitals included		63			64		
Total no. of unique patients with (at least) one vitamin B12 test		233.541			226.999		
Total no. of unique patients with (at least) one vitamin B12 test with an associated DBC code		103.540			141.507		
Total no. of unique patients with (at least) one non-indicated vitamin B12 test		83.549			111.986		
		Total	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**
Male	**No. of male patients receiving at least one vitamin-B12 test**	38.308	31.731	82,8	52.360	42.758	81,7
Age cat.	18-29	3.152	2.461	78,1	4.473	3.421	76,5
	30-49	9.785	7.967	81,4	13.346	10.598	79,4
	50-69	25.371	21.303	84,0	34.541	28.739	83,2
Neighbourhood	*Lowest socio-economic status*	8.070	6.693	82,9	11.346	9.271	81,7
SES cat.	*Socio-economic status below average*	8.484	7.022	82,8	11.347	9.236	81,4
	Average socio-economic status	7.790	6.472	83,1	10.372	8.472	81,7
	Socio-economic status above average	8.067	6.698	83,0	10.481	8.576	81,8
	Highest socio-economic status	5.639	4.625	82,0	8.426	6.860	81,4
	Missing	258	221	85,7	388	343	88,4
Female	**No. of female patients receiving at least one vitamin-B12 test**	65.232	51.818	79,4	89.147	69.228	77,7
Age cat.	18-29	9.138	7.155	78,3	12.781	9.751	76,3
	30-49	24.905	18.991	76,3	34.030	25.313	74,4
	50-69	31.189	25.672	82,3	42.336	34.164	80,7
Neighbourhood	*Lowest socio-economic status*	14.351	11.223	78,2	19.935	15.267	76,6
SES cat.	*Socio-economic status below average*	14.520	11.513	79,3	19.865	15.328	77,2
	Average socio-economic status	13.035	10.453	80,2	17.656	13.737	77,8
	Socio-economic status above average	13.122	10.459	79,7	17.006	13.245	77,9
	Highest socio-economic status	9.850	7.871	79,9	14.226	11.277	79,3
	Missing	354	299	84,5	459	374	81,5

	2017			2018			2019	
	66			69			68	
	239.033			240.063			239.351	
	162.900			178.933			183.204	
	128.872			141.914			145.039	
	Total	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**
	60.511	49.222	81,3	66.724	54.567	81,8	67.869	55.148	81,3
	5.262	4.094	77,8	5.921	4.579	77,3	6.468	4.886	75,5
	15.247	12.119	79,5	16.397	13.175	80,4	16.864	13.572	80,5
	40.002	33.009	82,5	44.406	36.813	82,9	44.537	36.690	82,4
	13.512	11.082	82,0	14.632	12.008	82,1	15.099	12.346	81,8
	13.601	11.078	81,4	14.686	12.001	81,7	14.962	12.116	81,0
	11.669	9.447	81,0	13.296	10.908	82,0	13.086	10.593	80,9
	11.738	9.480	80,8	12.914	10.527	81,5	13.150	10.687	81,3
	9.564	7.766	81,2	10.666	8.665	81,2	10.985	8.915	81,2
	427	369	86,4	530	458	86,4	587	491	83,6
	102.389	79.650	77,8	112.209	87.347	77,8	115.335	89.891	77,9
	14.863	11.415	76,8	16.799	13.011	77,5	17.371	13.487	77,6
	38.672	28.996	75,0	42.186	31.649	75,0	43.276	32.826	75,9
	48.854	39.239	80,3	53.224	42.687	80,2	54.688	43.578	79,7
	23.709	18.370	77,5	25.526	19.832	77,7	26.005	20.190	77,6
	23.072	17.820	77,2	25.157	19.555	77,7	25.860	20.112	77,8
	19.669	15.213	77,3	21.936	17.000	77,5	22.334	17.366	77,8
	19.098	14.899	78,0	21.366	16.625	77,8	21.984	17.079	77,7
	16.294	12.914	79,3	17.601	13.813	78,5	18.532	14.643	79,0
	547	434	79,3	623	522	83,8	620	501	80,8

Table 2 | Extended table 1 for vitamin D.

1. Vitamin D							
		2015			**2016**		
No. of hospitals included		62			61		
Total no. of patients with (at least) one vitamin D test		244.834			247.186		
Total no. of patients with (at least) one vitamin D test with an associated DBC code		101.932			148.660		
Total no. of patients with (at least) one non-indicated vitamin D test		84.414			124.162		
		Total	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**
Male	**No. of male patients receiving at least one vitamin-D test**	31.660	26.363	83,3	47.794	40.211	84,1
Age cat.	18-29	2.689	2.220	82,6	4.201	3.426	81,6
	30-49	8.684	7.658	88,2	13.086	11.536	88,2
	50-69	20.287	16.485	81,3	30.507	25.249	82,8
Neighbourhood	*Lowest socio-economic status*	6.358	5.326	83,8	10.349	8.813	85,2
SES cat.	*Socio-economic status below average*	6.758	5.584	82,6	9.964	8.342	83,7
	Average socio-economic status	6.331	5.233	82,7	9.311	7.755	83,3
	Socio-economic status above average	6.855	5.746	83,8	9.476	7.955	83,9
	Highest socio-economic status	5.164	4.309	83,4	8.405	7.092	84,4
	Missing	194	165	85,1	289	254	87,9
Female	**No. of female patients receiving at least one vitamin-D test**	70.272	58.051	82,6	100.866	83.951	83,2
Age cat.	18-29	8.105	7.198	88,8	11.485	10.030	87,3
	30-49	25.045	22.802	91,0	35.377	32.119	90,8
	50-69	37.122	28.051	75,6	54.004	41.802	77,4
Neighbourhood	*Lowest socio-economic status*	15.024	12.710	84,6	22.104	18.873	85,4
SES cat.	*Socio-economic status below average*	15.087	12.344	81,8	21.509	17.770	82,6
	Average socio-economic status	13.987	11.419	81,6	19.561	16.032	82,0
	Socio-economic status above average	14.302	11.739	82,1	19.405	15.976	82,3
	Highest socio-economic status	11.581	9.608	83,0	17.823	14.923	83,7
	Missing	291	231	79,4	464	377	81,3

2017			2018			2019		
63			65			65		
264.811			265.362			272.380		
172.342			189.423			196.452		
143.408			157.431			162.345		
Total	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**	**Total**	**Without (clear) indication**	**% without (clear indication**
56.211	47.313	84,2	61.749	51.845	84,0	63.295	52.656	83,2
5.185	4.179	80,6	5.992	4.779	79,8	6.529	5.052	77,4
15.614	13.701	87,7	16.964	14.750	86,9	17.483	15.123	86,5
35.412	29.433	83,1	38.793	32.316	83,3	39.283	32.481	82,7
12.430	10.559	84,9	13.295	11.272	84,8	13.578	11.323	83,4
12.199	10.201	83,6	13.249	11.079	83,6	13.818	11.496	83,2
10.677	8.851	82,9	12.292	10.249	83,4	12.287	10.092	82,1
10.833	9.090	83,9	11.678	9.739	83,4	12.134	10.134	83,5
9.735	8.315	85,4	10.825	9.151	84,5	11.025	9.238	83,8
337	297	88,1	410	355	86,6	453	373	82,3
116.131	96.095	82,7	127.674	105.586	82,7	133.157	109.689	82,4
14.038	12.120	86,3	16.091	13.864	86,2	17.271	14.855	86,0
40.982	37.028	90,4	45.932	41.407	90,1	47.884	42.896	89,6
61.111	46.947	76,8	65.651	50.315	76,6	68.002	51.938	76,4
26.254	22.286	84,9	28.219	23.903	84,7	29.182	24.613	84,3
25.087	20.641	82,3	27.646	22.868	82,7	28.971	23.760	82,0
22.374	18.184	81,3	24.947	20.169	80,8	25.582	20.690	80,9
21.663	17.576	81,1	24.007	19.542	81,4	25.303	20.528	81,1
20.248	17.011	84,0	22.235	18.618	83,7	23.500	19.606	83,4
505	397	78,6	620	486	78,4	619	492	79,5

Table 1 | Top 20 of diagnose codes associated with non-indicated vitamin B12-tests.

No. of non-indicated vitamin B12-tests	DTC specialism code	DTC diagnose code	DTC specialism description	DTC diagnose description
31,055	313	3	Internal medicine	Analysis general malaise/fatigue without diagnosis
30,807	313	222	Internal medicine	Diabetes mellitus with secondary complications
26,591	313	283	Internal medicine	Adiposity (obesity)
25,268	318	602	Gastro-enterology	Ulcerative colitis
20,667	313	221	Internal medicine	Diabetes mellitus without secondary complications
20,118	307	Z41	Obstetrics and gynaecology	Miscellaneous counselling during pregnancy
18,747	318	205	Gastro-enterology	Irritable bowel syndrome, diverticulosis
12,136	313	211	Internal medicine	Hypothyroidism
11,765	330	531	Neurology	Multiple sclerosis
11,168	313	6	Internal medicine	Analysis stomach complaints without diagnosis
10,003	313	223	Internal medicine	Diabetes mellitus chronic pump therapy
8,358	318	203	Gastro-enterology	Blood loss Gastrointestinal tract, unknown cause
8,057	324	701	Rheumatology	Arthralgia and/or myalgia
7,822	313	4	Internal medicine	Analysis anorexia, emaciation without diagnosis
7,770	318	204	Gastro-enterology	Chronic stomach-ache
7,425	330	402	Neurology	Miscellaneous cognitive and memory impairments
6,841	313	311	Internal medicine	Hypertension
6,758	324	101	Rheumatology	Rheumatoid arthritis
6,688	313	324	Internal medicine	Chronic kidney insufficiency eGFR 30-60 ml/min
6,339	313	299	Internal medicine	Miscellaneous endocrine and metabolic conditions

Supplementary file 4: Top 20 diagnose codes associated with non-indicated vitamin B12- and D-testing.

Table 2 | Top 20 of diagnose codes associated with non-indicated vitamin D-tests.

No. of non-indicated vitamin D-tests	DTC specialism code	DTC diagnose code	DTC specialism description	DTC diagnose description
38,213	303	342	General surgery	Morbid obesity BMI >45
34,842	313	283	Internal medicine	Adiposity (obesity)
30,175	307	Z41	Obstetrics and gynaecology	Miscellaneous counselling during pregnancy
29,857	324	701	Rheumatology	Arthralgia and/or myalgia
25,072	318	602	Gastro-enterology	Ulcerative colitis
23,263	313	3	Internal medicine	Analysis general malaise/fatigue without diagnosis
21,723	313	222	Internal medicine	Diabetes mellitus with secondary complications
21,076	324	101	Rheumatology	Rheumatoid arthritis
19,013	313	811	Internal medicine	Malignancy mama
18,346	330	531	Neurology	Multiple sclerosis
17,208	313	324	Internal medicine	Chronic kidney insufficiency eGFR 30-60 ml/min
15,050	313	461	Internal medicine	HIV infection with treatment indication
12,718	303	341	General surgery	Morbid obesity BMI <45
12,354	313	76	Internal medicine	Recipient kidney transplant
12,155	324	707	Rheumatology	Fibromyalgia
12,025	313	221	Internal medicine	Diabetes mellitus without secondary complications
11,182	313	211	Internal medicine	Hypothyroidism
10,926	313	311	Internal medicine	Hypertension
9,960	322	1201	Respiratory medicine	Asthma
8,639	318	205	Gastro-enterology	Irritable bowel syndrome, diverticulosis

Supplementary file S5: Hospital variation; overview of Intraclass correlation coefficient across years (ICC) of both the casemix adjusted and unadjusted models for vitamin B12- and D-testing included in our analysis (2015 – 2019)

Table 1 | Hospital variation; overview of Intraclass correlation coefficient across years (ICC) of both the casemix adjusted and unadjusted models for vitamin B12- and D-testing included in our analysis (2015 – 2019). Patient age, gender, socioeconomic status and an proxy for hospital size were included as casemix variables in the adjusted models.

	2015		2016		2017		2018		2019	
	Unadjusted	*Adjusted*	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*	*Unadjusted*	*Adjusted*
Vitamin-B12	0.083	0.072	0.094	0.085	0.088	0.081	0.089	0.079	0.095	0.082
Vitamin-D	0.101	0.095	0.096	0.083	0.085	0.081	0.095	0.090	0.099	0.095

Supplementary file 6: Overview of the raw data and proportions of the total outpatient visits that received a vitamin B12- and D- test for each of the included hospitals over 2019

Table 1 | Proportion (%) of outpatient visits that received a vitamin B12-tests without a clear indication.

No,	Hospital ID	Total no, of outpatient visits recorded in 2019	Total no, of vitamin B12 tests without a clear indication	Proportion (%) of outpatient visits that received a vitamin B12 test without clear indication
1	48	362,887	2	0.00
2	40	580,678	8	0.00
3	66	127,984	52	0.04
4	41	576,063	263	0.05
5	11	345,279	407	0.12
6	23	568,860	716	0.13
7	38	289,822	400	0.14
8	39	565,276	1,094	0.19
9	46	358,602	1,693	0.47
10	1	344,543	1,953	0.57
11	62	227,934	1,324	0.58
12	25	165,640	985	0.59
13	50	436,838	2,808	0.64
14	52	508,460	4,723	0.93
15	35	156,620	1,485	0.95
16	68	133,631	1,387	1.04
17	12	494,601	6,081	1.23
18	67	153,812	1,915	1.25
19	69	318,615	4,021	1.26
20	71	408,355	5,680	1.39
21	55	194,373	2,792	1.44
22	60	421,743	6,422	1.52

Continuation Table 1 | Proportion (%) of outpatient visits that received a vitamin B12-tests without a clear indication.

No,	Hospital ID	Total no, of outpatient visits recorded in 2019	Total no, of vitamin B12 tests without a clear indication	Proportion (%) of outpatient visits that received a vitamin B12 test without clear indication
23	6	384,801	6,032	1.57
24	59	395,402	6,608	1.67
25	27	389,827	7,214	1.85
26	54	383,123	7,109	1.86
27	36	363,166	7,477	2.06
28	16	163,322	5,133	3.14
29	19	131,553	4,361	3.32
30	57	280,541	9,441	3.37
31	63	427,252	15,000	3.51
32	29	216,550	7,672	3.54
33	4	315,289	11,366	3.60
34	65	237,265	8,670	3.65
35	31	309,391	11,457	3.70
36	28	113,439	4,239	3.74
37	42	268,758	10,220	3.80
38	45	384,873	14,745	3.83
39	64	342,101	13,120	3.84
40	17	155,415	6,008	3.87
41	30	370,550	14,413	3.89
42	7	384,199	14,944	3.89
43	33	606,974	23,806	3.92

44	20	147,262	5,997	4.07
45	49	259,604	10,903	4.20
46	26	182,212	7,660	4.20
47	47	185,195	7,875	4.25
48	56	192,410	8,244	4.28
49	34	429,614	18,618	4.33
50	10	276,454	12,102	4.38
51	8	223,140	9,775	4.38
52	14	567,264	25,389	4.48
53	70	253,998	11,381	4.48
54	24	654,329	29,546	4.52
55	43	536,979	24,622	4.59
56	51	443,539	21,314	4.81
57	44	356,220	17,236	4.84
58	58	154,585	7,521	4.87
59	15	165,151	8,097	4.90
60	13	466,598	23,040	4.94
61	5	401,794	20,260	5.04
62	37	171,709	8,665	5.05
63	32	337,152	19,183	5.69
64	61	174,166	10,466	6.01
65	3	113,329	7,276	6.42
66	22	251,812	17,093	6.79
67	21	186,579	13,004	6.97
68	2	379,408	104,750	27.61

Table 2 | Proportion (%) of outpatient visits that received a vitamin D tests without a clear indication.

No,	Hospital ID	Total no, of outpatient visits recorded in 2019	Total no, of vitamin D determinations without a clear indication	Proportion (%) of outpatient visits that received a vitamin D determination without clear indication
1	48	362,887	0	0.00
2	40	580,678	0	0.00
3	38	289,822	0	0.00
4	41	576,063	576,063	0.02
5	66	127,984	127,984	0.02
6	11	345,279	345,279	0.06
7	23	568,860	568,860	0.08
8	39	565,276	565,276	0.20
9	50	436,838	436,838	0.28
10	1	344,543	344,543	0.38
11	68	133,631	133,631	0.48
12	62	227,934	227,934	0.54
13	25	165,640	165,640	0.59
14	52	508,460	508,460	0.64
15	46	358,602	358,602	0.71
16	35	156,620	156,620	0.84
17	67	153,812	153,812	0.89
18	55	194,373	194,373	1.35
19	60	421,743	421,743	1.35
20	6	384,801	384,801	1.49
21	12	494,601	494,601	1.52
22	71	408,355	408,355	1.75

23	59	395,402	395,402	2.36
24	36	363,166	363,166	2.39
25	28	113,439	113,439	2.64
26	30	370,550	370,550	2.65
27	27	389,827	389,827	2.75
28	54	383,123	383,123	2.88
29	64	342,101	342,101	3.04
30	69	318,615	318,615	3.07
31	56	192,410	192,410	3.20
32	20	147,262	147,262	3.20
33	29	216,550	216,550	3.23
34	26	182,212	182,212	3.30
35	8	223,140	223,140	3.39
36	47	185,195	185,195	3.41
37	17	155,415	155,415	3.41
38	45	384,873	384,873	3.45
39	57	280,541	280,541	3.50
40	19	131,553	131,553	3.50
41	33	606,974	606,974	3.67
42	5	401,794	401,794	4.05
43	24	654,329	654,329	4.06
44	34	429,614	429,614	4.30
45	65	237,265	237,265	4.58
46	22	251,812	251,812	4.80
47	58	154,585	154,585	4.97

Continuation Table 2 | Proportion (%) of outpatient visits that received a vitamin D tests without a clear indication.

No,	Hospital ID	Total no, of outpatient visits recorded in 2019	Total no, of vitamin D determinations without a clear indication	Proportion (%) of outpatient visits that received a vitamin D determination without clear indication
48	42	268,758	268,758	5.00
49	15	165,151	165,151	5.07
50	3	113,329	113,329	5.07
51	16	163,322	163,322	5.27
52	51	443,539	443,539	5.44
53	13	466,598	466,598	5.46
54	4	315,289	315,289	5.68
55	21	186,579	186,579	5.84
56	49	259,604	259,604	5.91
57	32	337,152	337,152	5.91
58	70	253,998	253,998	6.17
59	63	427,252	427,252	6.31
60	61	174,166	174,166	6.33
61	14	567,264	567,264	6.50
62	10	276,454	276,454	6.83
63	31	309,391	309,391	7.04
64	44	356,220	356,220	7.42
65	37	171,709	171,709	8.37
66	43	536,979	536,979	9.00
67	7	384,199	384,199	9.37
68	2	379,408	379,408	34.79

Supplementary file 7: Density plots and Shapiro-Wilk normality test outcomes and the correlation test outcomes (before removal of the outliers)

Table 1 | Normality testing outcomes.

Variable	Shapiro-Wilk normality test	P-value (significance p<0.05)	Outcome
Rate of non-indicated vitamin-B12 tests	0.59	1.46e-12	Not normally distributed
Rate of non-indicated vitamin-D tests	0.58	1.19e -12	Not normally distributed

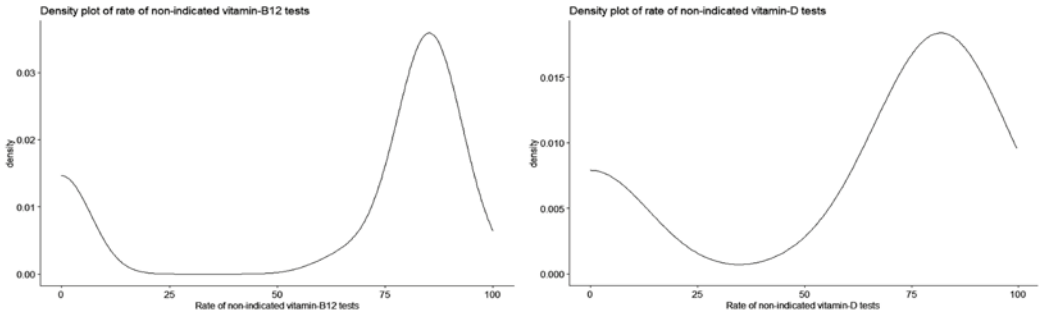


Figure 1 | Density plots of rates of vitamin B12- and D-testing among the hospitals included.

Table 2 | Spearman's rank correlation Rho test outcomes.

Rho	P-value
0.8571538	P < 2.2e-16

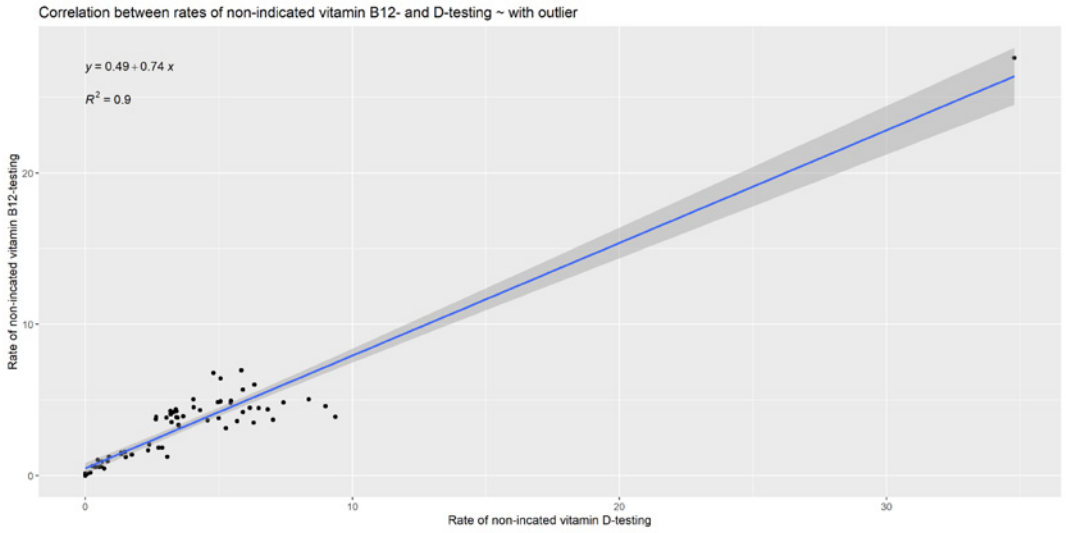


Figure 2 | Rates of non-indicated vitamin B12- and D-testing before removal of the outlier.

CHAPTER 6

Trends in low-value GP care during the COVID-19 pandemic: a retrospective cohort study

Joris L.J.M. Müskens
Tim C. Olde Hartman
Henk J. Schers
Reinier P. Akkermans
Gert P. Westert
Rudolf B. Kool
Simone A. van Dulmen

Published in BMC Primary Care, Volume 25, Article number 73. 2024

Abstract

Background: Several studies showed that during the pandemic patients have refrained from visiting their general practitioner (GP). This resulted in medical care being delayed, postponed or completely forgone. The provision of low-value care, i.e. care which offers no net benefit for the patient, also could have been affected. We therefore assessed the impact of the COVID-19 restrictions on three types of low-value GP care: 1) imaging for back or knee problems, 2) antibiotics for otitis media acuta (OMA), and 3) repeated opioid prescriptions, without a prior GP visit.

Methods: We performed a retrospective cohort study using registration data from GPs part of an academic GP network over the period 2017-2022. The COVID-19 period was defined as the period between April 2020 to December 2021. The periods before (January 2017 to April 2020) and after the COVID-19 period (January 2022 to December 2022) are the pre- and post-restrictions periods. The three clinical practices examined were selected by two practicing GPs from a top 30 of recommendations originating from the Dutch GP guidelines, based on their perceived prevalence and relevance in practice. [1] Multilevel Poisson regression models were built to examine changes in the incidence rates (IR) of both registered episodes and episodes receiving low-value treatment.

Results: During the COVID-19 restrictions period, the IRs of episodes of all three types of GP care decreased significantly. The IR of episodes of back or knee pain decreased by 12%, OMA episodes by 54% and opioid prescription rate by 13%. Only the IR of OMA episodes remained significantly lower (22%) during the post-restrictions period. The provision of low-value care also changed. The IR of imaging for back or knee pain and low-value prescription of antibiotics for OMA both decreased significantly during the COVID-restrictions period (by 21% and 78%), but only the low-value prescription rate of antibiotics for OMA remained significantly lower (by 63%) during the post-restrictions period. The IR of inappropriately repeated opioid prescriptions remained unchanged over all three periods.

Conclusions: This study shows that both the rate of episodes as well as the rate at which low-value care was provided have generally been affected by the COVID-19 restrictions. Furthermore, it shows that the magnitude of the impact of the restrictions varies depending on the type of low-value care.

This indicates that deimplementation of low-value care requires tailored (multiple) interventions and may not be achieved through a single disruption or intervention alone.

Introduction

The COVID-19 pandemic has greatly impacted healthcare. Governments introduced several social restrictions, such as lockdowns, to prevent the spread of COVID-19 and to mitigate pressure on healthcare systems. [2] Recent studies have shown that during COVID-19, patients have refrained from visiting their general practitioner (GP). [3-8] A report from the Dutch National Institute for Public Health and the Environment estimated that during first months of COVID-19 (March - June 2020), the number of GP consultations decreased by approximately 11% compared to the same period in 2019. [9] This decrease in visits has been linked to medical examinations and treatments being delayed, postponed or completely forgone. Additionally, the decrease in GP visits is only partially accounted for by an increase in telemedicine visits [10-14], indicating some patients did not receive the same care they would have received before the pandemic. However, the actual impact of these restrictions applied during the COVID-19 pandemic remains largely unknown.

The COVID-19 restrictions might have resulted in patients missing (necessary) care during the pandemic. Since the COVID-19 pandemic affected both the number of GP visits and provided care, it is also broadly hypothesized that COVID-19 could also have impacted the provision of low-value care among GPs. [15-21] Low-value care is defined as care which offers no net benefit for the patient and could be associated with harmful outcomes and wasteful spending. [22-24] The COVID-19 pandemic therefore might provide an unique opportunity to study changes in high- and low-value care provision, and where changes might be sustained or stopped. During the COVID-19 pandemic, patients could have been shielded from unnecessary or harmful medicine while they were unable to visit their GP or receive treatment. A process which is also referred to as quaternary prevention, thereby improving the quality of care these patients have received. [25, 26]. The provision of low-value care could lead to unnecessary time and costs due to additional prescriptions, laboratory tests, extra consultations and referrals. [1]

A study from the US indicated that COVID-19 reduced the amount of low-value care provided. [27] Using claims data Shahzad et al., showed that on average the amount of low-value services decreased by 56.2% during the initial month of the pandemic (April 2020), before rebounding to 83.1% of baseline by January of 2021. Unfortunately, apart from this study, knowledge regarding the impact of COVID-19 on the provision of low-value care is limited. Most studies to date have examined its impact on hospital care, Knowledge regarding its impact on (low-value) healthcare provision among GPs is lacking. We therefore studied the effect of COVID-19 on the provision of three types of low-value GP care derived from the Dutch GP guidelines in a primary care practice research network in the Netherlands, using routinely collected healthcare registration data:

1 Use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee. [28-30].
2 Prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms. [31]
3 Prescription of repeat opioid prescriptions, without a prior GP visit. [32]

Through quantification of the number and rates of both episodes, as well as episodes receiving low-value treatment, before, during and after the peak of COVID-19 to gain insight into its effect on the provision of (low-value) GP care.

Methods

Design and database

We conducted a retrospective cohort study using registration data from the database of the department of primary and community care of the Radboud university medical center. This database contains routinely collected registration data of approximately 40.000 registered patients of the GP network FaMe-net (32 GPs, six primary care practices). In FaMe-Net, all morbidity is registered in episodes of care. The title of an episode of care is the episode diagnosis, classified with the ICPC-2. The episode diagnosis can be modified during the episode of care. For example, when abdominal pain turns out to be a colon carcinoma on further diagnosis. Medication prescriptions are recorded using Anatomical Therapeutic Chemical (ATC) codes, and are linked to the relevant episode diagnosis. [33] Data collected between the 1st of January of 2017 and 31st of December 2022 were used to examine the impact of COVID-19 on both the number and rates of episodes and the provision of low-value services for three types of low-value care.

Outcome measures

The following outcome measures were used to quantify the impact of the COVID-19 pandemic on the occurrence of episodes and the provision of low-value GP care.

1 The rate of episodes or prescriptions recorded during the pre-, COVID-19 and post-restrictions period.
2 The rate of episodes or prescriptions that could be considered of low-value during the pre-, COVID-19 and post-restrictions period.

Incidence rates (IR) were calculated by dividing the total number of (low-value) episodes or prescriptions recorded by the total amount of years patients were present over each period. Thereby correcting for the time patients were able to visit the GP practice. However, apart from calculating the rates for each of these, we also first report on the raw numbers of episodes and low-value care provision recorded as supporting information.

Selection of the types of low-value GP care & cohort selection

In a previous study, a prioritization was made of "do-not-do" recommendations present in Dutch GP guidelines resulting in a top 30 of recommendations perceived as being highly relevant (through means of an online survey among 5000 GPs). [1] The resulting top 30 was presented to two authors (ToH and HS), whom are also active as GPs in clinical practice. They selected "do-not-do" recommendations based on their perceived relevance and occurrence in current daily practice. After having discussed the outcomes of their selections, the selected recommendations by both authors were clustered into three topics while multiple recommendations concerned similar topics (see Box 1). Details regarding the operationalization of the topics (e.g. the data definitions) can be found in Additional file 1.

Following operationalization of the different topics, all patients matching the diagnose codes included in our data definitions over the examined period (2017 – 2022) were extracted from the database. We did not limit ourselves to patients visiting the included practices with COVID-19 related complaints. All relevant contact were included for either of the included practices, especially since research has shown that during COVID-19 GP the way GPs were visited substantially changed. With more and more visits being conducted remotely, and less in a face-to-face manner. [34, 35] Furthermore, while within the Dutch healthcare system, all citizens are required by law to be registered at a GP and to have healthcare insurance (covering the costs of GP visits). Furthermore, patients can only gain access to (non-emergency) hospital care through referral of their GP. Guaranteeing that almost all relevant episodes of patients registered at the included practices (apart for medical emergencies) were included in the different cohorts of this study.

Box 1: short descriptions of the operationalization of the different types of low-value GP care, including from which guidelines they were derived (additional file 1 contains an elaborate description of the specific diagnose codes included):

1 | The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.
Dutch GP guidelines do not recommend to order imaging in case of non-specific knee or back pain. For our assessment we selected all episodes related to back or knee pain were selected. Next, all contacts with a code indicating they resulted in an imaging procedure were matched to each episode based on their unique episode identifier. The episodes with an associated contact indicating the performance of an imaging procedure were considered to have received low-value imaging. [28-30]

2 | The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.
Guidelines recommend not to prescribe antibiotics in case of otitis media in children without the patient being seriously ill or without them being at risk of complications. For our assessment, we selected all episodes of otitis media acuta among children (<18 years old). Next, all prescriptions of antimicrobial agents were matched to the distinct OMA episode based on episode number and prescription date. The Dutch GP guidelines only advice the prescription of an antibiotic in case of OMA when no improvement of both the present fever or pain occurs after three days of appropriate pain management. We therefore defined severe symptoms as children which had a reason for encounter for OMA of at least 72 hours.. In case a child did not have a reason for encounter of at least 72 hours, but had received a prescription for antibiotics within this time frame we marked that prescription as being of low-value. [31]

3 | Repeat opioid prescriptions, without a prior GP visit.
Guidelines advice repeat opioid prescriptions only to be prescribed following a consultation with a GP. We therefore included all opioid prescriptions over the examined period in our examination of low-value repeated opioid prescriptions in our assessment. In our analysis of appropriateness, we did not include the initial opioid prescriptions, while these simply cannot be considered repeat prescriptions. Next, the identified GP contacts were matched to each of the repeat opioid prescriptions based on their respective contact and prescription dates. These had to match in order for the repeat opioid prescription to be considered as being appropriate. Repeat prescriptions that did not have a contact associated to them were considered as being of low-value. [32]

Defining the assessment numerator and denominator: Assessment lenses

Two types of assessment lenses were used depending on the type of care examined: the patient-indication and service lens. [36] The patient-indication lens was applied in our assessment of the inappropriate use of imaging for musculoskeletal problems, and antibiotic prescriptions in case of OMA. Which implies that only patients with a certain indication were included in the denominator for these assessments, the numerator consisted of patients that received the types of low-value care for at least one episode. For our assessment of inappropriate repeat opioid prescriptions a service lens was used. Implying that all registered opioid prescriptions were included in the denominator, and all prescriptions considered to be inappropriately repeated in the numerator.

Definitions of the Pre-, COVID-19 and Post-restrictions periods

We defined the COVID-19 period as the period during which strong COVID-19 related restrictions, such as lockdowns, were imposed on the Dutch population as described on the website of the Dutch Government. [37] Resulting in the period between the 1st of April 2020 (the second quarter of 2020) and the 31st of December 2021 (the fourth quarter of 2021) to be referred to as the COVID-19 period. While the periods before (January 1st 2017 to April 2020) and after the COVID-19 period (the 1st of January 2022 up to the 31st of December 2022) are referred to as the pre- and post-restrictions periods. Figure 1 provides an overview of the timeline and restrictions. Additional file 2 presents a detailed overview of the restrictions used to define the COVID-19 period.

Data analysis

Assessing differences in incidence rates of episodes and low-value care provision

To test the differences in IRs between each of the periods, Poisson multilevel regression models were built, and checked for overdispersion. In case overdispersion was detected, negative binomial models were built to account for the over-dispersed data. Separate models were built to examine whether changes in the IR of episodes/prescriptions or the provision of low-value care could be detected between the three periods. In each of the models a fixed effects of period was included and we aimed to include random effects for both the patient and practice level when possible. Furthermore, we included an offset for the number of years a patient was present in each period, to correct for any differences in duration patients were present over the different periods. [38] Patient age and sex were included as case-mix variables in the models, while previous research has shown they could affect the amount of care a patient requires, receives or has access to. [39-41] Differences in IRs between periods were expressed as Rate Ratio (including 95% confidence intervals [95% CI]). The pre-COVID-19 period was taken as reference period.

A P-value smaller or equal to 0.05 was considered statistically significant for all analyses, based on two-sided testing. Data analysis and visualization was performed using R (version 4.1.3).

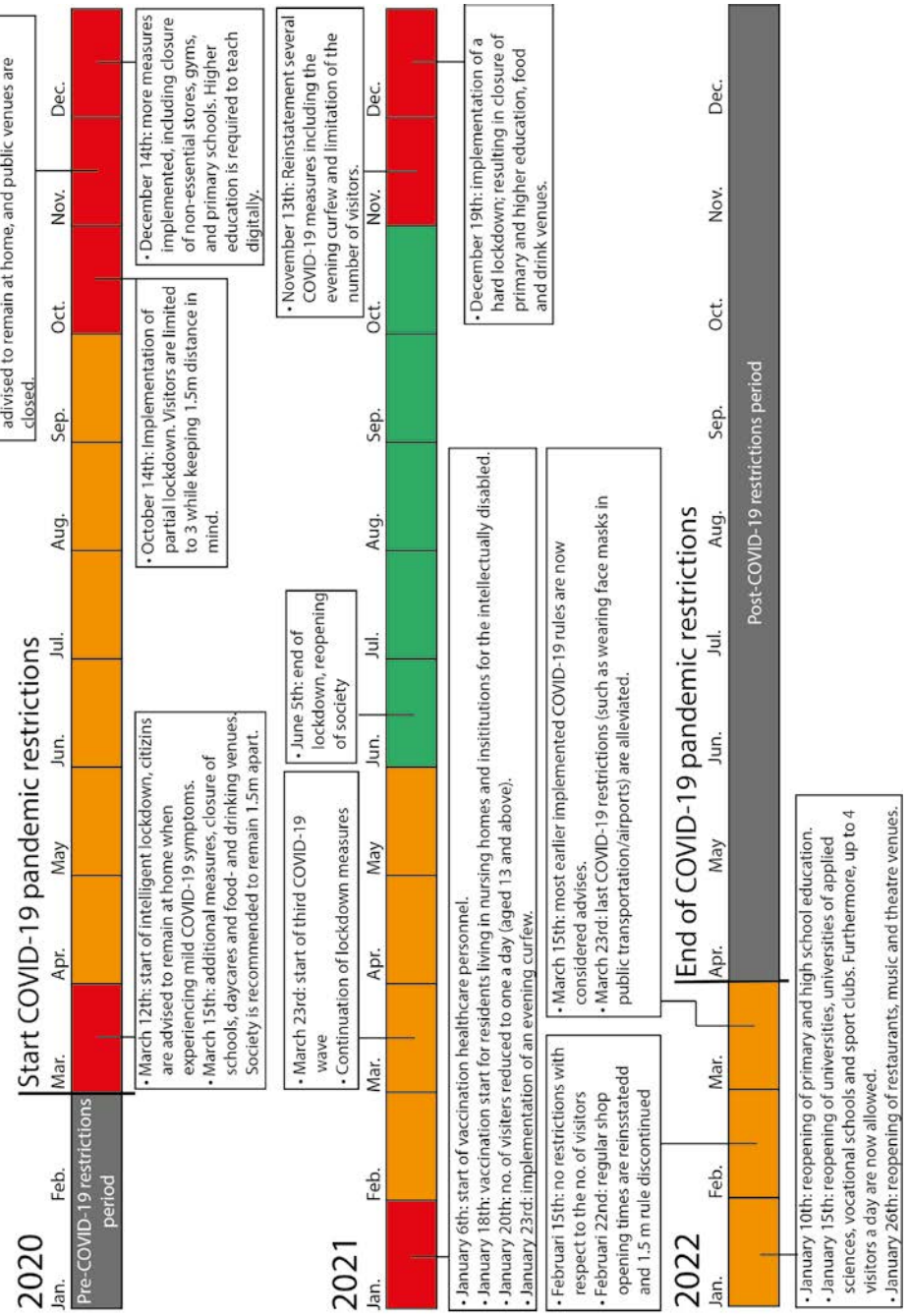


Figure 1 | Overview of the timline and restrictions implemented over the course of the COVID-19 pandemic.

Results

Trends in number of recorded episodes and recorded episodes receiving low-value care

Over the COVID-19 restrictions period, both the number of recorded episodes or prescriptions across the three types of GP care examined show distinct patterns. Table 1 contains an overview of the population characteristics of the populations used to examine each type of care. The average number of episodes and number of episodes receiving low-value care over the periods examined is shown in table 2 (additional file 3 contains an extended version of table 2). Both the number of episodes of back and knee pain and prescriptions of antibiotics for OMA sharply decrease at the onset of the COVID-19 period (plots and data concerning separate back and knee episodes are shown in additional file 4). With the average number of episodes of back and knee pain decreasing from 848.4 to 692.9, and episodes of OMA from 145.7 to 99.0.The number of opioid prescriptions also showed to slightly decrease from 988.3 to 1016.8, but already showed to slightly decrease before onset of the restrictions (as shown in additional file 3). The number of episodes of all three types of care show to gradually increase again over the course of the COVID-19 restrictions period.

Regarding low-value treatment (table 2), both the number of (low-value) imaging procedures and antibiotic prescriptions for OMA slightly decreased (from 80.5 to 62.6, and from 9.6 to 3.0 respectively. In both cases the observed decrease was reverted during the post-restrictions period. The low-value prescription of opioids was not affected by the introduction and removal of the restrictions since its low-value prescription remained high over the entire period remaining relatively high (with an average decrease from 249.2 to 217.3). However, since these raw numbers are not corrected for either exposure period nor any patient characteristics, we have performed our main analysis using the rates of provision as described below.

Table 1 | Overview of the population characteristics for the different types of care examined.

		Pre-COVID-19 restrictions period	COVID-19 restrictions period	Post-COVID-19 restrictions period
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee				
Total no. of unique patients, n		10,802	10,329	9,798
Female, n (%)		5,876 (54.4)	5,609 (54.3)	5,271 (53.8)
Age category, n (%)	*0 - 18*	1,057 (9.8)	1,049 (10.2)	1,023 (10.4)
	19 - 50	4,784 (44.3)	4,729 (45.8)	4,477 (45.7)
	50 - 70	3,387 (31.4)	3,193 (30.9)	3,078 (31.4)
	70+	1,574 (14.6)	1,358 (13.1)	1,220 (12.5)
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms				
Total no. of unique patients, n		1,684	1,875	1,823
Female, n (%)		807 (47.9)	881 (47.0)	859 (47.1)
Age category, n (%)	*0 - 1*	637 (37.8)	843 (45.0)	815 (44.7)
	1 - 5	690 (41.0)	683 (36.4)	669 (36.7)
	5 - 12	275 (16.3)	271 (14.5)	264 (14.5)
	12 - 18	82 (4.9)	78 (4.2)	75 (4.1)
3. Repeat opioid prescriptions. without a prior visit				
Total no. of unique patients, n		3,498	3,498	3,498
Female, n (%)		2,081 (59.5)	2,081 (59.5)	2,081 (59.5)
Age category, n (%)	*0 - 50*	1,161 (33.2)	1,161 (33.2)	1,161 (33.2)
	50 - 70	1,279 (36.6)	1,279 (36.6)	1,279 (36.6)
	70+	1,058 (30.2)	1,058 (30.2)	1,058 (30.2)

Table 2 | Average number of episodes and number and proportion of episodes receiving low-value care for the three types of GP care examined.

	Pre-COVID-19 restrictions period	COVID-19 restrictions period	Post-COVID-19 restrictions period
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.			
Average no. of episodes	848.4	692.9	713.5
Average no. episodes receiving low-value care	80.5	62.6	74.0
Average % of episodes receiving low-value treatment	9.5	9.0	10.4
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.			
Average no. of episodes	145.7	99.0	182.3
Average no. episodes receiving low-value care	9.6	3.0	5.8
Average % of episodes receiving low-value treatment	6.8	4.1	3.4
3. Repeat opioid prescriptions. without a prior visit			
Average no. of episodes	988.3	867.6	1,016.8
Average no. episodes receiving low-value care	249.2	217.3	244.0
Average % of episodes receiving low-value treatment	25.4	25.1	24.0

Trends in rates of episodes between the different periods

TThe IR of the episodes of the examined types of GP care all significantly decreased over the COVID-19 period (figure 2, and table 3). Both the IRs of back and knee pain and opioid prescriptions only moderately decreased by 12% ($p<0.001$) and 13% ($p<0.01$) over the restrictions period. However, these decreases did not sustain during the post-restrictions period. Both the IRs of episodes of back and knee pain and opioid prescriptions did not significantly differ from the pre-restrictions period ($p>0.05$). In case of OMA among children the IR of episodes decreased by 54% during the restrictions period ($p<0.001$). In contrast to the other two types of care, the IR of OMA episodes remained significantly (22%) lower during the post-restrictions period ($p<0.001$). The IR of OMA episodes shows a clear seasonal tendency. The rate of episodes peaks around the first quarter of the included years, apart from the first quarter of 2021.

Trends in incidence rates of low-value care between the different periods

The IRs of two out of the three types of low-value GP care significantly decreased during the COVID-19 restrictions period (figure 3 and table 3). The IR of episodes of back or knee receiving low-value imaging decreased by 21% ($p<0.001$) and the IR of OMA episodes receiving low-value antibiotics by 78% ($p<0.001$). The IR of low-value repeat opioid prescriptions also showed to have decreased by 7%, however did not significantly differ from the IR of the pre-restrictions period ($p>0.05$). During the post-restrictions period, the IR of low-value imaging for back and knee pain and low-value repeat opioid prescription both returned to pre-restrictions period levels. Conversely, the IR of low-value antibiotics prescriptions for OMA remained 63% lower during the post-restrictions period ($p<0.001$). In case of the IR of low-value antibiotic prescriptions for OMA in children we did not observe a clear seasonal trend as was the case with the rate of OMA episodes. Additional file 5 contains the IRs for the different types of care examined.

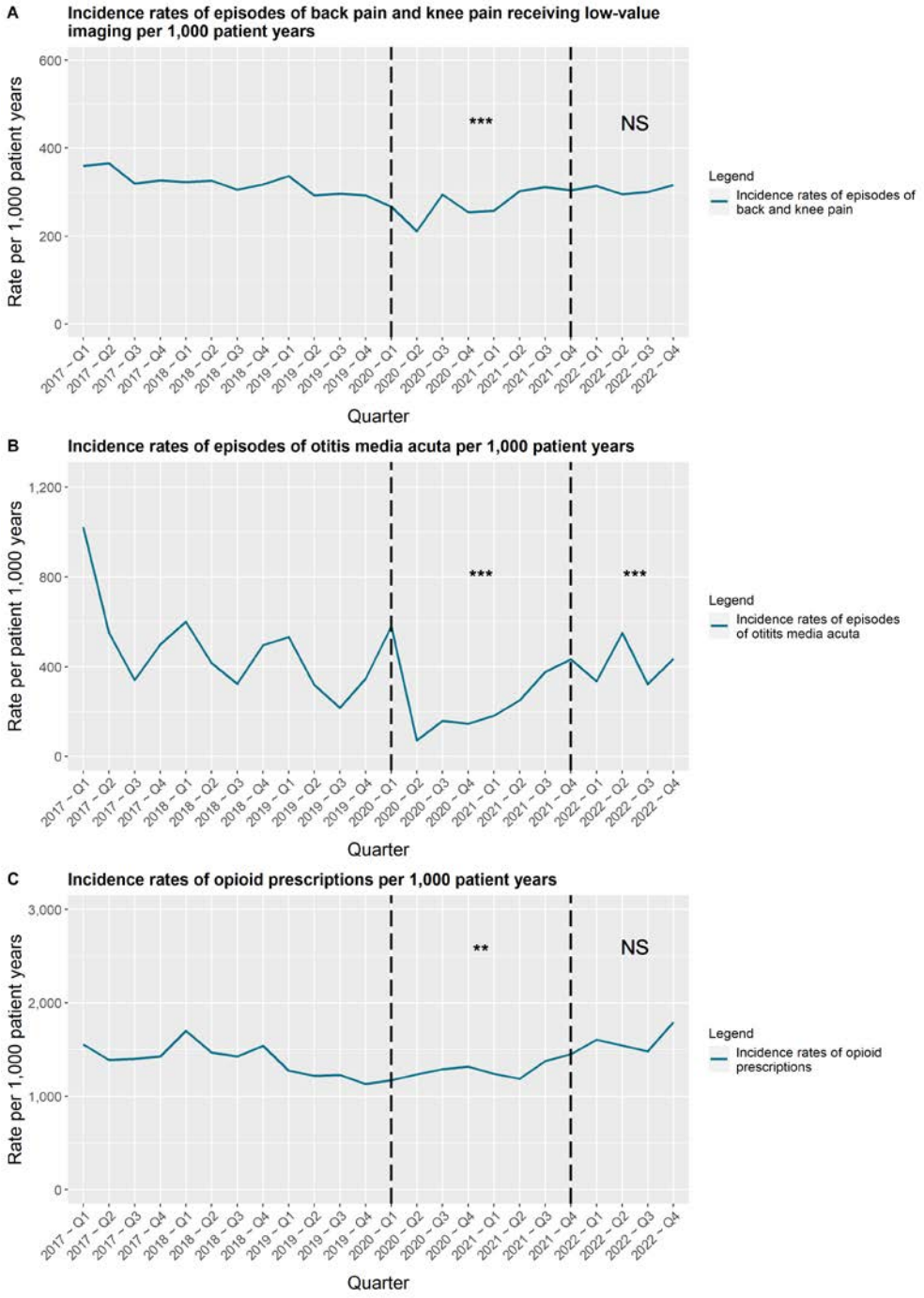


Figure 2 | Trends in incidence rates of episodes or prescriptions per 1,000 patient years corresponding to each of the three types care examined. Significance levels: * indicates significance at 0.05 level,** indicates significance at 0.01 level, *** indicates significance at 0.001 level, NS indicates no significance difference was found.

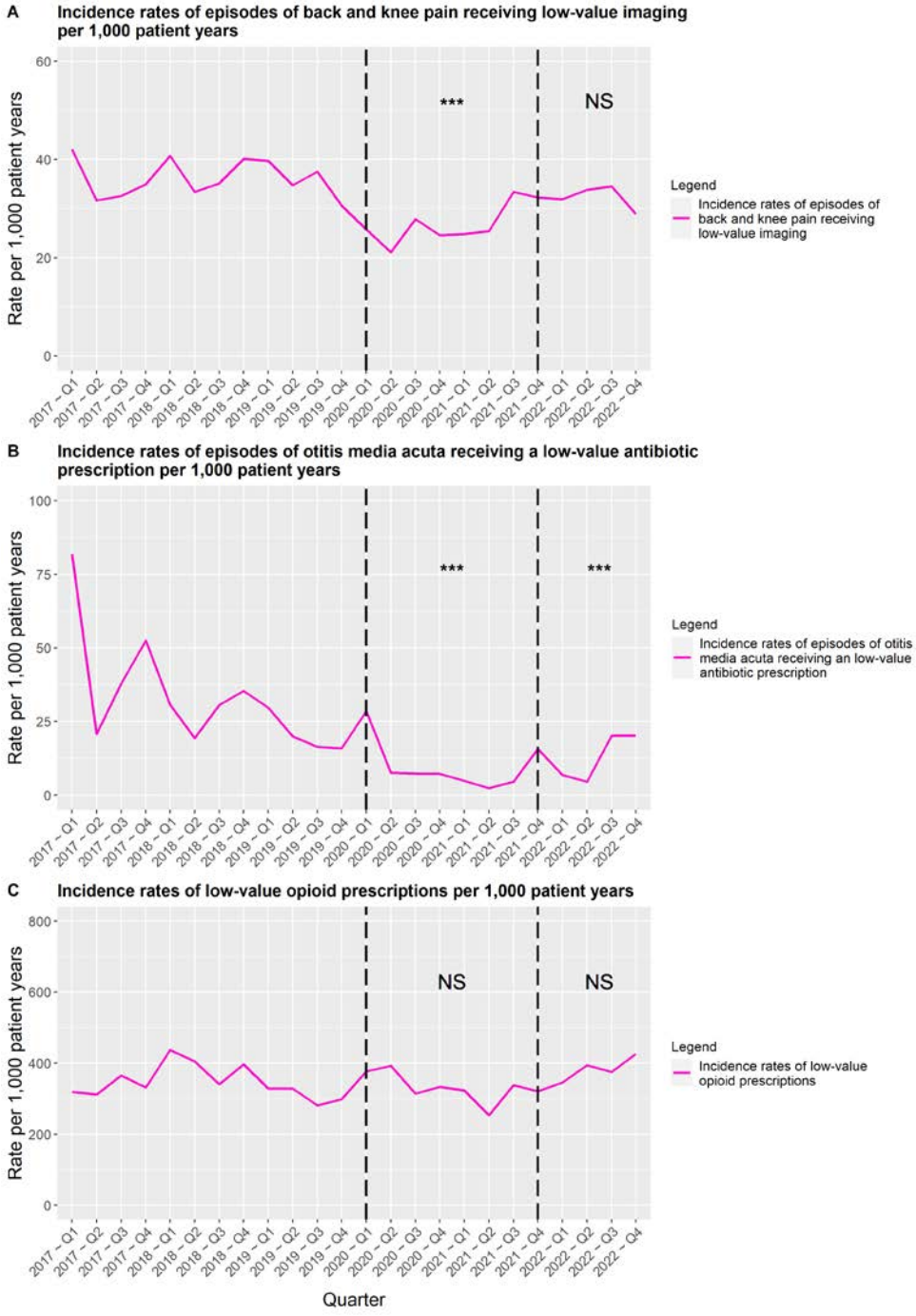


Figure 3 | Trends in incidence rates of low-value care provision per 1,000 patient years for the three types of care examined. Significance levels: * indicates significance at 0.05 level,** indicates significance at 0.01 level, *** indicates significance at 0.001 level, NS indicates no significant difference was observed.

Table 3 | Rate ratios over the different periods.

	Rate ratio of episodes / prescriptions (incl. 95% CI)			Rate ratio of low-value episodes / prescriptions (incl. 95% CI)		
	Pre COVID-19 restrictions period (reference period)	COVID-19 restrictions period	Post COVID-19 restrictions period	Pre COVID-19 restrictions period (reference period)	COVID-19 restrictions period	Post COVID-19 restrictions period
The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.	1.0	0.88 *** [0.85 – 0.91]	0.96 [0.92 – 1.0]	1.0	0.79 *** [0.71 - 0.88]	0.93 [0.83 - 1.06]
The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.	1.0	0.46 *** [0.42 – 0.51]	0.78 *** [0.71 – 0.86]	1.0	0.22 *** [0.14 – 0.36]	0.37 *** [0.23 – 0.61]
Repeat opioid prescriptions, without a prior GP visit.	1.0	0.87 ** [0.79 – 0.96]	1.06 [0.96 – 1.18]	1.0	0.93 [0.79 – 1.09]	1.05 [0.88 – 1.24]

* significant at 0.05 level,** significant at 0.01 level, *** significant at 0.001 level.

Discussion

Summary

This study shows that the COVID-19 related restrictions have affected both the IRs of episodes and low-value care provision among the clinical scenarios examined. The IRs of episodes of all three types of care significantly decreased during the COVID-19 restrictions period. Only the IR of antibiotic prescriptions for OMA remained significantly lower (22%) over the post-COVID-19 restrictions period. The rates at which low-value care was provided during the COVID-19 period, significantly decreased in case of back and knee imaging (21%) and the prescription of antibiotics for OMA (78%). During the post-restrictions period, only the decrease in IR of the low-value antibiotic prescription for OMA remained lower ($p<0.001$). The IR of inappropriately repeated opioid prescriptions remained unchanged over all three periods ($p>0.05$).

Strengths and limitations

This is the first study to examine the effects of the COVID-19 restrictions on the provision of low-value care among GPs. We used routinely collected, highly detailed and high quality clinical information. GPs part of the FaMe-Net meet regularly to discuss and

review the coding system to ensure that the quality and validity of data registration remains high. [42] The availability of such reliable and detailed clinical information enabled us to accurately distinguish the appropriateness of the examined clinical scenarios. However, this study is also prone to some limitations. First, this study was conducted using data collected from only 6 out of approximately 4,874 practices in the Netherlands. [43] However, the patient population of the GP practices included in the network has been shown to be representative for the Netherlands with respect to age, sex and social class. [44, 45] We therefore expect our findings to be generalizable to the entire Dutch population. Second, due to the use of a reasonably small network of practices, we were limited with respect to the depth of our analysis. For example, we observed that we had to few data points of our outcome measure over the included months or even quarters to include an seasonal effect in our models. Third, our assessment of low-value repeat opioid prescriptions could be an underestimate as we limited ourselves to the prescriptions prescribed within the time period examined. Some opioid prescriptions might therefore have been wrongly classified as appropriate, while we did not take into account prescriptions prescribed shortly before the inclusion period. However, we do not expect this to have a large impact on the found results while opioids prescriptions are generally short. Fourth, we were only able to include several patient characteristics in our models, but were not able to correct for GP characteristics (such as age and sex). Lastly, we want to note that indeed the operationalization of recommendations in guidelines is often a challenging task, mainly because we were limited with respect to the applicable Dutch guidelines for GPs and subsequent assumptions that had to be made. For example, in our examination of the prescription of antibiotics for OMA in children we had to rely on the information presented to us in the Dutch GP guidelines. Stating that OMA related complaints generally will resolve themselves within 48 – 72 hours, and that after 72 hours of complaints the prescription of antibiotics is considered appropriate. [31] We were unable to find any other information regarding the proper definition of severe symptoms in case of OMA, applicable to Dutch GP care. Additionally, some countries issued temporary modified guidelines for GP care during COVID-19, such as in long-term pain management with opioids. [46] However, the relevant GP bodies in the Netherlands have not adjusted their guidelines during COVID-19, resulting in us needing to use the existing guidelines.

Comparison with existing literature

Our findings regarding the observed differences in both trends in episodes [47-51] and low-value services are in line with previous studies in hospital care. [27, 52-54] Hence, both the number and rates of episodes and low-value care provision were largely affected during the first months of the COVID-19 pandemic. Furthermore, our finding that the pandemic differentially affected the provision of the different types of low-value service, complies with assessments regarding the impact of COVID-19 healthcare from the US, albeit it being conducted in hospitals. [27] Hence Shahzad et al. demonstrated that the pandemic had varying effects on low-value care provision, with some types not rebounding afterward.

Implications for research and/or practice

The results of our assessment show that the introduction of the COVID-19 restrictions have differentially affected low-value GP care. Reasons for which could be found in the severity of the complaints of the different clinical scenarios examined. In both the case of imaging for back or knee pain or the prescription of OMA, the implemented restrictions did not affect the patients' complaint status. Hence, the symptoms of a patient with back or knee pain do not diminish after having received an imaging procedure. Additionally, OMA related complaints often resolve themselves over time (e.g. 2-3 days) without the prescription of an antibiotic. In both cases, the patient conditions do not necessarily deteriorates but could potentially even improve. Conversely, in case of the prescriptions of opioids, generally the patient's condition deteriorates while these are often prescribed for patients suffering from long-term or chronic pain syndromes. This notion could provide an explanation as to why we observed that in case of opioids (almost) no change in prescription rates was observed, while the rates of the other types of care did show to change (and the number and distribution of patients remained the same over the period examined). Furthermore, the observation that the COVID-19 restrictions differentially affected low-value GP care provision supports the idea that deïmplementation of low-value care requires tailored interventions. [55, 56] A recently published review showed that among the existing studies examining the impact of deïmplementation strategies showed that strategies targeting healthcare providers, patients or organizational context are often more effective. [55] Suggesting that the provision of low-value care is often the result of an interplay of factors existing on multiple levels. For example, although healthcare providers often try to provide the best care possible, implemented systems on the level of the hospital could often hinder them in its provision. However, because the COVID-19 pandemic affected the entire healthcare system and was noticeable across all levels of healthcare provision it might have alleviated some of the barriers which earlier prevented the provision of appropriate care.

Future research should investigate both the (potential) mechanisms underlying the observed changes in the IR of low-value care provision over the COVID-19 period for some of the examined types of care, as well as the GPs' perspective as to why these changes in the IR of low-value care shows such different patterns. Hence, the IR of some of the examined types of care decreased during COVID-19, but rebounded afterwards in some cases (however, this was not the case in the IRs associated to OMA). Additionally, further examination of patient and physician characteristics associated with either the provision or reception of low-value GP care is warranted. While these insights could also be used to further develop interventions aiming to reduce low-value care. Furthermore, exploring whether similar trends can be observed in the use of high-value GP services could also be valuable.

Conclusion

This study shows that both the IRs of episodes and low-value care provision among Dutch GPs are affected by the COVID-19 restrictions, although differences between the clinical scenarios were identified Additionally, our findings indicate that only in some cases the COVID-19 restrictions could have had a lasting effect on the provision of low-value care. The combination of these findings confirm the idea that reducing low-value care is a complex challenge; which requires tailormade interventions and which is not easily nor quickly achieved.

References

1. van Dulmen SA, Tran NH, Wiersma T, Verkerk EW, Messaoudi JCL, Burgers JS, Kool RB. Identifying and prioritizing do-not-do recommendations in Dutch primary care. BMC Primary Care. 2022;23(1):141.
2. World Health Organization (WHO). Public Health and Social Measures in Response to COVID-19: World Health Organization (WHO),; 2023 [updated 09-03-2023. Available from: https://phsm.euro.who.int/.
3. Sarac NJ, Sarac BA, Schoenbrunner AR, Janis JE, Harrison RK, Phieffer LS, et al. A Review of State Guidelines for Elective Orthopaedic Procedures During the COVID-19 Outbreak. JBJS. 2020;102(11):942-5.
4. Anderson KE, McGinty EE, Presskreischer R, Barry CL. Reports of Forgone Medical Care Among US Adults During the Initial Phase of the COVID-19 Pandemic. JAMA Netw Open. 2021;4(1):e2034882.
5. Giannouchos TV, Brooks JM, Andreyeva E, Ukert B. Frequency and factors associated with foregone and delayed medical care due to COVID-19 among nonelderly US adults from August to December 2020. J Eval Clin Pract. 2022;28(1):33-42.
6. Baum A, Schwartz MD. Admissions to Veterans Affairs Hospitals for Emergency Conditions During the COVID-19 Pandemic. JAMA. 2020;324(1):96-9.
7. Maringe C, Spicer J, Morris M, Purushotham A, Nolte E, Sullivan R, et al. The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study. The Lancet Oncology. 2020;21(8):1023-34.
8. Schers H, van Weel C, van Boven K, Akkermans R, Bischoff E, olde Hartman T. The COVID-19 Pandemic in Nijmegen, the Netherlands: Changes in Presented Health Problems and Demand for Primary Care. Ann Fam Med. 2021;19(1):44-7.
9. M.S. L, M. H, L. J, M. M, S. V, J.; dJ. Avoiding GP care during the coronavirus pandemic. . Dutch National Institute for Public Health and the Environment 2021.
10. Whaley CM, Pera MF, Cantor J, Chang J, Velasco J, Hagg HK, et al. Changes in Health Services Use Among Commercially Insured US Populations During the COVID-19 Pandemic. JAMA Netw Open. 2020;3(11):e2024984.
11. Pronovost PJ, Cole MD, Hughes RM. Remote Patient Monitoring During COVID-19: An Unexpected Patient Safety Benefit. JAMA. 2022;327(12):1125-6.
12. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in Outpatient Care Delivery and Telemedicine During the COVID-19 Pandemic in the US. JAMA Internal Medicine. 2021;181(3):388-91.
13. Ateev Mehrotra et al. The Impact of the COVID-19 Pandemic on Outpatient Visits: Practices Are Adapting to the New Normal (Commonwealth Fund, June 2020)2020. Available from: https://www.commonwealthfund.org/publications/2020/jun/impact-covid-19-pandemic-outpatient-visits-practices-adapting-new-normal.
14. Mehrotra AC, M.E.; Linetsky, D.; Hatch, H.; Cutler, D.A.; Schneider, E.C. The Impact of COVID-19 on Outpatient Visits in 2020: Visits Remained Stable, Despite a Late Surge in Cases. The Commonwealth Fund; 2021.

15. Gyawali B, Poudyal BS, Eisenhauer EA. Covid-19 Pandemic—An Opportunity to Reduce and Eliminate Low-Value Practices in Oncology? JAMA Oncology. 2020;6(11):1693-4.
16. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. N Engl J Med. 2020;382(21):2049-55.
17. Marshall M, Howe A, Howsam G, Mulholland M, Leach J. COVID-19: a danger and an opportunity for the future of general practice. British Journal of General Practice. 2020;70(695):270-1.
18. Moynihan R, Johansson M, Maybee A, Lang E, Légaré F. Covid-19: an opportunity to reduce unnecessary healthcare. BMJ. 2020;370:m2752.
19. Auener S, Kroon D, Wackers E, van Dulmen S, Jeurissen P. COVID-19: A Window of Opportunity for Positive Healthcare Reforms. International Journal of Health Policy and Management. 2020;9(10):419-22.
20. Corral-Gudino L. Tools to Reduce Low-Value Care: Lessons From COVID-19 Pandemic Comment on "Key Factors that Promote Low-Value Care: Views of Experts From the United States, Canada, and the Netherlands". Int J Health Policy Manag. 2022.
21. Clarke M, Born K, Johansson M, Jørgensen KJ, Levinson W, Madrid E, et al. Making wise choices about low-value health care in the COVID-19 pandemic. Cochrane Database of Systematic Reviews. 2021(9).
22. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, Heath I, et al. Evidence for overuse of medical services around the world. The Lancet. 2017;390(10090):156-68.
23. Colla CH. Swimming against the current--what might work to reduce low-value care? N Engl J Med. 2014;371(14):1280-3.
24. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. Med J Aust. 2015;203(4):179-81.
25. Martins C, Godycki-Cwirko M, Heleno B, Brodersen J. Quaternary prevention: an evidence-based concept aiming to protect patients from medical harm. Br J Gen Pract. 2019;69(689):614-5.
26. Norman AH, Tesser CD. Quaternary prevention: a balanced approach to demedicalisation. Br J Gen Pract. 2019;69(678):28-9.
27. Shahzad M, Zirui Song MDP, Chernew M, Fendrick A. Changes in Use of Low-Value Services During the COVID-19 Pandemic. The American Journal of Managed Care. 2022;28.
28. Dutch College of General Practitioners. Guideline sciatica Dutch College of General Practitioners; 2015.
29. Dutch College of General Practitioners. Non-specific lower back pain. Dutch College of General Practitioners; 2017.
30. Dutch College of General Practitioners. Non-traumatic knee complaints. 2020.
31. Dutch College of General Practitioners. Otitis media acuta in children. 2023.
32. Dutch College of General Practitioners. Pain. 2021.
33. Department of primary and community. TransHis Electronic Patient Records 2022 [Available from: https://www.transhis.nl/over/over-transhis/transhis-elektronisch-patientendossier/.

34. Rijpkema C, Ramerman L, Homburg M, Meijer E, Muris J, olde Hartman T, et al. Care by general practitioners for patients with asthma or COPD during the COVID-19 pandemic. npj Primary Care Respiratory Medicine. 2023;33(1):15.
35. Groenewegen PP, van den Muijsenbergh M, Batenburg R, Van Poel E, van den Broek S, Bussche PV, Willems S. Quick adaptation of the organisation of general practices during the COVID-19 pandemic in the Netherlands. BMC Primary Care. 2023;24(1):170.
36. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
37. National Institute for Public Health and the Environment. Covid-19 measures timeline 2022 [cited 2022 25-11-2022]. Available from: https://www.rivm.nl/gedragsonderzoek/tijdlijn-maatregelen-covid.
38. Austin PC, Stryhn H, Leckie G, Merlo J. Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. Stat Med. 2018;37(4):572-89.
39. Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
40. David JL, Kaplan HB. Gender, social roles and health care utilization. Applied Behavioral Science Review. 1995;3(1):39-64.
41. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract. 2000;49(2):147-52.
42. Ballering AV, Muijres D, Uijen AA, Rosmalen JGM, Olde Hartman TC. Sex differences in the trajectories to diagnosis of patients presenting with common somatic symptoms in primary care: an observational cohort study. J Psychosom Res. 2021;149:110589.
43. Batenburg R, Flinterman, L., Vis, E., Schaaijk, A. van, Kenens, R.J., Duijkers, B. Cijfers uit de Nivel-registratie van huisartsen en huisartsenpraktijken: een actualisering voor de periode 2020-2022. Nivel; 2022.
44. Luijks H, van Boven K, olde Hartman T, Uijen A, van Weel C, Schers H. Purposeful Incorporation of Patient Narratives in the Medical Record in the Netherlands. The Journal of the American Board of Family Medicine. 2021;34(4):709-23.
45. van Boven K, Uijen AA, van de Wiel N, Oskam SK, Schers HJ, Assendelft WJJ. The Diagnostic Value of the Patient's Reason for Encounter for Diagnosing Cancer in Primary Care. J Am Board Fam Med. 2017;30(6):806-12.
46. Narouze HSSPCNSCLSEABS. Recommendations on Chronic Pain Practice during the COVID-19 Pandemic 2020 [cited 2024. Available from: https://www.asra.com/news-publications/asra-updates/blog-landing/legacy-b-blog-posts/2020/03/27/recommendations-on-chronic-pain-practice-during-the-covid-19-pandemic.
47. Choi SY, Yon DK, Choi YS, Lee J, Park KH, Lee YJ, et al. The Impact of the COVID-19 Pandemic on Otitis Media. Viruses. 2022;14(11):2457.
48. Hullegie S, Schilder AGM, Marchisio P, de Sévaux JLH, van der Velden AW, van de Pol AC, et al. A Strong Decline in the Incidence of Childhood Otitis Media During the COVID-19 Pandemic in the Netherlands. Front Cell Infect Microbiol. 2021;11:768377.

49. van de Pol AC, Boeijen JA, Venekamp RP, Platteel T, Damoiseaux RAMJ, Kortekaas MF, van der Velden AW. Impact of the COVID-19 Pandemic on Antibiotic Prescribing for Common Infections in The Netherlands: A Primary Care-Based Observational Cohort Study. Antibiotics. 2021;10(2):196.
50. Rikin S, Perez HR, Zhang C, Khalid L, Groeger J, Deng Y, Starrels JL. Changes in Outpatient Opioid Prescribing During the COVID-19 Pandemic: An Interrupted Time Series Analysis. J Prim Care Community Health. 2022;13:21501319221076926.
51. Sharma S, Traeger AC, Machado GC, Abdel Shaheed C, Jones C, Maher CG. Effect of COVID-19 on management of patients with low back pain in the emergency department. Australasian Emergency Care. 2022;25(2):154-60.
52. Levine DM, Samal L, Neville BA, Burdick E, Wien M, Rodriguez JA, et al. The Association of the First Surge of the COVID-19 Pandemic with the High- and Low-Value Outpatient Care Delivered to Adults in the USA. Journal of General Internal Medicine. 2022;37(15):3979-88.
53. Chen RC, Haynes K, Du S, Barron J, Katz AJ. Association of Cancer Screening Deficit in the United States With the COVID-19 Pandemic. JAMA Oncology. 2021;7(6):878-84.
54. Mehrotra A CM, Linetsky D, Hatch H, Cutler DA. The impact of the COVID-19 pandemic on outpatient visits: practices are adapting to the new normal. The Commonwealth Fund; 2021 11-01-2023.
55. Heus P, van Dulmen SA, Weenink JW, Naaktgeboren CA, Takada T, Verkerk EW, et al. What are Effective Strategies to Reduce Low-Value Care? An Analysis of 121 Randomized Deimplementation Studies. J Healthc Qual. 2023;45(5):261-71.
56. Ingvarsson S, Hasson H, von Thiele Schwarz U, Nilsen P, Powell BJ, Lindberg C, Augustsson H. Strategies for de-implementation of low-value care—a scoping review. Implementation Science. 2022;17(1):73.

Supplementary materials

Additional file 1: overview of both the included recommendations in each of the topics examined and their operationalization

1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.

The following do-not-do recommendations were included in the operationalization of the low-value use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.

- The use of imaging is not recommended in case of non-specific lower back pain. While no clear correlation exists between deviations of the Vertebral column visible on images and the presence of non-specific lower back pain.
- The ordering of imaging (X-ray, CT- or MRI-scan) in patients with sciatica is not recommended, unless there are clear signs of the sciatica originating from a severe cause.
- The use of imaging (such as x-ray or MRI-scan) to diagnose or exclude the existence of knee arthrosis, Prepatellar bursitis, Iliotibial band syndrome, Patellar Tendinitis/ Jumper's Knee or Osgood–Schlatter disease is not recommended, since the relation between deviations found during imaging procedures and the patient's complaints remains unclear.

ICPC-2 codes included	ICD-10 codes included	Operationalisation of recommendations
a. Back symptom/complaint (L02) b. Low back symptom/complaints [ex. L86] (L03) c. Back syndrome without radiating pain (L84) d. Acquired deformity of spine (L85) e. Back syndrome with radiating pain (L86) f. Sprain/strain of knee (L78) g. Knee symptom/complaint (L15) h. Bursitis/tendinitis/synovitis not otherwise specified (L87) i. Osteoarthrosis of knee (L90) j. Osgood–Schlatter disease (L94) k. Musculoskeletal disease other (L99)	a. Patellofemoral disorders (M22.2) b. Prepatellar bursitis (M70.4) c. Iliotibial band syndrome (M76.3) d. Osgood–Schlatter disease (M92.5) e. Patellar Tendinitis/ Jumper's Knee (M76.50)	The International Classification of Primary Care, 2nd edition (ICPC-2) and International Classification of Diseases and Related Health Problems (ICD-10) codes related to back and knee pain as shown on the left were included in our definition of musculoskeletal complaints. All episodes which had either of the described ICD-10 or ICPC-2 codes associated to them were extracted from the database and included in our analysis. Next, all contacts having a code indicating they resulted in a imaging procedure associated to them were matched to each episode based on the unique episode identifier. After which we marked all episodes that had such a contact associated to them, as having received imaging. And these therefore were considered as having received low-value imaging. The following age categories were used in our examination of this type of low-value use of imaging in case of musculoskeletal problems: 0-18, 18-50, 50-70 and 70+.

2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.
The following do-not-do recommendations were included in the operationalization of the low-value prescription of antibiotics for otitis media acuta in children.
In children with an episode of otitis media without major general symptoms of illness (fever) and without any risk of complications, treatment with an oral antimicrobial agent is not indicated.

ICPC-2 codes included	ATC codes included	Operationalisation of recommendations
a. Otitis Media Acuta / Myringitis (H71)	a. All ATC codes belonging to J01 (Antibacterials for systemic use): - J01A Tetracyclines - J01B Amphenicols - 01C Beta-lactam antibacterials, penicillins - J01D Other Beta-lactam antibacterials - J01E Sulfonamides and trimethoprim - J01F Macrolides, lincosamides and streptogramins - J01G Aminoglycoside antibacterials - J01M Quinolone antibacterials - J01R Combinations of antibacterials - J01X Other antibacterials	First, all children (age < 18) with an episodes of otitis media acuta (OMA) were extracted from the database. Next, all prescriptions of antimicrobial agents were matched to the distinct OMA episode based on episode number and prescription date. We defined children with severe symptoms as children which had a reason for encounter of at least 72 hours. This means that the child has experienced OMA related complaints for at least 72 hours. In case a child did not have a reason for encounter of at least 72 hours but did receive a prescription for antibiotics, we marked that prescription as being of low-value. In our examination of the prescription of antibiotics for otitis media acuta (OMA) in children we only included patients aged 18 or less. This resulted in us using the following age categories in our examination: 0-1, 1-5, 5-12, 12-18.

3. Repeat opioid prescriptions, without a prior GP visit.

The following do-not-do recommendations were included in the operationalization of repeat opioid prescriptions, without a prior GP visit.

- Do not automatically repeat opioids, without a prior GP consult. Ask for pain mitigation and stop with prescribing opioids when the pain mitigation is considered to be inadequate.
- Try to limit the time an opioid is prescribed to reduce the risk of side effects and excessive use because of habituation and dependency.

ICPC-2 codes included	ATC codes included	Operationalisation of recommendations
N.A.	a. All ATC codes belonging to N02A (Opioids): - N02AA: Natural opium alkaloids - N02AB: Phenylpiperidine derivatives - N02AC Diphenylpropylamine derivatives - N02AD Benzomorphan derivatives - N02AE Oripavine derivatives - N02AF Morphinan derivatives - N02AG Opioids in combination with antispasmodics - N02AJ Opioids in combination with non-opioid analgesics - N02AX Other opioids	All patients that received an opioid prescription over the examined period were included in our analysis. We did not include the initial opioid prescriptions in our analysis, while these simply cannot be considered a repeat prescriptions. In our analysis of the appropriateness of repeat opioid prescriptions, the following GP contacts were included as justifying an repeat prescription: consultations, visits, telephone and email contacts to both GP practices and after hours clinics. Next, the identified GP contacts were matched to their respective repeat opioid prescription based on the contact and prescription dates. These had to match in order for the repeat opioid prescription to be considered as being appropriate. Repeat prescriptions that did not have a contact associated to them were considered as being of low-value. In our examination of inappropriate repeat opioid prescriptions, we aimed to apply the same age categories as was the case for the inappropriate imaging assessment. However, there were too little number of prescriptions in the lowest age category. We therefore ended up having to merge the two lowest categories resulting in the following age division: 0-50, 50-70 and 70+.

Additional file 2: overview of COVID-19 restrictions implemented over the examined period

1 2020:

- March 12th: first restrictions implemented to prevent spread of the COVID-19 virus; intelligent lockdown, Dutch citizens are advised to remain at home when experiencing mild COVID-19 symptoms.
- March 15th: closure of food and drinking venues, schools and day-cares. Society is recommended to keep at least 1,5m distance to one another.
- October 14th: second wave of COVID-19, partial lockdown is implemented. Facemasks are made mandatory in public domains and transport, and a maximum number of three people that are allowed to be received at home. All food and drinking venues are required to close their business until further notice (take away still a possibility).
- November 4th: the initial restrictions do not reduce the amount of COVID-19 infections, more severe restrictions are therefore implemented. The Dutch population is advised to remain at home when possible. Furthermore, visitors both inside and outside are further limited to two a day, or one household and public accessible venues are closed. Exercising is limited to two persons (group workouts are prohibited), while remaining at 1.5 m distance (exception for children up to 17 years old and professional athletes)
- December 14th: Closure of non-essential stores, gyms, day-cares and primary schools. Other schools are required to teach digitally.

2 2021:

- January 6th: start of vaccination campaign among healthcare personnel.
- January 20rd: Visiting policy tightened to only 1 visitor (aged 13 and older), once a day.
- January 23rd: evening curfew for the entire Dutch population is implemented, and continuation of the lockdown that started in the previous month. Vaccination of the Dutch population is started.
- March 23rd: Start third COVID-19 wave, continuation of the lockdown restrictions.
- June 5th: end of lockdown, most restrictions are revoked and society is re-opened.
- November 13th: after a sharp increase in COVID-19 related infections, several COVID-19 restrictions are reinstated including an a (partial) evening lockdown (from 05:00 P.M. most non-essential venues are closed). Furthermore, society is advised to work from home again and the number visitors (aged 13 and over) is limited to four per household.
- December 19th: Hard lockdown is implemented, following detection of the Omicron variant of COVID-19. Resulting in the closure of primary and higher education (high schools, universities and vocational schools), food and drinking venues, non-essential stores, sport club, cultural venues and all public meeting places.

3 2022:

- January 10th: first alleviations of the COVID-19 restrictions implemented during the hard lockdown of 2021 are announced; reopening of both primary and high school education and day-cares.
- January 15th: reopening of universities, universities of applied sciences and vocational schools and most sport clubs and shops, shops are allowed to open until 5 pm in the afternoon. An general advise is given to wear facemasks when 1,5m distance cannot be uphold. Furthermore, up to 4 visitors are allowed each day, excluding children up until 12 years old.
- January 26th: reopening of restaurants, music and theatre venues (including zoos, cinemas, museums and amusement parks.
- February 15th: no more restrictions with respect to visitors. Furthermore, remote working is no longer mandatory, up to half the time one is now allowed to work at the office.
- February 22nd: Regular opening times for shops are reintroduced and the 1,5m distance rule is discontinued.
- March 15th: COVID-19 rules are now considered to be advises.
- March 23rd: last COVID-19 restrictions (such as mandatory masks required in public transportation) are alleviated.
- April 1st: No more COVID-19 restrictions implemented.

Additional file 3: extended versions showing both the total no. of episodes/prescriptions and low-value episodes/prescriptions as a whole and separated for each of the types of care examined

1. Total number of episodes and number and proportion of episodes receiving low-value care for the three types of GP care examined.

	Pre-COVID-19 restrictions period													COVID-19 restrictions period								Post-COVID-19 restrictions period			
Year and quarter	2017 Q1	2017 Q2	2017 Q3	2017 Q4	2018 Q1	2018 Q2	2018 Q3	2018 Q4	2019 Q1	2019 Q2	2019 Q3	2019 Q4	2020 Q1	2020 Q2	2020 Q3	2020 Q4	2021 Q1	2021 Q2	2021 Q3	2021 Q4	2022 Q1	2022 Q2	2022 Q3	2022 Q4
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.																								
No. of episodes	955	955	882	894	861	880	823	887	881	777	785	750	699	544	751	651	661	750	764	729	750	688	699	717
No. episodes receiving low-value care	92	69	71	79	92	78	81	98	86	79	89	70	62	49	68	58	58	58	77	70	74	78	78	66
% of episodes receiving low-value treatment	9.6	7.2	8.0	8.8	10.7	8.9	9.8	11.0	9.8	10.2	11.3	9.3	8.9	9.0	9.1	8.9	8.8	7.7	10.1	9.6	9.9	11.3	11.2	9.2
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.																								
No. of episodes	225	133	90	143	176	129	106	168	179	112	79	131	223	28	65	61	75	106	165	193	147	244	144	194
No. episodes receiving low-value care	18	5	10	15	9	6	10	12	10	7	6	6	11	3	3	3	2	1	2	7	3	2	9	9
% of episodes receiving low-value treatment	8.0	3.8	11.1	10.5	5.1	4.7	9.4	7.1	5.6	6.3	7.6	4.6	4.9	10.7	4.6	4.9	2.7	0.9	1.2	3.6	2.0	0.8	6.3	4.6
3. Repeat opioid prescriptions, without a prior visit																								
No. of episodes	1,096	997	1,023	1,049	1,233	1,080	1,067	1,151	866	839	855	787	805	843	883	893	814	780	908	952	1023	983	939	1,122
No. episodes receiving low-value care	225	224	267	244	317	298	255	296	224	226	196	208	259	268	215	226	212	166	223	211	220	251	238	267
% of episodes receiving low-value treatment	20.5	22.5	26.1	23.3	25.7	27.6	23.9	25.7	25.9	26.9	22.9	26.4	32.2	31.8	24.3	25.3	26.0	21.3	24.6	22.2	21.5	25.5	25.3	23.8

2. Number of (low-value) episodes/prescriptions separated for each of the types of care examined.
A) The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.

Year and quarter	2017 - Q1		2017 - Q2		2017 - Q3		2017 - Q4		2018 - Q1		2018 - Q2		2018 - Q3		2018 - Q4	
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.																
No. of episodes - age categoriess	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	22	22	32	27	16	22	29	21	31	23	20	30	22	29	36	29
19-50	181	265	153	248	174	234	196	224	166	237	171	238	151	206	169	224
50-70	136	162	143	192	130	147	120	165	110	160	128	154	117	155	126	160
70+	68	99	77	83	76	83	67	72	47	87	68	71	67	76	55	88
Total no. of episodes	407	548	405	550	396	486	412	482	354	507	387	493	357	466	386	501
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	3	0	1	2	1	2	2	1	4	2	2	1	2	2	4	4
% low-value	13.64	0.00	3.13	7.41	6.25	9.09	6.90	4.76	12.90	8.70	10.00	3.33	9.09	6.90	11.11	13.79
19-50	11	23	12	7	6	18	8	13	8	16	9	16	8	10	15	20
% low-value	6.08	8.68	7.84	2.82	3.45	7.69	4.08	5.80	4.82	6.75	5.26	6.72	5.30	4.85	8.88	8.93
50-70	9	22	13	15	10	13	15	22	10	27	12	23	21	17	15	19
% low-value	6.62	13.58	9.09	7.81	7.69	8.84	12.50	13.33	9.09	16.88	9.38	14.94	17.95	10.97	11.90	11.88
70+	11	13	6	13	11	10	5	13	6	19	7	8	10	11	9	12
% low-value	16.18	13.13	7.79	15.66	14.47	12.05	7.46	18.06	12.77	21.84	10.29	11.27	14.93	14.47	16.36	13.64
Total no of episodes with low-value treatment	34	58	32	37	28	43	30	49	28	64	30	48	41	40	43	55
% low-value	8.35	10.58	7.90	6.73	7.07	8.85	7.28	10.17	7.91	12.62	7.75	9.74	11.48	8.58	11.14	10.98

Year and quarter	2021 - Q1		2021 - Q2		2021 - Q3		2021 - Q4		2022 - Q1		2022 - Q2		2022 - Q3		2022 - Q4	
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	32	30	31	41	33	32	32	30	38	38	46	39	27	39	28	43
19-50	139	155	142	176	165	207	153	194	152	201	144	146	149	172	131	187
50-70	101	110	97	147	102	120	104	116	100	119	100	112	105	120	101	129
70+	38	56	51	65	51	54	51	49	43	59	35	66	35	52	40	58
Total no. of episodes	310	351	321	429	351	413	340	389	333	417	325	363	316	383	300	417
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	2	4	2	5	0	2	2	4	5	1	4	3	2	4	1	4
% low-value	6.25	13.33	6.45	12.20	0.00	6.25	6.25	13.33	13.16	2.63	8.70	7.69	7.41	10.26	3.57	9.30
19-50	9	13	9	9	11	18	7	15	12	19	10	8	12	24	14	17
% low-value	6.47	8.39	6.34	5.11	6.67	8.70	4.58	7.73	7.89	9.45	6.94	5.48	8.05	13.95	10.69	9.09
50-70	15	5	7	12	16	19	7	18	7	19	13	24	12	16	6	12
% low-value	14.85	4.55	7.22	8.16	15.69	15.83	6.73	15.52	7.00	15.97	13.00	21.43	11.43	13.33	5.94	9.30
70+	5	5	6	8	2	9	10	7	5	6	6	10	4	4	4	8
% low-value	13.16	8.93	11.76	12.31	3.92	16.67	19.61	14.29	11.63	10.17	17.14	15.15	11.43	7.69	10.00	13.79
Total no of episodes with low-value treatment	31	27	24	34	29	48	26	44	29	45	33	45	30	48	25	41
% low-value	10.00	7.69	7.48	7.93	8.26	11.62	7.65	11.31	8.71	10.79	10.15	12.40	9.49	12.53	8.33	9.83

Year and quarter	2021 - Q1		2021 - Q2		2021 - Q3		2021 - Q4		2022 - Q1		2022 - Q2		2022 - Q3		2022 - Q4	
1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	32	30	31	41	33	32	32	30	38	38	46	39	27	39	28	43
19-50	139	155	142	176	165	207	153	194	152	201	144	146	149	172	131	187
50-70	101	110	97	147	102	120	104	116	100	119	100	112	105	120	101	129
70+	38	56	51	65	51	54	51	49	43	59	35	66	35	52	40	58
Total no. of episodes	310	351	321	429	351	413	340	389	333	417	325	363	316	383	300	417
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-18	2	4	2	5	0	2	2	4	5	1	4	3	2	4	1	4
% *low-value*	6.25	13.33	6.45	12.20	0.00	6.25	6.25	13.33	13.16	2.63	8.70	7.69	7.41	10.26	3.57	9.30
19-50	9	13	9	9	11	18	7	15	12	19	10	8	12	24	14	17
% *low-value*	6.47	8.39	6.34	5.11	6.67	8.70	4.58	7.73	7.89	9.45	6.94	5.48	8.05	13.95	10.69	9.09
50-70	15	5	7	12	16	19	7	18	7	19	13	24	12	16	6	12
% *low-value*	14.85	4.55	7.22	8.16	15.69	15.83	6.73	15.52	7.00	15.97	13.00	21.43	11.43	13.33	5.94	9.30
70+	5	5	6	8	2	9	10	7	5	6	6	10	4	4	4	8
% *low-value*	13.16	8.93	11.76	12.31	3.92	16.67	19.61	14.29	11.63	10.17	17.14	15.15	11.43	7.69	10.00	13.79
Total no of episodes with low-value treatment	31	27	24	34	29	48	26	44	29	45	33	45	30	48	25	41
% low-value	10.00	7.69	7.48	7.93	8.26	11.62	7.65	11.31	8.71	10.79	10.15	12.40	9.49	12.53	8.33	9.83

B) The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.

Year and quarter	2017 - Q1		2017 - Q2		2017 - Q3		2017 - Q4		2018 - Q1		2018 - Q2		2018 - Q3		2018 - Q4	
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	50	27	30	22	22	12	44	26	39	41	45	24	25	22	36	49
1-5	51	48	25	28	8	19	30	27	27	40	26	19	12	11	31	28
5-12	28	16	9	12	11	14	5	7	10	17	5	5	16	11	9	7
12-18	1	4	5	2	1	3	0	4	2	0	2	3	2	7	3	5
Total no. of episodes	130	95	69	64	42	48	79	64	78	98	78	51	55	51	79	89
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	7	2	0	2	2	0	3	0	2	1	2	0	0	2	4	3
% low-value	14.00	7.41	0.00	9.09	9.09	0.00	6.82	0.00	5.13	2.44	4.44	0.00	0.00	9.09	11.11	6.12
1-5	4	2	1	0	0	3	9	2	3	1	1	1	0	2	2	2
% low-value	7.84	4.17	4.00	0.00	0.00	15.79	30.00	7.41	11.11	2.50	3.85	5.26	0.00	18.18	6.45	7.14
5-12	3	0	0	0	4	1	0	1	1	0	0	2	2	3	0	0
% low-value	10.71	0.00	0.00	0.00	36.36	7.14	0.00	14.29	10.00	0.00	0.00	40.00	12.50	27.27	0.00	0.00
12-18	0	0	0	2	0	0	0	0	1	0	0	0	0	1	0	1
% low-value	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	14.29	0.00	20.00
Total no of episodes with low-value treatment	14	4	1	4	6	4	12	3	7	2	3	3	2	8	6	6
% low-value	10.77	4.21	1.45	6.25	14.29	8.33	15.19	4.69	8.97	2.04	3.85	5.88	3.64	15.69	7.59	6.74

Year and quarter	2019 - Q1		2019 - Q2		2019 - Q3		2019 - Q4		2020 - Q1		2020 - Q2		2020 - Q3		2020 - Q4	
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	45	44	26	25	16	19	38	38	51	50	8	9	18	13	25	14
1-5	40	32	22	13	9	11	17	21	51	49	4	1	9	14	11	9
5-12	6	11	8	8	9	9	6	6	1	17	3	3	3	4	1	0
12-18	0	1	8	2	3	3	0	5	2	2	0	0	1	3	0	1
Total no. of episodes	91	88	64	48	37	42	61	70	105	118	15	13	31	34	37	24
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	4	1	0	2	1	3	1	1	4	1	0	0	0	1	0	2
% low-value	8.89	2.27	0.00	8.00	6.25	15.79	2.63	2.63	7.84	2.00	0.00	0.00	0.00	7.69	0.00	14.29
1-5	4	1	1	0	0	1	2	0	2	2	1	0	0	0	0	0
% low-value	10.00	3.13	4.55	0.00	0.00	9.09	11.76	0.00	3.92	4.08	25.00	0.00	0.00	0.00	0.00	0.00
5-12	0	0	3	0	0	1	0	2	1	1	0	2	2	0	0	0
% low-value	0.00	0.00	37.50	0.00	0.00	11.11	0.00	33.33	100.00	5.88	0.00	66.67	66.67	0.00	0.00	0.00
12-18	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
% low-value	0.00	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total no of episodes with low-value treatment	8	2	5	2	1	5	3	3	7	4	1	2	2	1	0	3
% low-value	8.79	2.27	7.81	4.17	2.70	11.90	4.92	4.29	6.67	3.39	6.67	15.38	6.45	2.94	0.00	12.50

Year and quarter	2021 - Q1		2021 - Q2		2021 - Q3		2021 - Q4		2022 - Q1		2022 - Q2		2022 - Q3		2022 - Q4	
2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	23	25	46	22	57	35	77	39	43	21	67	46	33	19	28	55
1-5	7	16	18	13	34	20	31	22	43	25	42	42	36	21	52	39
5-12	1	3	5	2	6	4	3	10	4	10	18	17	23	8	7	9
12-18	0	0	0	0	2	7	4	7	1	0	5	7	1	3	3	1
Total no. of episodes	31	44	69	37	99	66	115	78	91	56	132	112	93	51	90	104
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-1	0	0	1	0	0	0	1	2	2	0	1	1	1	1	0	3
% low-value	0.00	0.00	2.17	0.00	0.00	0.00	1.30	5.13	4.65	0.00	1.49	2.17	3.03	5.26	0.00	5.45
1-5	0	1	0	0	0	0	2	0	0	1	0	0	2	2	3	3
% low-value	0.00	6.25	0.00	0.00	0.00	0.00	6.45	0.00	0.00	4.00	0.00	0.00	5.56	9.52	5.77	7.69
5-12	0	1	0	0	1	0	0	1	0	0	0	0	2	0	0	0
% low-value	0.00	33.33	0.00	0.00	16.67	0.00	0.00	10.00	0.00	0.00	0.00	0.00	8.70	0.00	0.00	0.00
12-18	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0
% low-value	0.00	0.00	0.00	0.00	0.00	14.29	25.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total no of episodes with low-value treatment	0	2	1	0	1	1	4	3	2	1	1	1	5	4	3	6
% low-value	0.00	4.55	1.45	0.00	1.01	1.52	3.48	3.85	2.20	1.79	0.76	0.89	5.38	7.84	3.33	5.77

C) Repeat opioid prescriptions, without a prior visit.

Year and quarter	2017 - Q1		2017 - Q2		2017 - Q3		2017 - Q4		2018 - Q1		2018 - Q2		2018 - Q3		2018 - Q4	
3. Repeat opioid prescriptions, without a prior visit																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	68	133	57	118	73	119	69	93	80	146	69	173	75	174	116	214
50-70	217	176	182	183	167	182	154	198	174	242	163	209	158	193	175	189
70+	188	314	175	282	190	292	174	361	194	397	150	316	146	321	168	289
Total no. of episodes	473	623	414	583	430	593	397	652	448	785	382	698	379	688	459	692
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	10	22	10	26	21	25	26	14	14	24	20	45	24	37	50	52
% low-value	14.71	16.54	17.54	22.03	28.77	21.01	37.68	15.05	17.50	16.44	28.99	26.01	32.00	21.26	43.10	24.30
50-70	51	34	53	28	51	48	34	42	44	61	44	75	46	41	46	37
% low-value	23.50	19.32	29.12	15.30	30.54	26.37	22.08	21.21	25.29	25.21	26.99	35.89	29.11	21.24	26.29	19.58
70+	45	63	47	60	49	73	44	84	61	113	28	86	44	63	49	62
% low-value	23.94	20.06	26.86	21.28	25.79	25.00	25.29	23.27	31.44	28.46	18.67	27.22	30.14	19.63	29.17	21.45
Total no of episodes with low-value treatment	106	119	110	114	121	146	104	140	119	198	92	206	114	141	145	151
% low-value	22.41	19.10	26.57	19.55	28.14	24.62	26.20	21.47	26.56	25.22	24.08	29.51	30.08	20.49	31.59	21.82

Year and quarter	2019 - Q1		2019 - Q2		2019 - Q3		2019 - Q4		2020 - Q1		2020 - Q2		2020 - Q3		2020 - Q4	
3. Repeat opioid prescriptions, without a prior visit																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	82	169	79	146	79	144	70	159	84	161	89	168	88	168	118	132
50-70	134	187	135	166	129	177	147	140	129	108	117	199	114	214	140	212
70+	88	206	84	229	100	226	83	188	89	234	86	184	88	211	85	206
Total no. of episodes	304	562	298	541	308	547	300	487	302	503	292	551	290	593	343	550
No. episodes receiving low-value c are - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	30	42	20	40	10	26	11	38	18	38	27	41	15	25	31	19
% low-value	36.59	24.85	25.32	27.40	12.66	18.06	15.71	23.90	21.43	23.60	30.34	24.40	17.05	14.88	26.27	14.39
50-70	29	46	35	35	27	38	39	30	43	28	35	72	33	56	45	59
% low-value	21.64	24.60	25.93	21.08	20.93	21.47	26.53	21.43	33.33	25.93	29.91	36.18	28.95	26.17	32.14	27.83
70+	26	51	27	69	35	60	29	61	36	96	23	70	19	67	20	52
% low-value	29.55	24.76	32.14	30.13	35.00	26.55	34.94	32.45	40.45	41.03	26.74	38.04	21.59	31.75	23.53	25.24
Total no of episodes with low-value treatment	85	139	82	144	72	124	79	129	97	162	85	183	67	148	96	130
% low-value	27.96	24.73	27.52	26.62	23.38	22.67	26.33	26.49	32.12	32.21	29.11	33.21	23.10	24.96	27.99	23.64

Year and quarter	2021 - Q1		2021 - Q2		2021 - Q3		2021 - Q4		2022 - Q1		2022 - Q2		2022 - Q3		2022 - Q4	
3. Repeat opioid prescriptions, without a prior visit																
No. of episodes - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	100	131	100	140	104	158	112	195	146	179	110	185	102	205	129	251
50-70	103	168	104	174	123	238	109	214	97	245	140	231	114	232	124	292
70+	110	202	77	185	96	189	106	216	129	227	129	188	93	193	108	218
Total no. of episodes	313	501	281	499	323	585	327	625	372	651	379	604	309	630	361	761
No. episodes receiving low-value care - age categories	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**
0-50	21	29	18	26	16	43	20	51	33	41	21	41	24	56	29	59
% low-value	21.00	22.14	18.00	18.57	15.38	27.22	17.86	26.15	22.60	22.91	19.09	22.16	23.53	27.32	22.48	23.51
50-70	32	30	20	30	39	44	15	47	19	43	42	66	38	67	27	73
% low-value	31.07	17.86	19.23	17.24	31.71	18.49	13.76	21.96	19.59	17.55	30.00	28.57	33.33	28.88	21.77	25.00
70+	41	59	18	54	35	46	12	66	26	58	34	47	10	43	21	58
% low-value	37.27	29.21	23.38	29.19	36.46	24.34	11.32	30.56	20.16	25.55	26.36	25.00	10.75	22.28	19.44	26.61
Total no of episodes with low-value treatment	94	118	56	110	90	133	47	164	78	142	97	154	72	166	77	190
% low-value	30.03	23.55	19.93	22.04	27.86	22.74	14.37	26.24	20.97	21.81	25.59	25.50	23.30	26.35	21.33	24.97

Additional file 4: separate graphs of both the number of episodes of lower-back and the number of lower-back and knee pain episodes receiving low-value imaging

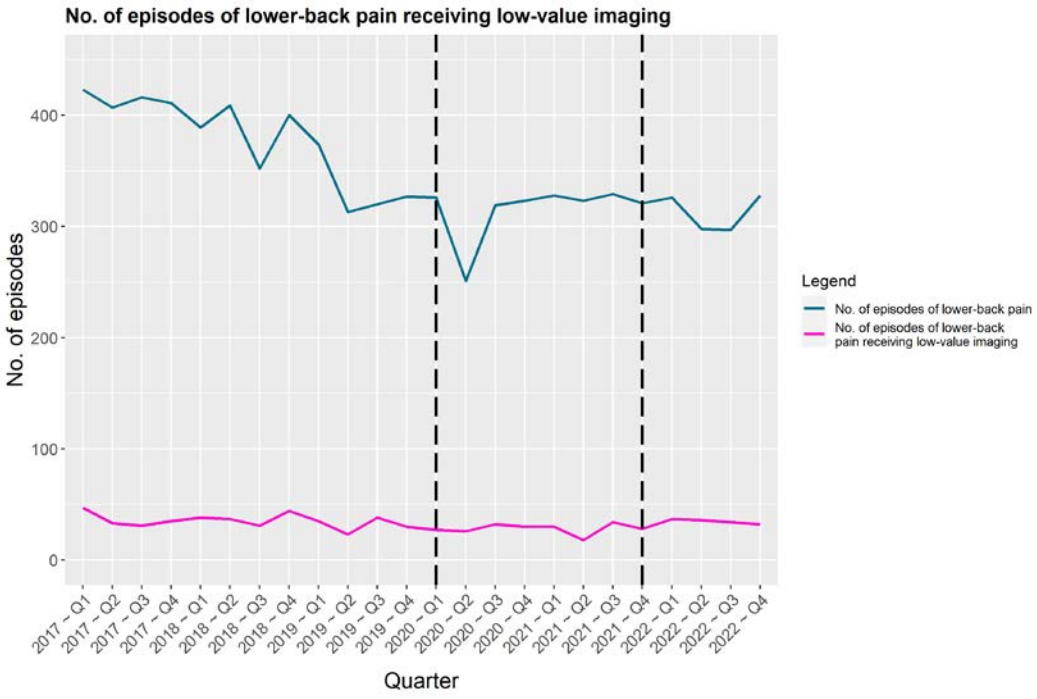


Figure S1 | Trends in both the number of episodes of lower-back pain and the number of episodes receiving low-value imaging.

Quarter	Total number of episodes of lower-back pain	Total number of episodes of lower-back pain receiving imaging
2017 Q1	423	47
2017 Q2	407	33
2017 Q3	416	31
2017 Q4	411	35
2018 Q1	389	38
2018 Q2	409	37
2018 Q3	352	31
2018 Q4	400	44
2019 Q1	374	35
2019 Q2	313	23
2019 Q3	320	38
2019 Q4	327	30
2020 Q1	326	27
2020 Q2	251	26
2020 Q3	319	32
2020 Q4	323	30
2021 Q1	328	30
2021 Q2	323	18
2021 Q3	329	34
2021 Q4	321	28
2022 Q1	326	37
2022 Q2	298	36
2022 Q3	297	34
2022 Q4	328	32

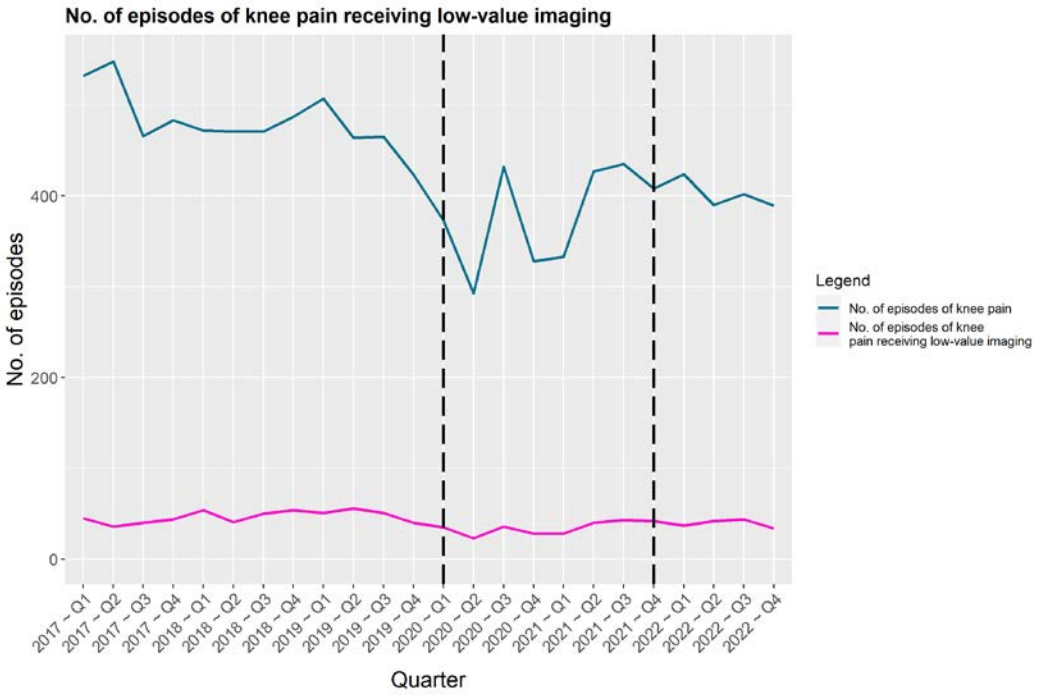


Figure S2 | Trends in both the number of episodes of knee pain and the number of episodes receiving low-value imaging.

Quarter	Total number of episodes of knee pain	Total number of episodes of knee pain receiving imaging
2017 Q1	532	45
2017 Q2	548	36
2017 Q3	466	40
2017 Q4	483	44
2018 Q1	472	54
2018 Q2	471	41
2018 Q3	471	50
2018 Q4	487	54
2019 Q1	507	51
2019 Q2	464	56
2019 Q3	465	51
2019 Q4	423	40
2020 Q1	373	35
2020 Q2	293	23
2020 Q3	432	36
2020 Q4	328	28
2021 Q1	333	28
2021 Q2	427	40
2021 Q3	435	43
2021 Q4	408	42
2022 Q1	424	37
2022 Q2	390	42
2022 Q3	402	44
2022 Q4	389	34

Additional file 5: Incidence rates of episodes and provision of low-value care for each type of care examined over the examined period corresponding to figure 1. Including rough calculations of the compared incidence rates over the entire periods (e.g. uncorrected for patient characteristics)

1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee

1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.

Year - Quarter	Total number of episodes	Total no. of episodes receiving low-value care	Total no. of patient years (per 1000)	Incidence rate episodes per 1000 patient years	Incidence rate low-value per 1000 patient years
2017 ~ Q1	778	91	2.164	359.52	42.05
2017 ~ Q2	818	71	2.241	365.02	31.68
2017 ~ Q3	735	75	2.303	319.15	32.57
2017 ~ Q4	766	82	2.344	326.79	34.98
2018 ~ Q1	753	95	2.333	322.76	40.72
2018 ~ Q2	781	80	2.396	325.96	33.39
2018 ~ Q3	748	86	2.448	305.56	35.13
2018 ~ Q4	782	99	2.466	317.11	40.15
2019 ~ Q1	779	92	2.314	336.65	39.76
2019 ~ Q2	689	82	2.357	292.32	34.79
2019 ~ Q3	711	90	2.398	296.50	37.53
2019 ~ Q4	707	74	2.420	292.15	30.58
2020 ~ Q1	643	62	2.408	267.03	25.75
2020 ~ Q2	510	51	2.414	211.27	21.13
2020 ~ Q3	719	68	2.442	294.43	27.85
2020 ~ Q4	621	60	2.441	254.40	24.58
2021 ~ Q1	613	59	2.379	257.67	24.80
2021 ~ Q2	726	61	2.402	302.25	25.40
2021 ~ Q3	756	81	2.426	311.62	33.39
2021 ~ Q4	736	78	2.418	304.38	32.26
2022 ~ Q1	738	75	2.351	313.91	31.90
2022 ~ Q2	697	80	2.365	294.71	33.83
2022 ~ Q3	714	82	2.376	300.51	34.51
2022 ~ Q4	745	68	2.355	316.35	28.87

1. The use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee.

Period	Sum of episodes	Sum of episodes receiving low-value care	Sum of patient years (per 1,000)	Uncorrected -incidence rate episodes over the entire period per 1,000 patient years	Uncorrected -incidence rate low-value over the entire period per 1,000 patient years
Pre-COVID-19	9,690	1,079	30.59	316.75	35.27
COVID-19	4,681	458	16.92	276.62	27.07
Post-COVID-19	2,894	305	9.45	306.34	32.29

2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms

2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.

Year - Quarter	Total number of prescrip-tions	Total no. of low-repeat prescriptions	Total no. of patient years (per 1,000)	Incidence rate prescriptions per 1,000 patient years	Incidence rate low-value per 1,000 patient years
2017 ~ Q1	225	18	0.22	1022.73	81.82
2017 ~ Q2	133	5	0.241	551.87	20.75
2017 ~ Q3	90	10	0.264	340.91	37.88
2017 ~ Q4	143	15	0.286	500.00	52.45
2018 ~ Q1	176	9	0.293	600.68	30.72
2018 ~ Q2	129	6	0.31	416.13	19.35
2018 ~ Q3	106	10	0.327	324.16	30.58
2018 ~ Q4	168	12	0.339	495.58	35.40
2019 ~ Q1	179	10	0.336	532.74	29.76
2019 ~ Q2	112	7	0.35	320.00	20.00
2019 ~ Q3	79	6	0.365	216.44	16.44
2019 ~ Q4	131	6	0.378	346.56	15.87
2020 ~ Q1	223	11	0.385	579.22	28.57
2020 ~ Q2	28	3	0.395	70.89	7.59
2020 ~ Q3	65	3	0.409	158.92	7.33
2020 ~ Q4	61	3	0.417	146.28	7.19
2021 ~ Q1	75	2	0.413	181.60	4.84
2021 ~ Q2	106	1	0.424	250.00	2.36
2021 ~ Q3	165	2	0.438	376.71	4.57
2021 ~ Q4	193	7	0.446	432.74	15.70
2022 ~ Q1	147	3	0.438	335.62	6.85
2022 ~ Q2	244	2	0.443	550.79	4.51
2022 ~ Q3	144	9	0.446	322.87	20.18
2022 ~ Q4	194	9	0.446	434.98	20.18

2. The prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms.

Period	Sum of episodes	Sum of episodes receiving low-value care	Sum of patient years (per 1,000)	Uncorrected -incidence rate episodes over the entire period per 1,000 patient years	Uncorrected -incidence rate low-value over the entire period per 1,000 patient years
Pre-COVID-19	1,894	125	4.09	462.63	30.53
COVID-19	693	21	2.94	235.55	7.14
Post-COVID-19	729	23	1.77	411.17	12.97

3. Repeat opioid prescriptions, without a prior visit

3. Repeat opioid prescriptions, without a prior visit

Year - Quarter	Total number of episodes	Total no. of episodes receiving low-value care	Total no. of patient years (per 1,000)	Incidence rate episodes per 1,000 patient years	Incidence rate low-value per 1,000 patient years
2017 ~ Q1	1,096	225	0.705	1554.61	319.15
2017 ~ Q2	997	224	0.718	1388.58	311.98
2017 ~ Q3	1,023	267	0.731	1399.45	365.25
2017 ~ Q4	1,049	244	0.736	1425.27	331.52
2018 ~ Q1	1,233	317	0.725	1700.69	437.24
2018 ~ Q2	1,080	298	0.736	1467.39	404.89
2018 ~ Q3	1,067	255	0.748	1426.47	340.91
2018 ~ Q4	1,151	296	0.747	1540.83	396.25
2019 ~ Q1	866	224	0.68	1273.53	329.41
2019 ~ Q2	839	226	0.688	1219.48	328.49
2019 ~ Q3	855	196	0.696	1228.45	281.61
2019 ~ Q4	787	208	0.696	1130.75	298.85
2020 ~ Q1	805	259	0.687	1171.76	377.00
2020 ~ Q2	843	268	0.683	1234.26	392.39
2020 ~ Q3	883	215	0.684	1290.94	314.33
2020 ~ Q4	893	226	0.678	1317.11	333.33
2021 ~ Q1	814	212	0.656	1240.85	323.17
2021 ~ Q2	780	166	0.656	1189.02	253.05
2021 ~ Q3	908	223	0.66	1375.76	337.88
2021 ~ Q4	952	211	0.657	1449.01	321.16
2022 ~ Q1	1,023	220	0.637	1605.97	345.37
2022 ~ Q2	983	251	0.637	1543.17	394.03
2022 ~ Q3	939	238	0.634	1481.07	375.39
2022 ~ Q4	1,122	267	0.626	1792.33	426.52

3. Repeat opioid prescriptions, without a prior visit

Period	Sum of episodes	Sum of episodes receiving low-value care	Sum of patient years (per 1,000)	Uncorrected -incidence rate episodes over the entire period per 1,000 patient years	Uncorrected -incidence rate low-value over the entire period per 1,000 patient years
Pre-COVID-19	12,848	3,239	9.29	1382.55	348.54
COVID-19	6073	1521	4.67	1299.32	325.42
Post-COVID-19	4067	976	2.53	1604.97	385.16

CHAPTER 7

Low-value chronic prescription of acid reducing medication among Dutch general practitioners: impact of a patient education intervention

Joris L.J.M. Müskens
Simone A. van Dulmen
Karin Hek
Gert P. Westert
Rudolf B. Kool

Published in BMC Primary Care, Volume 25, Article number 106, 2024

Abstract

Background: Dyspepsia is a commonly encountered clinical condition in Dutch general practice, which is often treated through the prescription of acid-reducing medication (ARM). However, recent studies indicate that the majority of chronic ARM users lack an indication for their use and that their long-term use is associated with adverse outcomes. We developed a patient-focussed educational intervention aiming to reduce low-value (chronic) use of ARM.

Methods: We conducted a randomized controlled study, and evaluated its effect on the low-value chronic prescription of ARM using data from a subset (n=26) of practices from the Nivel Primary Care Database. The intervention involved distributing an educational waiting room posters and flyers informing both patients and general practitioners (GPs) regarding the appropriate indications for prescription of an ARM for dyspepsia, which also referred to an online decision aid. The interventions' effect was evaluated through calculation of the odds ratio of a patient receiving a low-value chronic ARM prescription over the second half of 2021 and 2022 (i.e. pre-intervention vs. post-intervention).

Results: In both the control and intervention groups, the proportion of patients receiving chronic low-value ARM prescriptions slightly increased. In the control group, it decreased from 50.3% in 2021 to 49.7% in 2022, and in the intervention group it increased from 51.3% in 2021 to 53.1% in 2022. Subsequent statistical analysis revealed no significant difference in low-value chronic prescriptions between the control and intervention groups (Odds ratio: 1.11 [0.84 – 1.47] , $p>0.05$).

Conclusion: Our educational intervention did not result in a change in the low-value chronic prescription of ARM; approximately half of the patients of the intervention and control still received low-value chronic ARM prescriptions. The absence of effect might be explained by selection bias of participating practices, awareness on the topic of chronic AMR prescriptions and the relative low proportion of low-value chronic ARM prescribing in the intervention as well as the control group compared to an assessment conducted two years prior.

Introduction

Dyspepsia is one of the most commonly encountered clinical conditions in general practice, with a pooled prevalence ranging between 15 and 21% of the global population. [1-3] Dyspepsia is generally defined as a symptom complex characterised by a predominant pain or discomfort in the upper abdominal region, such as epigastric discomfort or pain, heartburn or regurgitation. [1] In the Netherlands alone, approximately 800,000 patients reporting symptoms of dyspepsia annually. [4] International assessments of the prevalence of dyspepsia reveal significant variation between countries, with rates ranging from less than 1% to as high as 57%. [1, 5, 6] As dyspepsia mostly is not caused by an identifiable disease or organic abnormalities, it is generally perceived as a harmless condition, in absence of alarm symptoms such as bleeding, anaemia, unintended weight loss, or dysphagia. [7-11]

Numerous studies have demonstrated an association between the development of dyspepsia and various lifestyle factors including diet, smoking, alcohol consumption, excessive body mass, and mental state. [9-12] Dutch guidelines for general practitioners (GPs) therefore recommend GPs to provide lifestyle advice prior to treatment with acid reducing medication (ARM), such as antacids or H2-receptor antagonist and proton-pump inhibitors (PPIs). [13, 14] However, a recent review indicated that around a quarter of the adult population worldwide uses ARM. [15] Additionally, ARM was the most frequently prescribed drug category in Dutch general practice in 2020, with over 2.2 million users. [16-19] Although short-term ARM prescriptions are an effective way to control acid-related disease, the chronic prescription of ARM is only indicated in specific situations. According to the guidelines for Dutch GPs, chronic prescriptions of ARMs should only be considered in patients with Barrett's oesophagus, Zollinger-Ellison syndrome, or in patients at high risk of gastrointestinal bleeding. [14] However, a recent study showed that around 88% of patients with a chronic ARM prescription in Dutch general practice lacked an appropriate indication, so called low-value prescription. [20] Although PPIs used to be considered effective and safe, there is growing concern regarding their long-term use as it is associated with numerous adverse effects such as vitamin deficiencies, development of multidrug resistance, decreased bone density, and enteric infections. [21-24] Moreover, the use of ARM can cover potential lifestyle risks. It therefore is necessary to reduce the (chronic) prescription of ARMs among Dutch general practitioners.

Previous studies have demonstrated the effectiveness of patient decision aids in reducing low-value treatment. Patient decision aids help patients comprehend the potential benefits and risks associated with their treatment options, empowering them to actively engage in healthcare decisions and make choices that align with their values. [25, 26] However, the effectiveness of the introduction of patient decision aids varies. Furthermore, in the context of chronic ARM provision, the existing evidence of their effectiveness is limited. Only one study by Krol et al., showed that the provision of patient information can effectively reduce low-value chronic ARM use through provision

of an educational flyer to chronic ARM users. [27] However, the educational materials used were limited to discussing the newly updated GP guidelines on dyspepsia management and did not provide information regarding potential underlying causes, associated risks and benefits of stopping ARM use, or appropriate indications. In this study, we therefore investigated the impact of an patient focused educational intervention containing these elements on the chronic prescription of ARM.

Methods

Study design, phases and setting

We conducted a randomized controlled interventional study and evaluated this using data derived from a subset of practices participating in the Nivel Primary Care Database (Nivel-PCD). The Nivel-PCD contains care data routinely collected from the electronic medical records from 529 GP practices throughout the Netherlands, representing approximately 2 million registered patients. [28] Furthermore, the database contains longitudinal information regarding patient characteristics such as age, sex, GP consultations, diagnoses, and drug prescriptions. Socioeconomic status (SES) scores (on the level of Dutch postal codes) were obtained from the Central Statistical Office (CBS). [29] Patients were assigned to one of five categories (lowest, below average, average, above average, highest) based on quintiles. Age categories were defined based on the available GP guidelines [13, 14] Diagnoses are recorded using the International Classification of Primary Care version 1 (ICPC-1). Prescriptions are recorded using the Anatomical Therapeutic Chemical classification system (ATC). This study was approved by the relevant governance bodies of the Nivel-PCD (nr. NZR00322.017) and by the Research Ethics Committee of the Radboud University Medical Centre (dossier number 2022-13579).

Intervention and recruitment

The intervention consisted of the distribution of a poster for the waiting room and flyers to be given to patients aiming to inform both patients and GPs with respect to the correct indications for treatment of dyspepsia (box 1 contains an elaborate description of the intervention materials). After signing up, practices assigned to the intervention group received a package containing 60 flyers and one waiting room poster to use during consultations. The flyer and poster provide a short description of the correct indications for treatment of dyspepsia. Additionally, both the flyers and posters contained a QR-code linking to a decision aid explaining the correct indications and causes of dyspepsia. The intervention materials are added as supplementary file 1.

Recruitment of the practices took place from July until October of 2022. Practices were recruited from the pool of practices that provide their registration data to the

Nivel. The recruitment involved sending several e-mails asking for their participation, with the promise of receiving a €20 gift card when consenting to participation. The responding practices were randomly assigned to either the intervention or control group. Throughout the recruitment process, no information regarding the exact research topic or the intervention was provided, ensuring the blind participation of the practices.

BOX 1: Intervention materials

WEB-BASED EDUCATIONAL DECISION-AID FOR PATIENTS
The educational web-based decision-aid was developed in collaboration with *thuisarts.nl* (*homedoctor.nl*), a Dutch website created by the Dutch GP association. The interactive tool provides patients with information about dyspepsia and its pathogenesis and explains treatment options as well as conservative management. The aim is to reassure patients, to give patients insight in their complaints and to learn them what they can do themselves to reduce complaints.

FLYERS AND POSTERS FOR PATIENTS
Flyers and posters were available to raise awareness about appropriate care in dyspepsia and inform patients about the available decision-aid. A QR-code led patients directly to the online tool.

Sample size calculation

Based on a z-test sample size calculation using the proportion of patients that received an inappropriate chronic ARM prescription observed in an earlier assessment in the Netherlands (88% of chronic ARMs users do not have an indication), an alpha of 0.05, power of 0.80 and an expected reduction of 10%, a minimum number of 28 GP practices (with a mean of 328 patients that are inappropriately using a chronic ARM) were required to achieve significance. [20]

Randomisation

The participating general practitioners were recruited in a blinded manner from the Nivel-PCD. Meaning that the GPs were approached by the Nivel-PCD without receiving information regarding the purpose of the study. After having consented to participation, GPs were randomly assigned to either the intervention or control group. When a GP was assigned to the intervention group, the entire practice was seen as being exposed. GPs assigned to the intervention group received the poster and flyers, to be shared with the patients suffering from dyspepsia. GPs assigned to the control group received nothing. However, it is important to note that the access to the decision aid was not limited to the GPs of the intervention group and their patients, it was freely accessible to anyone through the website Thuisarts.nl. [30]

Assessment of the low-value chronic prescription of acid reducing medication

Our assessment of the amount of ARM users was conducted using a patient-indication lens, as described by Chalmers et al. [31] Implying that all patients that were chronic ARM users were included in our denominator and all patients without indication for chronic use in our numerator. Individuals were considered chronic ARM users when they had received acid reducing medication for at least 180 days in the previous year. We defined a patient's chronic prescription as being of low-value when for at least 75% of all prescription days there was no clear indication for chronic ARM prescription present. [20] Supplementary file 2 contains an overview of the way we operationalised our assessment of low-value chronic ARM prescription. This part of the analysis was performed using STATA 16. [32]

Statistical analysis of the difference in prescribing over the two periods

To assess the differences in ARM prescriptions we compared the incidence rate of (inappropriate) chronic ARM prescriptions in the same 6 months before and after the intervention (i.e. last 6 months of 2021 and last 6 months of 2022). Our primary outcome therefore would be the odds ratio (OR) of patients receiving a low-value chronic ARM prescription between the pre- and post-intervention periods. For this purpose, we built a multilevel binomial model, with an interaction term between both the indicator of cohort (i.e. 2021 vs 2022) and an indicator indicating whether a patient was part of a practice belonging to the intervention or control group. We aimed to include random effects for both the patient and practice level when possible. However, we ended up using models only including a practice level because of the limited number of observations on the level of the patient. Generalised variance inflation factors (GVIF) were calculated to test for collinearity among the included variables before multilevel analysis was conducted (supplementary file 3). Patient age, socioeconomic status (SES) and sex were included as case-mix variables in the models, since previous research has shown they could affect the amount of care a patient requires, receives or has access to. [33-35] Patients for which either the age or socioeconomic status was unknown were excluded from the multilevel analysis, but were included in the table showing the general description of both cohorts (as presented in table 2). Following our analysis of the baseline characteristics of the included population, we were forced to exclude patients above the age of 80 from this analysis while no cases of low-value care provision were present, which would result in too little variation on the practice level. We therefore chose to exclude patients aged 80 and above from our analysis, prioritising the recognition of clustering at the practice level over the inclusion of this age group in our model. The pre-intervention period (2021) was taken as reference period. A P-value smaller or equal to 0.05 was considered statistically significant for all analyses, based on two-sided testing. Data analysis and visualisation was performed using R (version 4.4.2). [36]

Results

A total of 24 practices responded to our call for participation within the recruitment period. These 24 practices were randomly assigned to either the intervention or control group, resulting in 13 practices in the intervention group and 11 practices in the control group. To even out the number of practices in each of the groups, the two practices were randomly selected from the Nivel-PCD to be added to the control group, resulting in a total of 26 participating practices. These additional practices were selected based on the similarities in size and degree of urbanisation compared to the other practices included in our analysis. Table 1 and table 2 provide a general overview of the characteristics, and recorded number of (low-value) episodes in both the intervention and control group. The initial outcomes indicate a slight increase in chronic low-value ARM prescriptions for both the control and intervention groups. In the control group, the proportion of patients with a low-value chronic ARM decreased from 50.3% in 2021 to 49.7% in 2022, and in the intervention group, it increased from 51.3% in 2021 to 53.1% in 2022. Most patients were prescribed PPI's as subsequent analysis of the types of ARMs used over both periods revealed that the majority of patients used a PPI. In the 2021 and 2022 cohort, 99.7% and 99.3% of the patients received an PPI (ATC-codes starting with A02BC), while 2.1% and 2.4% of patients were prescribed another ARM. Furthermore, 35% of the ARM users included in the 2021 cohort were also present in the 2022 cohort. Conversely, 37% of patients included in our 2022 cohort were also present in the 2021 cohort. Our results also show that the number of prescription increases with age, however the proportion of inappropriate prescribing decreases. This can be explained by the notion that with increasing age, the number of indications for appropriate chronic ARM use also increases. Analysis of the VIF factors before performance of the multilevel analysis revealed that little or no collinearity exists among the variables included in our analysis (supplementary table 3).

Subsequent multilevel regression analysis revealed that albeit the proportions showing to have slightly increased in both the control and intervention group. no significant difference in low-value chronic ARM prescription between the two groups was observed. The odds of receiving a chronic low-value ARM prescription showed to not- significantly differ when comparing the control to the intervention group over the examined periods (Odds ratio: 1.11 [95% CI: 0.84 – 1.47], p>0.05). Table 3 contains an overview of the study outcomes after removal of the patients aged 80+, and table 4 contains the odds ratio resulting from the subsequent statistical analysis.

Table 1 | General overview of patient characteristics over both the 2021 and 2022 cohorts.

Variables	Control group		Intervention group	
	2021	2022	2021	2022
Median no. of patients per practice [25 - 75 percentile]	3,148 [2,743 – 4,007]	3,038 [2,609 – 4,194]	2,801 [2,492 – 4,079]	2,771 [2,433 – 4,055]
Average age [± SD]	40.0 [± 23.3]	40.2 [± 23.3]	42.2 [± 21.1]	42.3 [± 23.1]
Average socioeconomic status [± SD]	0.087 [± 0.23]	0.089 [± 0.23]	0.047 [± 0.21]	0.046 [± 0,21]

Table 2 | Cohort characteristics of both the control and interventions groups.

Variables	Control group				Intervention group			
	2021		2022		2021		2022	
	No. of patients with a chronic prescription	No. of patients with a low-value chronic prescription	No. of patients with a chronic prescription	No. of patients with a low-value chronic prescription	No. of patients with a chronic prescription	No. of patients with a low-value chronic prescription	No. of patients with a chronic prescription	No. of patients with a low-value chronic prescription
No. of patients with a chronic ARMs prescription	1,982	996 (50.3%)	1,894	942 (49.7%)	1,733	889 (51.3%)	1,694	899 (53.1%)
% female	54.1%	54.6%	53.3%	54.8%	62.2%	54.7%	54.8%	56.1%
No. of patients per age category								
0 – 49	229	225	229	227	155	152	165	163
50 – 59	299	291	258	255	275	271	270	266
60 – 69	456	328	438	302	452	333	435	326
70 – 79	555	152	560	158	487	133	492	144
80 +	443	0	409	0	364	0	332	0
No. of patients per SES category								
Lowest	359	178	315	164	587	335	545	324
Below average	269	139	545	241	319	158	298	160
Average	678	327	392	183	146	93	131	77
Above average	339	158	294	142	601	266	636	298
Highest	337	194	348	212	80	37	84	40

Table 3 | Overview of study outcomes. The number of chronic ARM users for each of the periods, including the proportion of these that receive a chronic low-value ARM prescription.

	2021 – Total	2021 – Low-value	% Low-value	2022 – Total	2022 – Low-value	% Low-value
Control	1,539	996	64.7%	1,485	942	63.4%
Intervention	1,369	889	64.9%	1,362	899	66.0%

Table 4 | Overview of the outcome of our analysis of the impact of our intervention on the odds of receiving an low-value chronic ARM prescription over the compared periods. The table contains both the proportions of chronic ARM users that received a low-value chronic ARM and the subsequently calculated odds ratio.

	Proportion low-value 2021 (%)	Proportion low-value 2022 (%)	Odds ratio of receiving a low-value ARM between control/intervention over 2021/2022 + [95% CI]
Control	64,7%	63,4%	1.11 [0.84 – 1.47]
Intervention	64,9%	66,0%	

Discussion

Our study shows that over the last half year of 2021 and 2022 in both the intervention and control group approximately half of the patients received low-value chronic ARM prescription. This indicates that ARM was still regularly prescribed over the investigated periods. Furthermore, no significant difference in the number of patients receiving a low-value chronic ARM prescription was observed between the control and intervention group (Odds ratio: 1.11 [0.84 – 1.47], p>0.05). Additional analysis revealed that in both the 2021 and in the 2022 cohort, the majority of patients used a PPI (ATC-codes starting with A02BC, prescribed to 99.7% and 99.3% of the patients respectively while only 2.1% and 2.4 of patients in either the 2021 and 2022 cohort were prescribed another ARM). This suggests that it is highly unlikely that the lack of an effect following our intervention cannot be ascribed due to a large proportion of patients stepping down from an PPI to antacids.

Comparison with other research

It seems that the intervention in itself did not alter the inappropriate prescription of ARM among the included GPs. This finding is not unique, however there is quite some variation in the effectiveness of similar interventions addressing low-value ARM prescribing using a patient educational tool exists.

In a study of Boster et al., the treating primary care physicians directly discussed the appropriate indications for ARM use with their patients. Using this method, they successfully reduced the patients' ARM dosage or completely stopped ARM usage in 44% of the identified ARM users within a military hospital over a 6-month period. [37] Apart from this one study, most studies regarding the reduction of ARM use rely on providing patients the tools needed for appropriate self-management of their dyspepsia. These tools included the provision of intensive support by a specialist nurse, the formulation of an action plan and an explanation of the appropriate indications as well as the benefits of decreasing or discontinuing ARM usage. However, the outcomes of these studies vary. For example, both the study by Murie et al. and the study by Coyle et al. managed to stop or reduce PPI use (by 83% and 35%, respectively) by providing patients the tools for self-management of their ARM use, such as formulating an action plan and providing information regarding appropriate ARM use. [38, 39] Conversely, the study by Dibly et al. provided similar support to ARM users, but their study did not show to change ARM use among the included patients. [40] This observation is consistent with a previous study by Batuwitage et al., which demonstrated that providing education to patients about the appropriate indications for ARM use did not lead to a significant change in ARM utilisation. [41] However, it is worth noting that none of these studies specifically focused on chronic ARM users in their intervention evaluation. As previously mentioned, only the study by Krol et al., specifically assess the impact of their intervention on chronic ARM users, and managed to reduce chronic ARM use by 24% in the intervention group compared to 7%

in the control group (24% reduction vs. 7%, respectively). [27] The difference between our study outcome and theirs can probably be that in our study the practices assigned to the control group in our study could also had access to the intervention materials, while these were freely accessible online. This could have led to exposure of the control practices to the intervention, which was not possible in the study by Krol et al., since they only actively approached the intervention practices. This difference could explain why we did not observe a difference in low-value chronic ARM prescribing between the control and intervention groups.

Analysis absence of effect

Our intervention did not lead to a significant reduction in low-value chronic ARM prescriptions between the intervention and control group. The present study does show a much lower percentage of low-value chronic ARM users compared to a previous assessment. Our earlier study, which examined chronic ARM use from 2016 to 2019, found that approximately 88% of chronic ARM users in the Netherlands lacked an indication. In the current study, this baseline was 66%. [20] Several possible reasons could explain the lower baseline for the included practices. First, since our previous assessment, a lot of (media) attention such as reports by national newspapers and an item during the eight o'clock news, has been generated on the appropriate use of ARM. Also, the publication of a report by the Dutch National Health Institute discussed the state of (appropriate) care provision for patients with dyspepsia early in 2021. This public attention might have had an effect on the prescription of ARM by GPs. [4] Second, the overarching national campaign started well before our distribution of the intervention materials among the intervention practices. Therefore, we cannot guarantee that before onset of our assessment the included practices (in both control and intervention groups) were not already affected. Third, the participating practices might already have affinity with improving the quality of care provision as they willingly joined the study unaware of the research topic or intervention. These practices might therefore already have a critical attitude towards the (chronic) prescription of ARMs, providing an explanation for the lower baseline observed in our study. Fourth, contact with the different intervention practices a few months after having distributed the materials revealed that the degree of exposure to the intervention varied amongst the intervention practices. Most GPs indicated that they were aware of the existence of the decision-aid. However, we do not know to what extent all GPs in the intervention practices have used the materials when seeing patients with dyspepsia. The fifth and final reason which could explain the absence of an effect following our intervention could be that our intervention was not sufficiently tailored to be effective. Hence, our intervention focussed on explaining the potential causes of dyspepsia and appropriate indications for ARMs use to both GPs and patients. However, as previous research indicated, the provision of low-value care is often the result of an interplay of multiple factors existing on multiple levels (e.g. the patient, healthcare provider and organizational or even medical society context). [42, 43] Additionally, it shows that the effectiveness of deimplementation strategies and interventions depend on

contextual factors, such as workplace culture or economic factors. Factors which we could not control in our intervention. Potentially, our intervention could have shown an effect if we had proactively put more emphasis on the use and implementation of the materials as well as improving knowledge of the existing guidelines. While in the current setup, our intervention heavily relies on the pro-active participation of the participating healthcare providers to improve ARM prescribing; something which has proven hard to monitor.

Strengths & Limitations

A strength of this study is that it used routinely collected administrative data containing high-quality and clinical information. This use of highly detailed data enabled us to accurately differentiation between appropriate and inappropriate prescriptions of ARM among patients. However, our study is also prone to limitations. Firstly, we were unable to reach the required number of practices to achieve significance according to our power calculation. Despite extensive efforts, we only managed to include 26 of the required 28 practices, making it challenging to draw definitive conclusions regarding the effectiveness of our intervention. Second, there were also some methodological limitations regarding our assessment of low-value chronic ARM prescription among GPs, as discussed in our previous study. [20] There is an inherent uncertainty in identifying whether a prescription is of low-value. Recommendations contain terms that do not map directly to data variables; also, diagnosis and procedure codes may not precisely identify patients for whom care is of low value. For instance, the recommendations regarding chronic ARM use lacked enough detail or required variables which are absent in the data to accurately distinguish appropriate from inappropriate prescribing. An illustrative example is the guideline stating that gastro-protection using a non-selective non-steroidal anti-inflammatory drug (NSAID) is justified if a patient is using a high dosage of a NSAID. However, information regarding the dosage of the prescribed NSAIDs was unavailable in the data used. We were also unable to identify patients suffering from chronic heartburn, as we only had access to diagnosis established within data of the years (and one year prior) included in our analysis. Patients diagnosed with heartburn outside of this period could therefore potentially be missed. More crucially, heartburn often only persists until patients take ARM (albeit via a prescription or obtained over the counter). The use of ARM often resolves the patients' symptoms, resulting in the removal of the heartburn diagnosis from their medical records, making defining chronic heartburn challenging. Third, unfortunately we are unable to monitor the number of patients that actually accessed or used the monitor following a visit to their GP. We did contact participating practices to obtain an indication of whether or not patients used the decision-aid. Unfortunately, the participating GPs indicated that they did not have insight into whether the patients actually did use the decision aid and reported that patient never mentioned its use in any of the subsequent visits. Finally, the persistent relatively high prevalence of inappropriate chronic ARM prescriptions could be attributed to the perception of ARMs as relatively harmless. ARMs are readily available over the counter at most drugstores in the Netherlands. Thus, it is likely that our

assessment still underestimates the true extent, as we could not capture all chronic ARM users in this study, particularly those using non-prescription ARMs.

Conclusion

Our educational intervention did not result in a change in the low-value chronic prescription of ARM, suggesting that (low-value) chronic prescribing ARM remains an important issue in current medical practice. Future research therefore should focus on what is needed for practices to successfully adopt the use of a patient-centred decision aid and reduce low-value chronic prescribing ARM.

References

1. Ford AC, Marwaha A, Sood R, Moayyedi P. Global prevalence of, and risk factors for, uninvestigated dyspepsia: a meta-analysis. Gut. 2015;64(7):1049-57.
2. Eusebi LH, Ratnakumaran R, Yuan Y, Solaymani-Dodaran M, Bazzoli F, Ford AC. Global prevalence of, and risk factors for, gastro-oesophageal reflux symptoms: a meta-analysis. Gut. 2018;67(3):430-40.
3. Nirwan JS, Hasan SS, Babar ZU, Conway BR, Ghori MU. Global Prevalence and Risk Factors of Gastro-oesophageal Reflux Disease (GORD): Systematic Review with Meta-analysis. Sci Rep. 2020;10(1):5814.
4. ZorginstituutNederland. Verbetersignalement Maagklachten. in Zinnige Zorg 2021.
5. Barberio B, Mahadeva S, Black CJ, Savarino EV, Ford AC. Systematic review with meta-analysis: global prevalence of uninvestigated dyspepsia according to the Rome criteria. Aliment Pharmacol Ther. 2020;52(5):762-73.
6. Mahadeva S, Goh KL. Epidemiology of functional dyspepsia: a global perspective. World J Gastroenterol. 2006;12(17):2661-6.
7. Talley NJ, Vakil N. Guidelines for the management of dyspepsia. Am J Gastroenterol. 2005;100(10):2324-37.
8. Ikenberry SO, Harrison ME, Lichtenstein D, Dominitz JA, Anderson MA, Jagannath SB, et al. The role of endoscopy in dyspepsia. Gastrointest Endosc. 2007;66(6):1071-5.
9. Feinle-Bisset C, Azpiroz F. Dietary and lifestyle factors in functional dyspepsia. Nature Reviews Gastroenterology & Hepatology. 2013;10(3):150-7.
10. Pesce M, Cargiolli M, Cassarano S, Polese B, De Conno B, Aurino L, et al. Diet and functional dyspepsia: Clinical correlates and therapeutic perspectives. World J Gastroenterol. 2020;26(5):456-65.
11. Zhang M, Hou ZK, Huang ZB, Chen XL, Liu FB. Dietary and Lifestyle Factors Related to Gastroesophageal Reflux Disease: A Systematic Review. Ther Clin Risk Manag. 2021;17:305-23.
12. Huang ZP, Wang K, Duan YH, Yang G. Correlation between lifestyle and social factors in functional dyspepsia among college freshmen. J Int Med Res. 2020;48(8):300060520939702.
13. Nederlands Huisartsen Genootschap. NHG-Standaard: Maagklachten: NHG; 2021 [Available from: https://richtlijnen.nhg.org/standaarden/maagklachten.
14. Genootschap NH. NHG behandelrichtlijn Preventie van maagcomplicaties door geneesmiddelgebruik: Nederlandse Huisartsen Genootschap; 2021 [Available from: https://richtlijnen.nhg.org/behandelrichtlijnen/preventie-van-maagcomplicaties-door-geneesmiddelgebruik.
15. Shanika LGT, Reynolds A, Pattison S, Braund R. Proton pump inhibitor use: systematic review of global trends and practices. Eur J Clin Pharmacol. 2023;79(9):1159-72.
16. Nielen M, Hek, K., Korevaar, J., Dijk, L. van, Weesie, Y. . Cijfers huisartsen - Voorgeschreven geneesmiddelen. www.nivel.nl 2021 [updated 18-11-2021. Available from: https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/voorgeschreven-geneesmiddelen-cijfers-huisartsen.

17. Nielen M, Hek, K. Jaarcijfers aandoeningen - Huisartsenregistraties www.nivel.nl 2022 [updated 07-02-2022. Available from: https://www.nivel.nl/nl/nivel-zorgregistraties-eerste-lijn/jaarcijfers-aandoeningen-huisartsenregistraties.
18. GIPdatabank. Aantal gebruikers 2017-2021 voor ATC-subgroep A02B : Middelen bij ulcus pepticum en gastro-oesofageale reflux gipdatabank.nl: Zorginstituut Nederland; 2022 [Available from: https://www.gipdatabank.nl./databank?infotype=g&label=00-totaal&tabel=B_01-basis&geg=gebr&item=A02B.
19. Lanas A. We Are Using Too Many PPIs, and We Need to Stop: A European Perspective. Am J Gastroenterol. 2016;111(8):1085-6.
20. Müskens JL, van Dulmen SA, Wiersma T, Burgers JS, Hek K, Westert GP, Kool RB. Low-value pharmaceutical care among Dutch GPs: a retrospective cohort study. Br J Gen Pract. 2022;72(718):e369-e77.
21. Jung SB, Nagaraja V, Kapur A, Eslick GD. Association between vitamin B12 deficiency and long-term use of acid-lowering agents: a systematic review and meta-analysis. Intern Med J. 2015;45(4):409-16.
22. Nochaiwong S, Ruengorn C, Awiphan R, Koyratkoson K, Chaisai C, Noppakun K, et al. The association between proton pump inhibitor use and the risk of adverse kidney outcomes: a systematic review and meta-analysis. Nephrol Dial Transplant. 2018;33(2):331-42.
23. Willems RPJ, van Dijk K, Ket JCF, Vandenbroucke-Grauls C. Evaluation of the Association Between Gastric Acid Suppression and Risk of Intestinal Colonization With Multidrug-Resistant Microorganisms: A Systematic Review and Meta-analysis. JAMA Intern Med. 2020;180(4):561-71.
24. Trifan A, Stanciu C, Girleanu I, Stoica OC, Singeap AM, Maxim R, et al. Proton pump inhibitors therapy and risk of Clostridium difficile infection: Systematic review and meta-analysis. World J Gastroenterol. 2017;23(35):6500-15.
25. Stacey D, Légaré F, Lewis K, Barry MJ, Bennett CL, Eden KB, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2017;4(4):Cd001431.
26. Ankolekar A, Dekker A, Fijten R, Berlanga A. The Benefits and Challenges of Using Patient Decision Aids to Support Shared Decision Making in Health Care. JCO Clinical Cancer Informatics. 2018(2):1-10.
27. Krol N, Wensing M, Haaijer-Ruskamp F, Muris JW, Numans ME, Schattenberg G, et al. Patient-directed strategy to reduce prescribing for patients with dyspepsia in general practice: a randomized trial. Aliment Pharmacol Ther. 2004;19(8):917-22.
28. Hendriksen J, Hooiveld, M., Nielen M., Hek, K Methoden vaststellen cijfers Nivel Zorgregistraties Eerste Lijn 2021 [Available from: https://www.nivel.nl/nl/zorgregistraties-eerste-lijn/methoden.
29. (CBS) CSO. Sociaal-economische status per postcode 2019. In: (CBS) CSO, editor. 2022.
30. Thuisarts. Decision-aid dyspepsia 2023 [Available from: https://www.thuisarts.nl/keuzehulp/maagklachten.
31. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.

32. LLC S. Stata Statistical Software: Release 16. In: LLC S, editor. 16 ed. College Station, TX: StataCorp LLC.: StataCorp LLC; 2019. p. StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.
33. Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147-56.
34. David JL, Kaplan HB. Gender, social roles and health care utilization. Applied Behavioral Science Review. 1995;3(1):39-64.
35. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract. 2000;49(2):147-52.
36. R Core Team. R: A language and environment for statistical computing. v3.6.3 ed. Vienna, Austria: R foundation for Statistical Computing; 2019.
37. Boster J, Lowry LE, Bezzant ML, Kuiper B, Surry L. Reducing the Inappropriate Use of Proton Pump Inhibitors in an Internal Medicine Residency Clinic. Cureus. 2020;12(1):e6609.
38. Murie J, Allen J, Simmonds R, de Wet C. Glad you brought it up: a patient-centred programme to reduce proton-pump inhibitor prescribing in general medical practice. Qual Prim Care. 2012;20(2):141-8.
39. Coyle C, Symonds R, Allan J, Dawson S, Russell S, Smith A, et al. Sustained proton pump inhibitor deprescribing among dyspeptic patients in general practice: a return to self-management through a programme of education and alginate rescue therapy. A prospective interventional study. BJGP Open. 2019;3(3).
40. Dibley LB, Norton C, Jones R. Non-pharmacological intervention for gastro-oesophageal reflux disease in primary care. Br J Gen Pract. 2010;60(581):e459-65.
41. Batuwitage BT, Kingham JG, Morgan NE, Bartlett RL. Inappropriate prescribing of proton pump inhibitors in primary care. Postgrad Med J. 2007;83(975):66-8.
42. Heus P, van Dulmen SA, Weenink JW, Naaktgeboren CA, Takada T, Verkerk EW, et al. What are Effective Strategies to Reduce Low-Value Care? An Analysis of 121 Randomized Deimplementation Studies. J Healthc Qual. 2023;45(5):261-71.
43. Ingvarsson S, Hasson H, von Thiele Schwarz U, Nilsen P, Powell BJ, Lindberg C, Augustsson H. Strategies for de-implementation of low-value care—a scoping review. Implementation Science. 2022;17(1):73.

Supplementary materials

Supplementary file 1: Overview of the campaign materials

Figure 1 | Front view of the folder used in our intervention.

Maagklachten

Maagklachten komen veel voor. U kunt last hebben van pijn boven in de buik, brandend maagzuur of opboeren. De oorzaak van de klachten is niet altijd duidelijk. De meeste mensen met maagklachten hebben geen ziekte en geen beschadigde maag. Meestal gaan de klachten vanzelf over en zijn medicijnen niet nodig.

Wat kunt u doen?

Gezond leven en voedingsadviezen kunnen maagklachten verminderen. Dit is belangrijk bij alle maagklachten. Ook als u maagmedicijnen gebruikt of gaat gebruiken. De keuzehulp van Thuisarts.nl kan u helpen.

Doe de keuzehulp!

De keuzehulp:

* geeft informatie over maagklachten
* legt uit hoe de maag werkt
* geeft informatie of onderzoeken nodig zijn
* geeft u persoonlijk advies.

Let op: De keuzehulp is niet geschikt voor kinderen, als u zwanger bent of een maagverkleining heeft gehad.

Bij ernstige klachten belt u direct de huisarts of huisartsenpost.

Voor toegang tot de keuzehulp:
Scan de QR-code met de camera van uw telefoon of ga naar **www.thuisarts.nl/keuzehulp/maagklachten**

Meer informatie over maagklachten leest u op **www.thuisarts.nl**

Doen of laten? THUISARTS.NL

Figure 2 | Back view of the folder used in our intervention.

Figure 3 | Example of the poster used in the intervention.

Table 13 | ATC and ICPC codes used to define the study populations for each of the examined recommendations.

Recommendation	Additional information	ICPC codes relevant for the analysis	ATC codes relevant for the analysis
Do not chronically prescribe or continue acid-reducing medication (ARM), without proper indication. • Numerator: no. of patients with a chronic prescription of ARM. • Denominator: no. of patients with no clear indication for their chronic ARM prescription. • A more elaborate description of the different conditions in which ARMs are indicated is presented below this table.	- We defined chronic acid-reducing medication users as individuals that received an acid-reducing medication for at least 180 days in the previous year. - We were unable to identify chronic ARM users suffering from chronic heartburn. - We were unable to identify patients using high-dose high dose non-selective NSAID, while no clear definition of high-dose was present in the guideline nor did we have information regarding doses of prescriptions.	D85: Duodenal ulcer D86: Peptic ulcer other D03: Heartburn L88: Rheumatoid / seropositive arthritis K77: Hearth failure T90: Diabetes non-insulin dependent	**Co-medications associated with increased risk of stomach complications:** ○ Coumarin derivative (B01AA) ○ Direct oral anticoagulants (B01AF02, B01AE07, B01AF03, B01AF01) ○ P2Y12-inhibitors (B01AC04, B01AC22, B01AC24) ○ Acetylsalicylic acid derivatives: Acetylsalicylic acid (Aspirin): A01AD05, B01AC06, B01AC56, C10BX01, C10BX02, C10BX04, C10BX05, M01BA03, N02BA01, N02BA51, N02BA71. ○ Systemic glucocorticoids (H02AB) ○ Selective serotonin reuptake inhibitors (SSRI) (N06AB) ○ Venlafaxine (N06AX16) ○ Duloxetine (N06AX21) ○ Trazodone (N06AX05) ○ Spironolactone (C03DA01) **Acid-reducing medication:** A02BA: H2-receptor antagonists A02BB: Prostaglandins A02BC: Proton pump inhibitors M01AB: Acetic acid derivatives and related substances M01AE52: Naproxen and misoprostol

More elaborate description of the ARM recommendation

Do not chronically prescribe or continue acid-reducing medication (ARM), without proper indication. ARM prescriptions are indicated in the following cases according to the guideline:

- Gastro-protection with an proton pump inhibitor (PPI) in case of a non-selective nonsteroidal anti-inflammatory drug (NSAID)
  ii. Age of 70 years or older;
  iii. Presence of an Ulcus Duodeni (D85) or Ulcus pepticum (D86) in their medical history, irrespective of their age.
  iv. When two or more of the following factors are applicable (the risk of complications increases with increasing number of factors present):
    - Age between 60 and 70 years.
    - Severe disabling rheumatoid arteritis (L88), Hearth failure (K77) or diabetes (T90).
    - Use of high dose non-selective NSAID
    - Use of comedication which increase the risk of stomach complications.

- Gastro-protection with an proton pump inhibitor (PPI) in case of a Acetylsalicylic acid derivative as platelet aggregation inhibitor and in absence of a non-selective NSAID is indicated in case of:

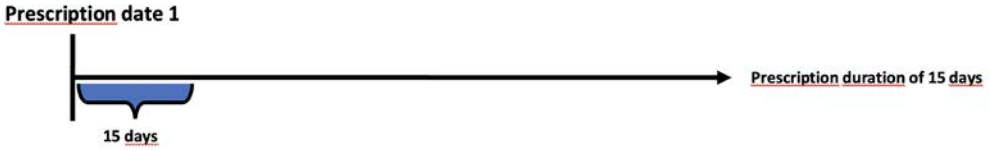
i. Age 80 years or older.
ii. Age 70 or older combined with use of comedication which increases the risk of stomach complications (except Acetylsalicylic acid derivatives).
iii. Age of 60 or older combined with the presence of an Ulcus Duodeni (D85) or Ulcus pepticum (D86) in their medical history.

- Patients suffering from chronic heartburn, which do not sufficiently benefit from alternative acid-reducing medication (or in which these have not been tried).
  i. Heartburn (D03); we could not identify patients with chronic heartburn because we only received data from the requested data period and therefore only could identify patients which received a diagnosis of chronic heartburn within the 4 years examined.

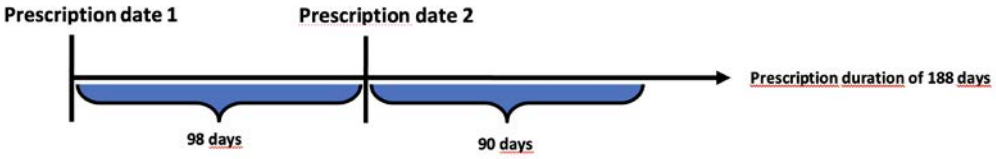
Calculation of prescription duration

Within Nivel-PCD only the dates on which the general practitioner prescribed the medication in question is recorded. The database does not contain the end date of a certain prescription. Therefore, in order to be able to define chronic use, we used the following method to calculate prescription duration :

When patients only have had a single prescription, we assumed these patients only received a 'start prescription'. In the Netherlands, a start prescription has a duration of 15 days and since there is only one prescription registered, we assumed that these patients only received this single prescription and therefore only received acid-reducing medication for the duration of 15 days.

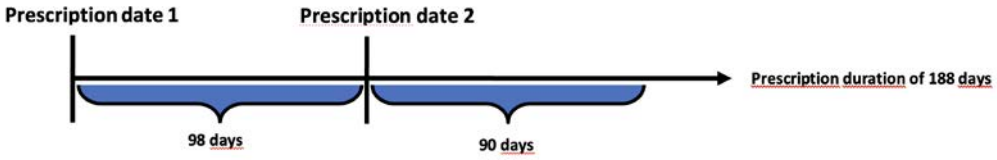


However, when a patient has received more than one prescription we have made the assumption that in between both dates the patient used the prescription continuously when the in between time is less or equal to 180 days. We chose to use 180 days since most prescriptions have a duration of 90 days, but medication is not always picked up after exactly 90 days. Furthermore, to assure that we also included patients that use their medication every other day instead of daily or patients which use their medication as needed we chose to double the prescription duration (e.g. 2 x 90 days) in order for those patients to be included. An additional 90 days are added to the latest prescription date of the two prescriptions, while a regular refill prescription has a duration of 90 days and the latest date of the two prescription dates us also the start date of the second prescription.



The same principle applies to a patient whom has 3 prescriptions, but who's prescriptions are less than 180 days apart. The duration between prescription 1 and 2 and the duration between prescription 2 and 3 are added up, and an additional 90 days are added for the duration of the third prescription.

In case a patient has more than 180 days between consecutive prescriptions, the calculate will be performed in the following manner:



When two prescriptions are more than 180 days apart, then a new prescription will commence which in turn will be treated according to the same rules as described above.

Supplementary file 3: Overview calculated generalized variance inflation factors (GVIF)

Table 1 | Generalized Variation Inflation Factors (GVIF) calculated for the model made.

	GVIF	Degrees of freedom	GVIF^(1/(2*Df))
Studygroup (control/intervention)	1.715815	1	1.309891
Cohort (Cohort2021/Cohort2022)	1.913168	1	1.383173
Gender	1.010374	1	1.005174
Age category	1.012668	3	1.002100
Socioeconomic status category	1.024646	1	1.012248
Interaction Studygroup*Cohort	2.632136	1	1.622386

CHAPTER 8

General discussion

In this thesis we aimed to gain insight into the presence of low-value care in the Dutch healthcare system and explore the methods used in its assessment. Obtaining accurate insight into both its presence and variation is essential to create a sense of urgency and a foundation to address it. We therefore conducted several assessments of the extent of low-value care in the Dutch healthcare system using routinely collected data. In this chapter, the outcomes of the performed assessments will be briefly summarised, after which the current knowledge regarding the presence of low-value care in the Netherlands will be discussed. Thereafter, we will present prerequisites for the performance and clear interpretation of low-value care assessments, based on the lessons learned from the assessments part of this thesis. Lastly, we will discuss the implications of this thesis and end with concluding remarks.

Main findings

In chapter two we examined the underlying reasons why assessments of similar low-value diagnostic tests yielded vastly different results. We performed a systematic review, in which we included studies that performed assessments regarding low-value diagnostic testing originating from countries part of the Organisation for Economic Co-operation and Development (OECD). Our results showed that differences in included population, definitions of low-value care (e.g. use of guidelines vs. expert opinion) and methodology were potential reasons underlying the differences observed between assessment outcomes. Especially the use of different methods, particularly the type of "*assessment lens*", and the type of data used in an assessment showed to greatly impact the outcome of an assessment. This suggests that these factors should (preferably) be well regarded when designing an assessment or in the interpretation of its outcomes.

In chapter 3 to 6 we put the lessons learned from chapter 2 into practice. We performed several assessments using different data sources, among both general practitioners and hospital clinicians. Chapter 3 describes our assessment of three types of low-value pharmaceutical care among Dutch general practitioners (GPs) between 2016 and 2019. This assessment, using registration data from around 300 Dutch GP practices, showed that the magnitude of and variation in low-value care provision largely differs between clinical scenarios and GP practices over the examined period. Between 2016 and 2019, between 53%-61% of patients received an inappropriate antibiotics prescription for their conjunctivitis episode, around 3% of patients with lower back pain (LBP) received an inappropriate benzodiazepine prescription and 88% inappropriately received an chronic prescription of acid reducing medication (ARMs). Chapter 4 describes our examination of (potentially low-value) vitamin B12- and D-testing among Dutch GPs using claims data obtained from one of the four large healthcare insurers of the Netherlands. The results of this study show that between 2015 and 2019, the number of vitamin B12 tests increased by 98.1%, and the number of vitamin D tests by 112.0%. Furthermore, the proportion of patients per practice that received a vitamin B12 test increased from

4.8% to 8.4%, and the proportion of patients with a vitamin D test from 4.7% to 9.1%. The observed increase in both the number of determinations and proportion of patients receiving one, could be indicative of increased low-value vitamin testing. Because no substantial changes in indications for vitamin B12- or D-testing have been made in the GP guidelines over the examined time period. However, due to the limitations of the claims data, we were unable to reliably assess the appropriateness of each determination. Therefore, further examination of low-value vitamin B12- and D-testing among GPs is warranted. Chapter 5 describes our examination of low-value vitamin B12- and D-testing among hospital clinicians. For this examination we were able to use more rich clinical data obtained from hospitals (Dutch National Basic Hospital Care Registration), thereby enabling us to provide an indication of appropriateness of each test. Our examination showed that at least 79% of vitamin B12- and 82% of D-determinations were ordered inappropriately over the years examined. Furthermore, our examination revealed that a wide range of associated diagnoses not warranting a determination were associated to them. The observed proportion of low-value vitamin testing combined with the wide range of associated diagnose codes provide all information necessary to start a discussion regarding the suitability of either the existing guidelines, or the use of vitamin B12- and D-determinations among clinicians. In chapter 6 we examined the effect of the COVID-19 pandemic on the provision of three types of low-value care by GPs. The following types of low-value care were selected in collaboration with the participating GPs: the use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee, the prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms and repeat opioid prescriptions, without a prior GP visit. During the COVID-19 restrictions period, the incidence rates (IR) of episodes related to back or knee pain decreased by 12%, otitis media acuta (OMA) episodes by 54% and opioid prescription rate by 13%. However, only the IR of OMA prescriptions remained significantly lower (22%) during the post-restrictions period. The IRs related to the provision of low-value care also showed to change over the different periods. The IR of imaging for back or knee pain and low-value prescription of antibiotics for OMA both decreased significantly during the COVID-19 restrictions period (by 21% and 78%), but only the low-value prescription rate of antibiotics for OMA remained significantly lower (by 63%) during the post-restrictions period. The IR of inappropriately repeated opioid prescriptions remained unchanged over all three periods. These findings indicate that both the number of episodes and low-value care practices are differentially affected by the COVID-19 restrictions. Suggesting that even if the provision of low-value care is reduced, maintaining its reduction could prove challenging and will not automatically occur.

Chapter 7 provides a description of the evaluation of the campaign regarding the reduction of inappropriate chronic use of acid reducing medication (ARM) among Dutch GPs. Following our findings from chapter 3 that around 88% of chronic ARM users did not have an indication, we started a campaign aiming to improve its appropriate use. We provided a group of GPs with educational flyers and a poster referring to a decision aid regarding the appropriate use of acid reducing medication. Our analysis of the odds ratio

associated with receiving a low-value chronic ARM prescription between the two periods revealed no significant change in chronic low-value ARM use between the control and intervention groups (Odds ratio: 0.95 [95% CI: 0.72 – 1.25], p>0.05). However, while this study did not find a significant reduction in low-value chronic ARM prescriptions, it does suggest the number of chronic ARM users has decreased. In the present study, we found that approximately half of the chronic ARM users lacked an indication, compared to the 88% identified in a previous assessment.

Low-value care in the Netherlands

The last years, low-value care has received increasing attention within the Netherlands, resulting in some assessments being conducted. These studies examine a wide variety of different healthcare services and show great variation with respect to the reported outcomes and the assessment methods used. For example, in a study by Kool et al., from 2020, the presence of three types of low-value care were examined among Dutch general practitioners using claims data from one of the four major healthcare insurers. [1] This study reported that approximately 8% of the identified patients received a low-value doppler or plethysmography procedure for the diagnosis of varices, and around 0.4% received low-value screening for colorectal cancer. This study also provided an indication that approximately 1 in 12 GPs ordered low-value x-ray once a week for lower back pain. Another study from 2020, by Laan et al., indirectly showed that a large proportion of patients within seven hospitals received inappropriate peripheral intravenous or urinary catheter using data obtained through chart review [2] Based on the indications described in both national and international guidelines, they were able to show that 22.0% of patients lacked an appropriate indication for their peripheral intravenous and 32.4% for an urinary catheter during the baseline period. A study by Koggel. et al from 2022 showed that around 66% of patients that started a proton pump inhibitor (PPI) lacked an appropriate indication, using clinical registration data. [3] Although only a few examples have been presented here, these studies generally provide highly detailed insights into the provision of specific low-value services and provide detailed descriptions of the underlying assessment methods used. This often includes elaborate descriptions of both the used (data) definitions of the types of low-value care examined. Unfortunately, the included samples are often relatively small and limited to one or a few hospitals, practices or regions of the Netherlands.

Apart from contributions published in peer-reviewed journals to unrevealing the provision of low-value care in the Netherlands, there are several reports of national low-value care provision by Dutch public institutes and government bodies, such as the National Health Care institute or the Netherlands Institute for Health Services Research (NIVEL). These reports often vary in quality regarding the underlying method, but do still provide valuable insights regarding the current status of healthcare. The NIVEL, for example, has examined the provision of several types of low-value care among Dutch

GPs, such as the appropriate use of antibiotics or tympanostomy tubes in children with OMA or appropriate treatment of osteoarthritis. [4] These reports showed that the amount of inappropriate care is decreasingly being provided, but still requires attention. [5-7] Another report from the Nivel examined the proportion of patients with knee and hip osteoarthritis (KHOA) that received joint replacement surgery. International guidelines published from 2012 onwards recommend to only consider joint replacement surgery for KHOA when maximal conservative treatments rendered insufficient results. Conservative treatments include education, (lifestyle) advice, pain medication, intra-articular injections and physio/exercise therapy (PET). The report showed that between 2013 and 2019, both the proportion of patients receiving joint replacement surgery for knee and hip osteoarthritis decreased while the use of conservative treatment had increased following a change in health insurance coverage. [5] This study thereby indirectly assesses the volume of low-value care, while simultaneously showing the effect of implementation of policy measures. Additionally, governmental agencies, such as The National Health Care Institute, also regularly publish reports discussing the state of healthcare provision for specific topics. These reports sometimes also include assessments of low-value care in for example patients with asthma, lower respiratory tract infections, chronic obstructive pulmonary disease, urinary tract infections and osteoporosis. [8-12] These assessments, often conducted on a larger scale, confirm the notion that the low-value care provision varies greatly between different types of care examined and different providers.

Overall, we can conclude that, to date, some assessments of low-value care have been conducted in the Netherlands. However, insight into low-value care provision in the Netherlands remains limited, which is comparable to many other countries. Worldwide, countries have only scratched the surface with respect to gaining insight into the total magnitude of the problem of low-value care. Additionally, most existing assessments lack a detailed description of the methods and definitions used to conduct their assessment. Thereby making it difficult to judge their quality, accuracy and usefulness.

Conceptual roadmap for the assessment of low-value care

The large variation in both the methods used and reported outcomes might partially be explained by the absence of a clear and unambiguous method of assessing low-value care. The review as shown in chapter 2 revealed that differences in both methods (assessment lenses) and definitions used in an assessment can greatly impact their outcomes. Unfortunately, as previously mentioned in the introduction, to date only one framework by Miller et al. regarding the assessment of low-value care exists. [13] This framework, however, appears to be unsuitable for guiding the assessment of low-value are. The main reasons for this are, first, that this framework has been developed to assess the total expenditure of low-value care within health systems, states or on the national level, and not the volume of low-value care per se. While it does require to assess the volume of low-value care, its focus is on identifying the total expenditure on low-value care rather than

to facilitating the identification of the proportion of services to be of low-value. Second, it proposes that the identified proportion of low-value treatment (e.g. expenditure) among the high-expenditure services should be considered as a predictor of the proportion of low-value care among low-expenditure services. However, the provision of low-value care shows to large variation between different types of low-value procedures, as showed in the different studies in this thesis. For example, both our review and analysis of three types of low-value care among Dutch GPs (chapter 3) revealed that both the utilisation of low-value care and its variation largely differs between types of care and healthcare providers. We, therefore, argue that due to the variability in low-value care provision among procedures, one should never use one assessment as a predictor for others. We therefore do not expect the methods as described by Miller et al., will yield a reliable estimate of low-value care provision.

The existence of large variation in healthcare utilisation is not a novel finding and is already widely known and recognized. In 2018, Westert et al. described a cycle which explains which steps should be taken to address and reduce unwarranted variation. [14] Even though this cycle does not primarily focus on addressing low-value care provision, it does have connection to it while the presence of large variation in healthcare utilisation often (indirectly) indicates the provision of low-value care. The described *Value Improvement Cycle (VIC)*, especially shows large similarities to the process of the assessment of low-value care. Based on the similarities of the VIC and the lessons from the studies described in this thesis, we have formulated a roadmap for the design and performance of a successful and accurate assessment of low-value care in the next paragraphs. Figure 1 depicts the steps considered essential for the assessment of low-value care based on the lessons learned in this thesis. However, we would like to stress that these steps should still be validated and tested before being viewed as a valid method of assessing low-value care.

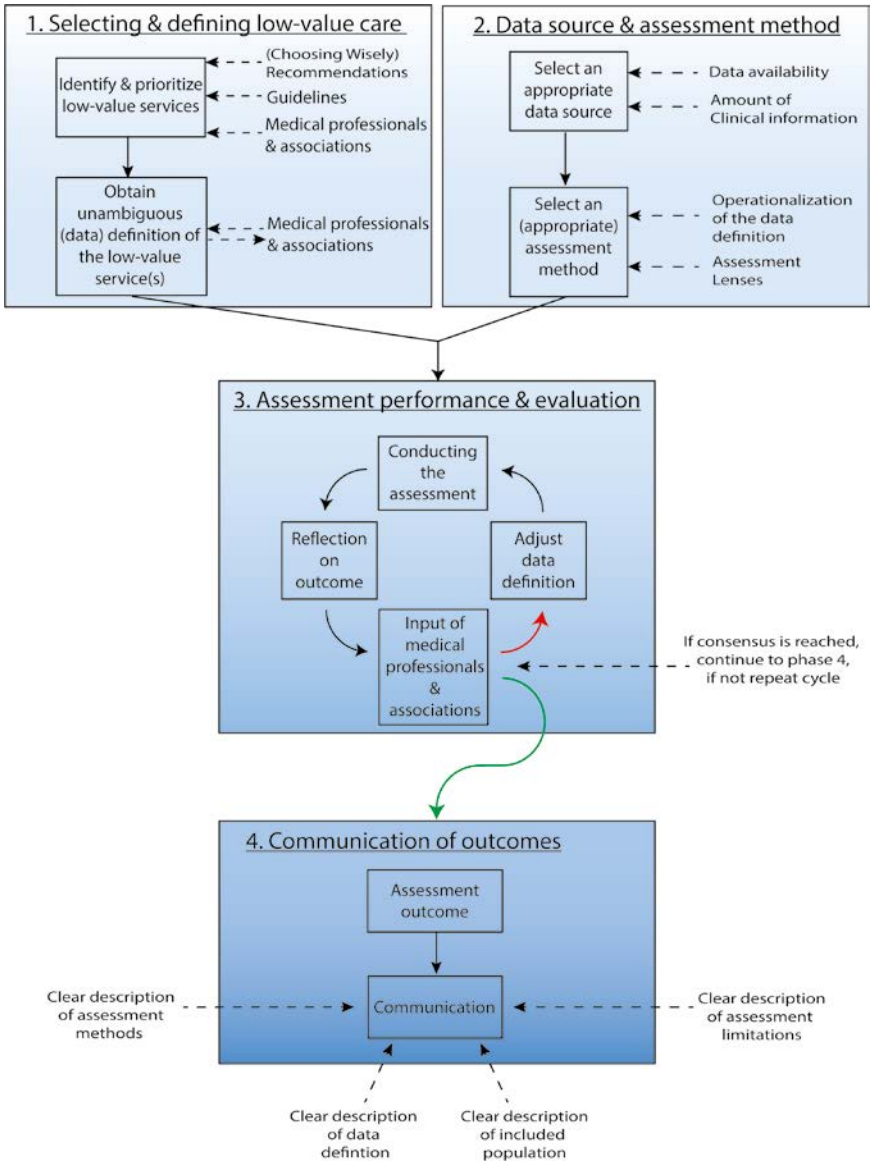


Figure 1 | schematic depiction of the phases of the assessment of low-value care.

Phase 1: Selecting types of low-value care and obtaining an unambiguous definition

The first step in the assessment of low-value care is selecting the type(s) of care to be examined. For the selection of the types of low-value care to be examined, most assessments rely on either guidelines published by professional associations [1, 15-18], Choosing Wisely (CW) recommendations [1, 19-24] or a multidisciplinary iterative process with the involved (medical) stakeholders. [25] However, whichever method one chooses to apply, it is crucial to involve healthcare professionals or their associations when making your selection. Healthcare professionals will have insight into whether or not the healthcare services provided, the appropriate indications and whether it is conceived as a problem in practice. Additionally, because of their close interactions with patients and their knowledge of (low-value) treatments, they can provide an indication with respect to which types of low-value care should be prioritized. he involvement of healthcare professionals early in the assessment process also provides the opportunity to probe whether or not the planned assessment is conceived as being feasible or relevant. In selecting the type of low-value care it is also essential to assess the evidence behind recommendations labelling certain care as low-value. Especially since for a large proportion of available (low-value) care procedures are backed up by little to no evidence. [25, 26] Selecting a type of low-value care backed up by a large amount of evidence will greatly improve the validity of your assessment and the ability to build a clear and unambiguous definition.

After having selected the type of low-value care to be examined, it is crucial to build a correct, complete and unambiguous data definition for the low-value service(s) to be examined. All relevant indications which justify the use of the examined service and all relevant exclusions should be included in the final draft. Often guidelines and CW recommendations form the basis of data definitions used in the assessment of low-value care. However, as argued before, both of these sources are not exhaustive with respect to the amount of information they contain and provide. Therefore, it is wise to consult the relevant healthcare professionals and associations with respect to the completeness of the data definition.

Phase 2: Selecting a suitable data source and an appropriate assessment method

After having obtained a suitable data definition it is important to select an appropriate data source for a valid assessment of low-value care. Especially since not all data sources contain the needed (clinical) information, to distinguish appropriate from inappropriate care. [27] Additionally, not all information is also easily accessible or even (publicly) available. It is therefore important to consider the limitations of each database before making your selection. While the chosen data source will also greatly impact the amount of (clinical) detail that can be included in the data definition built in phase one (and while both phase 1 and 2 have start rather simultaneously).

Besides selecting a suitable data source, it is also important to provide a clear description of the methods used in the assessment. Providing a clear description of the exact analyses to be performed, the outcome variables, the included population and procedures all aid in the correct interpretation of the assessment results. For example, if a multilevel analysis will be used, the type and which levels or confounders that will be included should be specified. Additionally, it should also be clear how you eventually will operationalized each of the types of low-value care to be examined. For example, which Anatomical Therapeutic Chemical (ATC), International Classification of Primary Care (ICPC) or International Classification of Disease (ICD) or Diagnosis Treatment Combination (DTC) codes will be used to identify the relevant patient groups, and how will these be coupled to the individual patients. [28-31] Second it also has to be clear which type of assessment lens will be applied (patient-indication, patient-population or service lens), as described by Chalmers et al., in 2017. [32] While different assessment lenses use different denominators, and as previously shown these can affect the amount of low-value care that could be detected. Clear understanding of the denominator that will be used is essential to put assessment outcomes in perspective, thereby enabling the correct interpretation.

Phase 3: Performance and evaluation of assessment outcomes

After having carefully defined your definition(s), operationalization and assessment methods, phase 3 relates to the performance of the assessment. After having performed your assessment and obtained the results, it is of prime importance to carefully evaluate them. Before publishing your results, it is recommended to consult relevant healthcare professionals and involved in building your data definition to provide feedback on your results. These feedback rounds can provide some explanations or insights with respect to the observed outcomes or methods and the interpretation. Which in turn could indicate any limitations or lacks in your data definition which previously had been overlooked or passed through unnoticed. Therefore, these feedback moments provide an excellent opportunity to adjust your data definition and assessment methods, or provide valuable input for the next phase of the assessment. Thereby preventing any codes or insights from being missed, and eventually ensuring the validity of your assessment outcomes.

Phase 4: Communication regarding assessment outcomes

Following the iterative process of validating the assessment outcomes with the involved healthcare professionals, it is time to communicate the outcomes. The most important thing to keep in mind is to ensure there is sufficient information for the correct interpretation of your findings. Both the assumptions made by the research team, as well as the limitations of your assessment should be clearly communicated in the light of your findings. Especially since inherent uncertainty exists with respect to distinguishing a low-value treatment in the assessment of low-value care using database data. Which partly can be explained by recommendations of low-value care often containing terms that do not map directly to data variables, diagnosis and procedure codes not precisely

identifying patients for whom care is of low value or the existence of multiple codes for the registration of similar diagnosis. Transparency with respect to both the measuring methods and the used data definition are essential for the correct interpretation of the assessment outcomes.

The importance of accurate assessments of low-value care

In a utopian world, low-value care would be absent from the entire healthcare system. However, in reality entirely removing low-value care is considered to be impossible and also should not be regarded as the goal. Most people see the problem of low-value care as something that can easily be fixed; e.g. physicians should just stop providing care that has no net benefit, but they do not fully grasp the complexity of low-value care. Hence, almost no type of care is considered of low-value in all clinical scenarios. And although assessments of low-value care are a way to gain insight into the presence of low-value care, their outcomes are also restricted with respect to their validity.

Despite their shortcomings, assessments of low-value care are still considered an important first step in the reduction of low-value care. [33] The reason for the importance of assessments of low-value care is twofold. First, although physicians are generally aware that they provide low-value care to their patients, they often do not know to what extent and what the impact of its provision is on the healthcare system. [34, 35] By conducting assessments we provide healthcare providers with some general indication with respect to the extent of low-value care provision. This provides an opportunity for healthcare professionals and policy makers to judge whether there is room for improvement or not. The second reason is that these assessments can facilitate discussions on how to improve the examined services. This is particularly valuable as they often offer insights into the factors associated with the provision of low-value care, which can then be used to address it. In contrast, assessments aimed at providing insight into the overall presence of low-value care within an entire health system or country do not yield these insights; instead, they may create incorrect expectations. As previously mentioned, estimates of the total volume of low-value care are often based on extrapolations from assessments of individual low-value services. However, previous research has demonstrated significant variations in the provision of low-value care across clinical practices, healthcare providers, health systems, and countries [1, 16, 17, 19-22, 36-38] Therefore, it's expected that estimates of the total magnitude of low-value care based on assessments of a subset of clinical practices, such as those by Shrank et al. [39] and Schwartz et al. [19] , may yield inaccurate results. More importantly, such assessments do not contribute to its reduction because they do not provide actionable insights or levers for addressing it.

After the assessment of low-value care it is essential that some party, either government, professional association or healthcare providers themselves undertake action to address the provision of low-value care. Unfortunately addressing the provision of low-value

care is highly intricate, and no single quick and universal fix exists. [40] The main reason for this is that low-value care provision often is affected by a multitude of drivers. [34, 41, 42] Addressing the problem of low-value care therefore requires a tailor-made de-implementation strategy, targeting multiple actors and facets in order to be successful. [43-45] The reduction of low-value care often requires substantial system and culture changes, even in case of the seemingly easy targets (such as low-value screening for prostate cancer). [43, 46-48] However, by starting your de-implementation trajectory with an accurate assessment, provides you with the solid foundation (and potentially the support) needed to further develop your de-implementation strategy.

Implications for practice

Based on our studies we see two main requirements which could contribute to securing a successful and effective process of assessing low-value care. First, it is important to improve both the quality and validity of the data. Hence, as previously discussed, not all data sources contain equal amounts of clinical information, and are therefore not equally suited for the assessment of low-value care. [27] Because of this, we should consider establishing which type of data should be used for which purpose. As previously discussed, claims data lack the clinical information required to distinguish appropriate from inappropriate care. These data therefore should preferably not be used with the purpose to accurately trying to assess the presence of low-value care. [13, 33] We therefore suggest if that in the case of a detailed assessment, one should always opt for the use of (clinical or administrative) registration, and not for claims, data if these are available. However, in case of explorative studies, claims data are a great option for a first scan to identify potential areas of low-value care provision (e.g. examining variations in healthcare utilisation between healthcare providers).

Apart from requiring data containing sufficient clinical detail, the validity of the available information should be of a higher standard. Currently, the available databases are prone to (some extent of) misregistration and are often incomplete. [33] It is therefore of importance to improve the quality of data registries, and improving their ease of use and general functionality of the registration programs and methods used. Both the application of (external) validation of registration databases and increased consensus regarding the registration of diseases or treatments are important factors which facilitate this improvement. [49-53] By improving data quality and standardization of the registration methods, we believe both the quality and validity of assessments will also increase. Additionally, both efforts could also aid in reducing the administrative burden healthcare professionals currently face. [54-58]

Second, the willingness among healthcare providers in hospitals, general practitioners and professional associations with respect to the reduction of low-value care and changing practice is important and should be nurtured. Throughout the process of

assessing low-value care, the input and time of healthcare professionals are essential. These professionals play a vital role in both obtaining a correct data definition and assisting with the correct interpretation of the outcomes, often doing so voluntarily. Moreover, while assessments of low-value care provide essential insights into its presence, the actual change can only be realised by healthcare professionals themselves. Therefore, both the government and professional societies should therefore actively support clinical leaders (and healthcare professionals in general) aiming to examine and reduce low-value care. This support can be achieved through allocation of funding and general support (e.g. data availability, support from professional societies) to clinical leaders aiming to address low-value care. Future research should focus on building a framework that describes an accurate and uniform method for assessing low-value care. In the discussion of this thesis, we have proposed the initial steps towards such a framework. Even though assessments of low-value care primarily serve gain insight and are a starting point for reducing its provision, it is important to acknowledge that low-value care will probably never be completely removed from the healthcare system. The reasons for this are inherently tied to the current state of healthcare provision. First, healthcare professionals face the enormous task to continuously stay up-to-date with relevant guidelines and evidence. However, due to the continuous development of the underlying evidence for different treatments this has become an almost impossible task, given the time they spend providing care to their patients. [25, 59-61] Second, recent evidence shows that over the past decades, physicians have perceived increased pressure from their patients to provide low-value care. Some patients hold (inaccurate) assumptions or personal believes, such as 'more care is always better' or 'newer technologies are better than older ones'. These pervasive assumptions or beliefs, often started or reinforced by either the patients' peers, the (social) media, or the internet, often lead to patients requesting low-value care from their healthcare providers. [34, 35, 41, 45, 62-64] Although the motives of healthcare professionals to help or reassure their patients through providing a low-value service are understandable, they still do not justify it. In the end, not providing low-value care is often, if not always, more beneficial for the patient.

Conclusion

The general conclusion of this thesis is that it is possible to assess the extent of low-value care in the Netherlands. This assessment should be a meticulous process. It requires a clear definition of low-value care, selection of an adequate database, operationalization of the definition and the general assessment methods (e.g. lenses) of assessment. Furthermore, we can conclude that a large variation exists in the provision of low-value care between both clinical practices and healthcare providers. We would therefore like to argue that general assessments of the total magnitude of low-value care within countries or healthcare systems do not contribute to the discussion of how to reduce low-value care provision. Hence, these assessments often generalise the findings of a subset of care practices to the entire healthcare system, without regarding the differences in occurrence

of low-value care between different practices. In order to address the problem of low-value care, more nuanced and specific assessments at the level of individual clinical practices are required. These nuanced assessments should include detailed descriptions of the assessment methods employed, the definitions of low-value care that were used, their operationalization (including assumptions), and should provide ample information to correctly interpret their results. But most importantly, these assessments should be performed in collaboration with the appropriate healthcare professionals. While they have the (medical) knowhow to assist in building the correct definitions, and can provide context to the eventual assessment outcomes. In conclusion, we believe that nuanced assessments are an important first step to sparking a discussion regarding the problem of low-value care and improving the quality of healthcare.

References

1. Kool RB, Verkerk EW, Meijs J, van Gorp N, Maessen M, Westert G, et al. Assessing volume and variation of low-value care practices in the Netherlands. European Journal of Public Health. 2020;30(2):236-40.
2. Laan BJ, Maaskant JM, Spijkerman IJB, Borgert MJ, Godfried MH, Pasmooij BC, et al. De-implementation strategy to reduce inappropriate use of intravenous and urinary catheters (RICAT): a multicentre, prospective, interrupted time-series and before and after study. Lancet Infect Dis. 2020;20(7):864-72.
3. Koggel LM, Lantinga MA, Büchner FL, Drenth JPH, Frankema JS, Heeregrave EJ, et al. Predictors for inappropriate proton pump inhibitor use: observational study in primary care. Br J Gen Pract. 2022;72(725):e899-e906.
4. Nivel zorgregistraties. Nivel Zorgregistraties. 2019.
5. Dros JT, van Dijk CE, Bos I, Meijer WM, Chorus A, Miedema H, et al. Healthcare utilization patterns for knee and hip osteoarthritis before and after changes in national health insurance coverage: a data linkage study from the Netherlands. Health Policy. 2023;133:104825.
6. Wong A, Stam, M., Dijk, C. van, Verheij, R., Flinterman, L., Bos, I. Trommelvliesbuisjes en antibioticagebruik in de zorgtrajecten van kinderen met middenoorontsteking. 2021.
7. van Dijk C, Dik, J.W., Noor Heim, N., Verheij, R., Leenen, C., Flinterman, L., Bos, I. Controleafspraken bij patiënten met nieuw event diepe veneuze trombose of longembolie. Zorginstituut Nederland, Nederlands Instituut voor Onderzoek van de Gezondheid (NIVEL); 2021.
8. Zorginstituut Nederland. Zinnige Zorg - Verbetersignalement COPD. 2019.
9. Zorginstituut Nederland. Zinnige Zorg - Verbetersignalement Maagklachten. 2021.
10. Zorginstituut Nederland. Zinnige Zorg - Verbetersignalement Astma. 2021.
11. Zorginstituut Nederland. Zinnige Zorg - Verbetersignalement Onderste luchtweginfecties. 2021.
12. Zorginstituut Nederland. Zinnige Zorg - Verbetersignalement Urineweginfecties. 2021.
13. Miller G, Rhyan C, Beaudin-Seiler B, Hughes-Cromwick P. A Framework for Measuring Low-Value Care. Value Health. 2018;21(4):375-9.
14. Westert GP, Groenewoud S, Wennberg JE, Gerard C, DaSilva P, Atsma F, Goodman DC. Medical practice variation: public reporting a first necessary step to spark change. Int J Qual Health Care. 2018;30(9):731-5.
15. Pendrith C, Bhatia M, Ivers NM, Mecredy G, Tu K, Hawker GA, et al. Frequency of and variation in low-value care in primary care: a retrospective cohort study. CMAJ Open. 2017;5(1):E45-E51.
16. Mafi JN, Reid RO, Baseman LH, Hickey S, Totten M, Agniel D, et al. Trends in Low-Value Health Service Use and Spending in the US Medicare Fee-for-Service Program, 2014-2018. JAMA Network Open. 2021;4(2):e2037328-e.
17. Mafi JN, Russell K, Bortz BA, Dachary M, Hazel WA, Jr., Fendrick AM. Low-Cost, High-Volume Health Services Contribute The Most To Unnecessary Health Spending. Health Aff (Millwood). 2017;36(10):1701-4.

18. Chami N, Simons JE, Sweetman A, Don-Wauchope AC. Rates of inappropriate laboratory test utilization in Ontario. Clin Biochem. 2017;50(15):822-7.
19. Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring Low-Value Care in Medicare. JAMA Internal Medicine. 2014;174(7):1067-76.
20. Schwartz AL, Zaslavsky AM, Landon BE, Chernew ME, McWilliams JM. Low-Value Service Use in Provider Organizations. Health Serv Res. 2018;53(1):87-119.
21. Colla CH, Morden NE, Sequist TD, Schpero WL, Rosenthal MB. Choosing wisely: prevalence and correlates of low-value health care services in the United States. J Gen Intern Med. 2015;30(2):221-8.
22. Chalmers K, Pearson SA, Badgery-Parker T, Brett J, Scott IA, Elshaug AG. Measuring 21 low-value hospital procedures: claims analysis of Australian private health insurance data (2010-2014). BMJ Open. 2019;9(3):e024142.
23. Chalmers K, Smith P, Garber J, Gopinath V, Brownlee S, Schwartz AL, et al. Assessment of Overuse of Medical Tests and Treatments at US Hospitals Using Medicare Claims. JAMA Network Open. 2021;4(4):e218075-e.
24. Bouck Z, Pendrith C, Chen XK, Frood J, Reason B, Khan T, et al. Measuring the frequency and variation of unnecessary care across Canada. BMC Health Serv Res. 2019;19(1):446.
25. 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.
26. Helou MA, DiazGranados D, Ryan MS, Cyrus JW. Uncertainty in Decision Making in Medicine: A Scoping Review and Thematic Analysis of Conceptual Models. Acad Med. 2020;95(1):157-65.
27. Isaac T, Rosenthal MB, Colla CH, Morden NE, Mainor AJ, Li Z, et al. Measuring overuse with electronic health records data. Am J Manag Care. 2018;24(1):19-25.
28. World Health Organisation. Anatomical Therapeutic Chemical (ATC) Classification: World Health Organisation; 2023 [Available from: https://www.whocc.no/atc_ddd_index/.
29. World Health Organisation. International Classification of Primary Care, 2nd edition (ICPC-2) 2003 [Available from: https://www.who.int/standards/classifications/other-classifications/international-classification-of-primary-care.
30. World Health Organisation. International Classification of Diseases 11th Revision 2022 [Available from: https://icd.who.int/en.
31. Nederlandse Zorgautoriteit. DBC-zorgproducten 2023 [Available from: https://www.opendisdata.nl/msz/zorgproduct.
32. Chalmers K, Pearson SA, Elshaug AG. Quantifying low-value care: a patient-centric versus service-centric lens. BMJ Qual Saf. 2017;26(10):855-8.
33. Marcotte LM, Schuttner L, Liao JM. Measuring low-value care: learning from the US experience measuring quality. BMJ Quality & Safety. 2020;29(2):154-6.
34. Kool RB, Verkerk EW, Winnemuller LJ, Wiersma T, Westert GP, Burgers JS, van Dulmen SA. Identifying and de-implementing low-value care in primary care: the GP's perspective-a cross-sectional survey. BMJ Open. 2020;10(6):e037019.
35. Buist DS, Chang E, Handley M, Pardee R, Gundersen G, Cheadle A, Reid RJ. Primary Care Clinicians' Perspectives on Reducing Low-Value Care in an Integrated Delivery System. Perm J. 2016;20(1):41-6.

36. Do LA, Koethe BC, Daly AT, Chambers JD, Ollendorf DA, Wong JB, et al. State-Level Variation In Low-Value Care For Commercially Insured And Medicare Advantage Populations. Health Aff (Millwood). 2022;41(9):1281-90.
37. Schwartz AL, Jena AB, Zaslavsky AM, McWilliams JM. Analysis of Physician Variation in Provision of Low-Value Services. JAMA Internal Medicine. 2019;179(1):16-25.
38. Cheikh-Moussa K, Caro Mendivelso J, Carrillo I, Astier-Peña MP, Olivera G, Silvestre C, et al. Frequency and estimated costs of ten low-value practices in the Spanish primary care: a retrospective study. Expert Opin Drug Saf. 2022;21(7):995-1003.
39. Shrank WH, Rogstad TL, Parekh N. Waste in the US Health Care System: Estimated Costs and Potential for Savings. JAMA. 2019;322(15):1501-9.
40. Mafi JN, Parchman M. Low-value care: an intractable global problem with no quick fix. BMJ Qual Saf. 2018;27(5):333-6.
41. Verkerk EW, Van Dulmen SA, Born K, Gupta R, Westert GP, Kool RB. Key Factors that Promote Low-Value Care: Views of Experts From the United States, Canada, and the Netherlands. Int J Health Policy Manag. 2021;11(8):1514-21.
42. Landon SN, Padikkala J, Horwitz LI. Identifying drivers of health care value: a scoping review of the literature. BMC Health Services Research. 2022;22(1):845.
43. Willson A. The problem with eliminating 'low-value care'. BMJ Quality & Safety. 2015;24(10):611-4.
44. Ingvarsson S, Hasson H, von Thiele Schwarz U, Nilsen P, Powell BJ, Lindberg C, Augustsson H. Strategies for de-implementation of low-value care—a scoping review. Implementation Science. 2022;17(1):73.
45. Norton WE, Chambers DA. Unpacking the complexities of de-implementing inappropriate health interventions. Implement Sci. 2020;15(1):2.
46. van Dulmen SA, Naaktgeboren CA, Heus P, Verkerk EW, Weenink J, Kool RB, Hooft L. Barriers and facilitators to reduce low-value care: a qualitative evidence synthesis. BMJ Open. 2020;10(10):e040025.
47. Leigh JP, Sypes EE, Straus SE, Demiantschuk D, Ma H, Brundin-Mather R, et al. Determinants of the de-implementation of low-value care: a multi-method study. BMC Health Services Research. 2022;22(1):450.
48. Parchman ML, Henrikson NB, Blasi PR, Buist DS, Penfold R, Austin B, Ganos EH. Taking action on overuse: Creating the culture for change. Healthc (Amst). 2017;5(4):199-203.
49. Registratie aan de bron. Facilitating clinical documentation at the point of care. Sharing and optimising healthcare information. 2016.
50. Arts DGT, de Keizer NF, Scheffer G-J. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. Journal of the American Medical Informatics Association. 2002;9(6):600-11.
51. Shemtob L, Beaney T, Norton J, Majeed A. How can we improve the quality of data collected in general practice? BMJ. 2023;380:e071950.
52. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, Godtliebsen F. Challenges and opportunities beyond structured data in analysis of electronic health records. WIREs Computational Statistics. 2021;13(6):e1549.

53. van der Bij S, Khan N, ten Veen P, de Bakker DH, Verheij RA. Improving the quality of EHR recording in primary care: a data quality feedback tool. Journal of the American Medical Informatics Association. 2016;24(1):81-7.
54. Bardach NS, Cabana MD. The unintended consequences of quality improvement. Curr Opin Pediatr. 2009;21(6):777-82.
55. Mannion R, Braithwaite J. Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. Intern Med J. 2012;42(5):569-74.
56. Powell AA, White KM, Partin MR, Halek K, Christianson JB, Neil B, et al. Unintended consequences of implementing a national performance measurement system into local practice. J Gen Intern Med. 2012;27(4):405-12.
57. Mathias JS, Baker DW. Developing quality measures to address overuse. Jama. 2013;309(18):1897-8.
58. Zegers M, Veenstra GL, Gerritsen G, Verhage R, van der Hoeven HJG, Welker GA. Perceived Burden Due to Registrations for Quality Monitoring and Improvement in Hospitals: A Mixed Methods Study. Int J Health Policy Manag. 2022;11(2):183-96.
59. Kim DD, Do LA, Daly AT, Wong JB, Chambers JD, Ollendorf DA, Neumann PJ. An Evidence Review of Low-Value Care Recommendations: Inconsistency and Lack of Economic Evidence Considered. Journal of General Internal Medicine. 2021;36(11):3448-55.
60. Powers BW, Jain SH, Shrank WH. De-adopting Low-Value Care: Evidence, Eminence, and Economics. JAMA. 2020;324(16):1603-4.
61. Tenny S, Varacallo M. Evidence Based Medicine. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2023, StatPearls Publishing LLC.; 2023.
62. Verkerk EW, Boekkooi JAH, Pels EGM, Kool RB. Exploring patients' perceptions of low-value care: An interview study. Patient Education and Counseling. 2023;111:107687.
63. Lugtenberg M, Vries P, Evertse A, Schaick J, Westert G, Burgers J. Welke barrières ervaren huisartsen bij de toepassing van aanbevelingen uit NHG-Standaarden? Huisarts en wetenschap. 2012;2010:13-9.
64. Sypes EE, de Grood C, Clement FM, Parsons Leigh J, Whalen-Browne L, Stelfox HT, Niven DJ. Understanding the public's role in reducing low-value care: a scoping review. Implementation Science. 2020;15(1):20.

Addendum

Summary

Low-value care is defined as care that is unlikely to benefit the patient given the harms, costs and available alternatives. Despite physicians' aiming to provide care that achieves the best results with the most efficient use of resources, international studies show that low-value care is still abundantly provided. The provision of low-value care provision is problematic since it unnecessarily exposes patients to potential harm and could waste the already limited healthcare resources. Its reduction is therefore indispensable to improve both the quality and affordability of healthcare. However, in order to reduce it, one first needs to gain insight into whether or not it is actually present and posts a problem. Unfortunately, the magnitude of its provision in the Netherlands is still largely unknown. We therefore aimed to assess the volume of several types of low-value care within the Dutch healthcare system.

To achieve our aim, we were first needed to examine the methods used to assess the volume of low-value care and gain insight into how differences in methods can impact assessment outcomes. **Chapter 2** describes our systematic analysis of the outcomes and underlying methods of low-value diagnostic testing. The results reveal the presence of significant variation in outcomes of assessments of (often similar) low-value diagnostic tests. Closer examination revealed that the observed differences in assessment outcomes can largely be attributed to differences in definitions of low-value tests used, their operationalisation and the used assessment methods (in particular assessment lenses). Additionally, we also observed that these factors are often not taken into account when describing or comparing assessments outcomes, making their interpretation challenging. Therefore, we argue that to make assessments of low-value care usable and comparable, standardized definitions and assessment methods are required.

After our examination of the methods used to assess the volume of low-value care, we put the newly acquired insights into practice. We conducted multiple assessments among Dutch general practitioners (**chapter 3, 4 and 6**) and hospital clinicians (**chapter 5**) in order to gain insight into the extent of low-value care in the Netherlands.

Chapter 3 describes our retrospective analysis of three types of low-value pharmaceutical care among Dutch general practitioners (GPs) using Dutch GP clinical registration data from the Nivel. We examined the low-value prescription of antibiotics for conjunctivitis, benzodiazepines for lower-back pain, and chronic prescription of acid reducing medication. Through the use of data containing clinical detail, we were able to accurately assess the provision of three types of low-value pharmaceutical care between 2016 and 2019. The assessment outcomes revealed that between 53% - 61% of patients received an low-value antibiotics prescription for conjunctivitis, approximately 3% of patients with lower back pain received an inappropriate benzodiazepine prescription, and 88% received an inappropriate chronic ARM prescription. Furthermore it showed that also large variation in low-value utilisation existed between the included practices. These

observations highlight that low-value care provision differs between different types of care and practices, each having its own distinct patterns and utilisation rates. Thereby emphasizing the need for customizing de-implementation efforts to the type(s) of low-value care you aim to reduce.

Chapter 4 presents the outcomes of our assessment of vitamin B12- and D-testing among Dutch general practitioners, using claims data from of one of the four big Dutch health insurers (VGZ). Between 2015 and 2019, we observed a 98.1% increase in vitamin B12 tests and a 112.0% increase in vitamin D tests. The percentage of patients per practice receiving a vitamin B12 test increased from 4.8% to 8.4%, and for vitamin D tests, it increased from 4.7% to 9.1% over the examined period. These observed increase in both vitamin B12 and D-testing, as well as the rising proportion of patients receiving these tests are indicative of increased non-indicated testing. However, due to a lack of clinical information in the used claims data, we were unable to assess the appropriateness of the included vitamin tests.

Chapter 5 describes our examination of low-value vitamin B12- and D-testing among Dutch hospital clinicians using administrative registration data from the Dutch Hospital Data. The assessment outcomes reveal that over 2015-2019, at least 79.0% of vitamin B12- and 82.0% of D-test ordered in patients aged 18-70 were considered low-value. The total number of vitamin B12-tests increased by 2.0%, and the number of D-tests by 12.2% over the examined period. Furthermore, our study showed that the proportion of the unexplained variation in non-indicated vitamin B12- and D-tests attributed to differences between hospitals remained consistently low during the examined period. We also observed the presence of a fairly strong positive correlation in non-indicated vitamin B12- and D-testing among the included hospitals (Rho: 0.86). Combined these findings suggest that vitamin B12- and D-tests are frequently utilized inappropriately among clinicians, and that high rates of inappropriate utilisation are often observed within the same hospital. Suggesting that all hospitals face a similar challenge, which is to reduce the number of non-indicated B12- and D-tests.

Chapter 6 describes our assessment of the provision of three types of low-value care among Dutch GPs before, during, and after the COVID-19 restrictions using clinical registration data from 2017 to 2022. The following three types of low-value care were selected in collaboration with the participating GPs: the use of imaging in the diagnosis of musculoskeletal complaints related to the back or knee, the prescription of antibiotics for otitis media acuta (OMA) in children without severe symptoms and repeat opioid prescriptions, without a prior GP visit. During the COVID-19 restrictions period, the Incidence Rates (IRs) of episodes related to all three types of GP care decreased significantly. The IR of episodes of back or knee pain decreased by 12%, OMA episodes by 54% and opioid prescription rate by 13%. Only the IR of OMA prescriptions remained significantly lower (22%) during the post-restrictions period. The provision of low-value care also changed. The IR of imaging for back or knee pain and prescription of antibiotics

for OMA both decreased significantly during the COVID-restrictions period (by 21% and 78%, respectively), but only the rate of the low-value prescription of antibiotics for OMA remained significantly lower (by 63%) during the post-restrictions period. The IR of inappropriately repeated opioid prescriptions remained unchanged over all three periods. These findings indicate that both the presence and the effect of the COVID-19 restrictions differ between the different types of low-value care. The combination of the findings of this study confirm the notion that reducing low-value care is a complex challenge that requires tailored interventions and is not easily nor quickly achieved.

In **chapter 7** we examined the impact of an educational intervention on the low-value (chronic) prescription of acid reducing medication (ARM) in case of dyspepsia. The intervention involved distributing an educational waiting room posters and flyers informing both patients and GPs regarding the appropriate indications for prescription of an ARM for dyspepsia, which also referred to an online decision aid. Using registration data obtained from the Nivel, we next examined whether the intervention had changed the amount of chronic ARM users between the second halves of 2021 and 2022. Our analysis of the odds ratio associated with receiving a low-value chronic ARM prescription between the two period, revealed no significant difference in low-value chronic ARM use between the control and intervention group (Odds ratio: 0.95 [95% CI: 0.72 – 1.25], $p>0.05$). However, even though this study did not find a significant reduction in low-value chronic ARM prescriptions, it does provide an indication that the number of chronic ARM users has reduced. In the present study we found that among the included practices, approximately half of the chronic ARM users lacked an indication compared to the 88% found in a previous assessment.

In **chapter 8** of this thesis, we conclude that low-value is also widely provided within the Netherlands, displaying wide variation in its utilisation between different types of care. Moreover, we observed throughout this thesis that the methods used in assessments in general are rarely standardized or clearly described, thereby hindering the meaningful comparison. We therefore propose a conceptual roadmap for the assessment of low-value care, to improve the quality and comparability of assessments of low-value care. Based on the studies presented in this thesis, we believe that there are four important phases that constitute a comprehensive and reliable assessment. In the first phase, low-value care should be selected and a clear data definition built. This selection (or prioritization) process should preferably be based on high quality evidence and performed in collaboration with medical professionals. In the second phase, an appropriate database and assessment method is selected. It is crucial to ensure that the selected database contains the necessary amount of clinical detail required to accurately distinguish appropriate from inappropriate care as described in the data definition. Similarly, the choice of assessment method should align with the level at which you intend to communicate the assessment outcomes, as this will impact the assessment denominator and, ultimately, the assessment outcomes. The third phase involves the process of actually performing your assessment. In this phase it is important to thoroughly evaluate

the results obtained, and engage in discussions with the involved stakeholders regarding their meaning. These feedback moments could provide valuable insights and may lead to adjustments in your data definition or methods, thereby improving the validity of your assessment which is crucial for the final phase. In the fourth and final phase, the assessment outcomes are ready for publication and can be communicated. It is of vital importance to provide sufficient information accompanying the information, to ensure the correct interpretation of your findings. Transparency and clarity are vital aspects in this communication process. We think that following the abovementioned steps will enhance both the comparability and quality of assessments of low-value care. This will ultimately contribute to the improvement of the quality of healthcare.

Dutch summary (Samenvatting)

Niet-gepaste zorg is zorg die weinig toegevoegde waarde heeft voor de patiënt gezien de schadelijke effecten, kosten, beschikbare alternatieven en patiëntvoorkeur. Voorbeelden zijn antibiotica bij virale luchtweginfecties of röntgenfoto's bij lage-rugklachten zonder alarmsymptomen. Ondanks dat artsen de beste zorg willen bieden met een zo efficiënt mogelijk gebruik van middelen, tonen internationale studies aan dat niet-gepaste zorg veelvuldig voorkomt. Het verlenen van niet-gepaste zorg moet worden voorkomen omdat het patiënten onnodig blootstelt aan schade en de beperkte middelen voor zorg verspilt. Het verminderen van niet-gepaste zorg is daarom cruciaal om zowel de kwaliteit als de betaalbaarheid van de gezondheidszorg te verbeteren. Om niet-gepaste zorg te verminderen, moet het duidelijk zijn of de zorg daadwerkelijk wordt geleverd en of het een probleem vormt. Er is nog weinig bekend over de omvang van niet-gepaste zorg in Nederland. Het doel van deze thesis was daarom om het volume van verschillende soorten niet-gepaste zorg in het Nederlandse gezondheidssysteem in kaart te brengen.

Hiervoor hebben we eerst de methoden onderzocht die gebruikt worden bij het bepalen van het volume van niet-gepaste zorg. Zo kunnen we zicht krijgen op de invloed hiervan op de uitkomsten. **Hoofdstuk 2** beschrijft een systematische analyse van de invloed van verschillen in meetmethoden op de uitkomsten van metingen over niet-gepaste diagnostische testen. De resultaten laten zien dat er aanzienlijke variatie in metingsuitkomsten bestaat tussen metingen over (vaak vergelijkbare) diagnostische testen. Nader onderzoek toonde aan dat de waargenomen verschillen in beoordelingsresultaten grotendeels kunnen worden toegeschreven aan verschillen in de gebruikte definities van niet-gepaste testen, hun operationalisatie en de onderliggende meetmethoden. En dan met name het gebruik van verschillende beoordelingslenzen. Bij het gebruik van de patiënt-indicatielens wordt het percentage patiënten dat de niet-gepaste zorg ontvangt gerapporteerd. Bij een service-indicatielens worden alle handelingen in een bepaalde praktijk/database meegenomen en onderzocht op juiste indicaties. Bovendien viel ook op dat de bovengenoemde factoren (definities, operationalisatie en meetmethoden) vaak niet worden meegenomen bij het beschrijven of vergelijken van de uitkomsten. Dit maakt een correcte interpretatie hiervan uitdagend. Wij betogen daarom dat voor het verkrijgen van bruikbare en vergelijkbare metingen van niet-gepaste zorg, gestandaardiseerde definities en beoordelingsmethoden vereist zijn.

Hierna hebben we de nieuwverworven inzichten in de praktijk gebracht. We hebben meerdere studies uitgevoerd onder zowel Nederlandse huisartsen (**hoofdstuk 3, 4 en 6**) als medisch specialisten (**hoofdstuk 5**), om inzicht te krijgen in de omvang van niet-gepaste zorg in Nederland.

Hoofdstuk 3 beschrijft onze retrospectieve analyse van drie soorten niet-gepaste farmaceutische zorg onder Nederlandse huisartsen tussen 2016 en 2019. Voor deze analyse hebben wij gebruik gemaakt van registratiegegevens van het Nivel uit

het huisartseninformatiesysteem van Nederlandse huisartsen. We onderzochten het onnodig voorschrijven van antibiotica voor een ooginfectie (conjunctivitis), kalmeringsmiddelen (benzodiazepinen) voor lage rugpijn en chronisch voorschrijven van maagzuurremmers. Doordat de gebruikte data veel klinische informatie bevat, was het mogelijk om het voorkomen van deze drie vormen van niet-gepaste farmaceutische zorg nauwkeurig te beoordelen. De resultaten lieten zien dat in de periode 2016 - 2019 er tussen de 53% en 61% van de patiënten een niet-gepast antibiotica-recept kregen voor conjunctivitis. Verder bleek dat er grote variatie was in het onnodig voorschrijven tussen de geïncludeerde praktijken. Ongeveer 3% van de patiënten met lage rugpijn kreeg een niet-gepaste benzodiazepine-recept en 88% een niet-gepast langdurig recept voor maagzuurremmers. Deze observaties benadrukken dat het aanbieden van niet-gepaste zorg verschilt per soort zorg en per praktijk. Dit wijst erop dat het noodzakelijk is om de inspanningen voor het verminderen van het gebruik van niet-gepaste zorg aan te passen aan het type niet-gepaste zorg dat men wil verminderen.

Hoofdstuk 4 bevat de resultaten van onze analyse van het gebruik van vitamine B12- en D-bepalingen onder Nederlandse huisartsen tussen 2015 en 2019. In deze studie is gebruik gemaakt van declaratiegegevens van een van de vier grote Nederlandse zorgverzekeraars (VGZ). Tussen 2015 en 2019 observeerden we een toename van 98,1% in vitamine B12-bepalingen, en een toename van 112,0% in vitamine D-bepalingen. Het percentage patiënten per praktijk dat een vitamine B12-bepaling kreeg, steeg van 4,8% naar 8,4%, en voor vitamine D-bepaling van 4,7% naar 9,1% over de onderzochte periode. De toename in zowel vitamine B12- als D-testen, evenals het stijgende percentage patiënten dat deze bepalingen kreeg, duiden op een toename van niet-geïndiceerde testen. Vanwege het gebrek aan klinische informatie in de gebruikte data, was het echter niet mogelijk om inzicht te krijgen in de gepastheid van de vitaminebepalingen.

Hoofdstuk 5 beschrijft de uitkomsten van het onderzoek naar niet-gepast gebruik van vitamine B12- en D-bepalingen onder medisch specialisten in Nederlandse ziekenhuizen. We gebruikten hiervoor administratieve registratiegegevens van Dutch Hospital Data. De resultaten hiervan tonen aan dat tussen 2015-2019 ten minste 79,0% van de vitamine B12- en 82,0% van de D-bepalingen aangevraagd onder patiënten tussen de 18 en 70 jaar oud als niet-gepast konden worden beschouwd. Het aantal vitamine B12-bepalingen nam toe met 2,0% en het aantal D-bepalingen met 12,2%. Verder toonde onze studie aan dat het aandeel van de onverklaarde variatie in niet-geïndiceerde vitamine B12- en D-bepalingen dat toegeschreven kan worden aan verschillen tussen ziekenhuizen, consistent laag bleef over de onderzochte periode. Daarnaast was er sprake van een vrij sterke positieve correlatie in niet-geïndiceerde vitamine B12- en D-bepalingen. Nederlandse ziekenhuizen staan voor een vergelijkbare uitdaging om het aantal niet-geïndiceerde B12- en D-bepalingen te verminderen.

Hoofdstuk 6 beschrijft de evaluatie van het leveren van drie soorten niet-gepaste zorg onder Nederlandse huisartsen vóór, tijdens en na de COVID-19-beperkingen met

behulp van klinische registratiegegevens van 2017 tot 2022 van een huisartsennetwerk. In samenwerking met enkele huisartsen hebben we hiervoor de volgende drie soorten niet-gepaste zorg geselecteerd: het gebruik van beeldvorming bij de diagnose van klachten aan rug of knie, het voorschrijven van antibiotica voor middenoorontsteking (otitis media acuta, OMA) in kinderen zonder ernstige symptomen en het herhalen van opioïderecepten zonder een voorafgaand huisartsenbezoek. Gedurende de periode van COVID-19-beperkingen daalden de incidentie ratios (IR's) van episodes die verband houden met alle drie de soorten huisartsenzorg significant. De IR van episodes van rug- of kniepijn daalde met 12%, episodes van OMA met 54% en het voorschrijfpercentage van opioïden met 13%. Alleen de IR van voorschriften voor OMA bleef significant lager (22%) nadat de beperkingen waren opgeheven. Ook de verstrekking van niet-gepaste zorg veranderde in de meetperiode. De IR van beeldvorming voor rug- of kniepijn en het voorschrijven van antibiotica voor OMA daalden beide significant ten tijde van de COVID-beperkingen (met respectievelijk 21% en 78%). Echter, alleen de IR van het onnodig voorschrijven van antibiotica voor OMA bleef significant lager (met 63%) gedurende de post-beperkingsperiode. De IR van ongepaste herhaalde opioïde recepten bleef onveranderd gedurende alle drie de perioden. Deze bevindingen geven aan dat zowel de aanwezigheid als het effect van de COVID-19-beperkingen verschillen tussen de verschillende soorten niet-gepaste zorg. De combinatie van de bevindingen van dit onderzoek bevestigt het idee dat het verminderen van niet-gepaste zorg een complexe uitdaging is die op maat gemaakte interventies vereist.

In **hoofdstuk 7** onderzochten we de impact van een educatieve interventie op het chronisch voorschrijven van maagzuurremmers in geval van bovenbuikklachten (dyspepsie). De educatieve interventie bestond uit het verstrekken van flyers en een poster aan huisartsen om uit te delen aan patiënten, met daarop een verwijzing naar een online keuzehulp. De keuzehulp had als doel patiënten gerust te tellen en te informeren wat ze er zelf aan kunnen doen behalve het slikken van geneesmiddelen. Met behulp van registratiegegevens verkregen van het Nivel hebben we vervolgens onderzocht of de interventie het aantal gebruikers van chronische maagzuurremmers had verminderd tussen de tweede helft van 2021 en 2022. Onze analyse van de odds ratio geassocieerd met het ontvangen van een niet-gepast chronisch voorschrift voor maagzuurremmers tussen de twee periodes, toonde geen significante vermindering tussen de controlegroep en de interventiegroep (Odds ratio: 1,02 [95% CI: 0,76 - 1,37], p>0,05). Het onderzoek geeft wel een indicatie dat het aantal gebruikers van chronische maagzuurremmers is afgenomen. Dit onderzoek laat zien dat ongeveer de helft van de mensen die chronische maagzuurremmers gebruiken geen indicatie had, in vergelijking met de 88% die werd gevonden in een eerdere studie.

In **hoofdstuk 8** van deze thesis concluderen we dat niet-gepaste zorg ook in Nederland wordt verleend, en dat er aanzienlijke variatie aanwezig is in de aanwezigheid ervan tussen verschillende vormen van zorg. Ook concluderen we dat de gebruikte methoden voor het meten over het algemeen niet gestandaardiseerd zijn of voldoende duidelijk

zijn beschreven. Dit maakt een zinvolle vergelijking van metingen en duiding van de resultaten moeilijk. aarom formuleren we, gebaseerd op de ervaringen opgedaan in de studies onderdeel van deze thesis, een conceptueel raamwerk voor het meten van niet-gepaste zorg. We verwachten dat zo de kwaliteit en vergelijkbaarheid van metingen van niet-gepaste zorg kan verbeteren. Het raamwerk bestaat uit vier belangrijke fasen die een uitgebreide en betrouwbare meting vormen. In de eerste fase moet de te onderzoeken vorm van niet-gepaste zorg worden geselecteerd en een duidelijke definitie van de gegevens worden opgebouwd. Dit selectie- (of prioriterings)proces moet bij voorkeur gebaseerd zijn op hoogwaardig bewijs dat de zorg niet-gepast is en worden uitgevoerd in samenwerking met zorgverleners. In de tweede fase moet een geschikte database en beoordelingsmethode worden gekozen. Het is cruciaal om ervoor te zorgen dat de geselecteerde database de benodigde hoeveelheid klinische details bevat om nauwkeurig onderscheid te maken tussen passende en niet-gepaste zorg zoals beschreven in de datadefinitie. Ook moet de keuze van de beoordelingsmethode in lijn zijn met het niveau waarop de resultaten worden gecommuniceerd. Dit heeft immers impact op de noemer van de meting en uiteindelijk op de resultaten. De derde fase omvat het daadwerkelijke uitvoeren van de meting. In deze fase is het belangrijk om de verkregen resultaten grondig te evalueren en in gesprek te gaan met de betrokken belanghebbenden over hun betekenis. Deze feedbackmomenten kunnen waardevolle inzichten opleveren en kunnen leiden tot aanpassingen in de datadefinitie of methoden. Hierdoor wordt de validiteit van de meting verbeterd wat cruciaal is voor de laatste fase. In de vierde en laatste fase zijn de metingsresultaten klaar voor publicatie en kan erover worden gecommuniceerd. Het is belangrijk om voldoende informatie te verstrekken bij de uitkomsten, om de bevindingen juist te interpreteren en te duiden. Transparantie en duidelijkheid zijn essentiële aspecten in dit communicatieproces. We denken dat het volgen van de hierboven genoemde stappen zowel de vergelijkbaarheid als de kwaliteit van metingen van niet-gepaste zorg zullen verbeteren. Dit zal uiteindelijk bijdragen aan de verbetering van de kwaliteit van de gezondheidszorg.

Dankwoord

Vier en een half jaar heb ik met veel plezier mogen werken aan dit proefschrift. Ik ben blij met het resultaat, maar weet ook dat ik het nooit had gehaald zonder de steun en hulp van iedereen om mij heen. Ik wil daarom allen uitgebreid bedanken voor hun ondersteuning tijdens dit traject. Echter, in het licht van de onderliggende boodschap uit mijn proefschrift – meer is niet altijd beter – wil ik graag een aantal personen in het bijzonder bedanken. Ik wil echter nog wel vooropstellen dat ik eenieder die heeft bijgedragen in de vorm van enige (inhoudelijke) input, zeer dankbaar ben.

Allereerst, ben ik veel dank verschuldigd aan mijn promotieteam: **Prof. dr. Tijn Kool**, **dr. Simone van Dulmen** en **prof. dr. Gert Westert**. Jullie enthousiasme en expertise rondom het onderwerp '*Niet-gepaste zorg*' hebben mij oprecht aangestoken en gemotiveerd om dit geheel tot een goed en mooi einde te willen brengen. Ik wil jullie graag bedanken voor jullie vertrouwen, steun in zowel werk als privé gerelateerde onderwerpen. Geen probleem was te groot of ik kon het wel met jullie bespreken; er stond altijd wel een van jullie klaar om mij verder te helpen en te werken naar een oplossing. Gezamenlijk had ik geen beter team kunnen wensen, mede dankzij jullie was ik in staat om het geheel naar behoren af te ronden. Jullie enthousiasme, betrokkenheid en kritische blik hielden mij scherp en hebben persoonlijke groei gestimuleerd, waarvoor nogmaals dank. Gelukkig is het mij niet gelukt om jullie gedurende deze 4,5 jaar te vermoeien met mijn eindeloze lappen tekst die ik bij vlagen rondstuurde. Als dank daarvoor houd ik dit laatste stukje wat beknopter; dat hebben jullie wel verdiend!

Geachte leden van de promotiecommissie, **prof. dr. M.G.M. Olde Rikkert**, **prof. dr. R. Verheij** en **prof. dr. D.L.M. Zwart**. Ik wil jullie allen danken voor het kritisch doornemen van het proefschrift en uw bereidheid te opponeren tijdens de openbare verdediging.

Onderzoek doe je wat mij betreft nooit alleen. Naast natuurlijk de steun vanuit mijn promotieteam waren er natuurlijk nog velen anderen die hebben bijgedragen aan dit traject. Allereerst, natuurlijk mijn mede '*Doen of Laten*' teamleden. Naast **Tijn** en **Simone** waren ook **Eva**, **Daniëlle** en **Angelique** gedurende mijn promotietraject altijd van de partij. Jullie kritische blikken, luchtige praatjes, (snelle) adviezen en vooral de gezelligheid hebben zeker bijgedragen aan mijn boekje.

Dan volgen kort daarop natuurlijk al snel mijn IQ-collega's. Ik wil graag **Irah**, **Reinier**, **Femke**, **Janine** en de vele anderen bedanken voor de (statistische) ondersteuning bij mijn onderzoek en de algemene gesprekken. Echter verdient **Irah** natuurlijk nog wel een speciaal bedankje. Jouw warme welkom elke ochtend en de vele gesprekken die wij hebben gehad, zijn altijd zeer gewaardeerd en zullen zeker worden gemist. Ik hoop dat je snel een nieuwe secondant vindt die mijn plekje over kan nemen om je te ondersteunen (hoewel dat wel grote schoenen zijn om te vullen).

De junioren van IQ hebben allen een speciaal plekje in mijn hart, en in het bijzonder: **Koen**, **Julie**, **Ester**, **Nynke**, **Laura**, **Toine**, **Carlijn**, **Sijmen**, **Gijs** (en alle andere **kelderkanjers**). De vermakelijke gesprekken tijdens de lunchwandelingen door Park Brakkenstein, borrels, samenwerkingen en vooral (soms) nutteloze discussies hebben allen bijgedragen aan het draagbaar en leuk houden van de solistische reis dat een promotietraject heet. Hiervoor ben ik jullie allen zeer dankbaar, en ik hoop dat er snel een nieuwe promovendus de "troon van ellende" overneemt in kamer 0.10. Enkele ex-collega's verdienen daarnaast ook zeker een eervolle vermelding: **Joëlle** en **Anna**. Ondanks jullie (jammerlijke) afwezigheid in de laatste anderhalf jaar van het traject hebben jullie echt impact op mij gehad tijdens jullie aanwezigheid.

Naast IQ-collega's heb ik gedurende mijn promotie ook velen anderen ontmoet die mij ondersteund en geïnspireerd hebben. Met name **Karin** en **Maarten** verdienen veel dank voor het klaarzetten van de vele datasets en het meedenken met de analyses. Gedurende mijn tripjes naar zowel het Nivel als de DHD heb ik mij bij vlagen geschaamd om het wederom moeten laten aanpassen van de gevraagde datasets. Jullie hadden gelukkig begrip voor het feit dat dit bij onderzoek doen hoort, en hebben mij hierin elke keer weer hartelijk ontvangen en gesteund.

Ook alle coauteurs van de artikelen die zijn opgenomen in dit proefschrift wil ik graag bedanken. Jullie (medisch) inhoudelijke kennis, kritische blik en input hebben allen bijgedragen aan de totstandkoming van de inhoud van deze thesis waarvoor mijn dank.

Lieve familie en vrienden, voor wie het nog niet duidelijk was, wil ik graag benadrukken dat ik dit proefschrift nooit had kunnen afronden zonder jullie. **Pap**, **Mam**, **Rick**, **Rens**, **Annika**, **Cas** en **Noor**; hoewel inhoudelijk jullie mij vaak niet goed konden volgen, heb ik toch altijd mijn hart bij jullie kunnen luchten. Jullie luisterend oor was iets wat ik geregeld nodig had, en daar ben ik dan ook dankbaar voor. Jullie onvoorwaardelijke steun, kritische blikken, nieuwe invalshoeken en de daaruit volgende fantastische discussies hebben op hun gekke manier ook bijgedragen aan dit proefschrift.

En dan last but not least, mijn dierbare vrienden. Jullie stonden mogelijk allen nog het dichts op de vuurlinie door jullie makkelijke bereikbaarheid. Startende met **Dirk**, oftewel mijn nepbroer. We kennen elkaar al zo lang dat ik mij eigenlijk niet herinner dat je niet aan mijn zijde stond. Samen hebben we veel meegemaakt en ook in dit traject was je er voor mij. Ik kan je niet vertellen hoe zeer ik je daarvoor dankbaar ben, maar hoop dat je dit beseft. Jouw manier van steunen en adviseren is misschien niet de meest alledaagse manier, maar wel iets dat bij mij zeer goed werkt en wat ik ook zeker nodig heb (bij vlagen). **Anthony**, **Luke**, **Peter**, **Wouter**, **Lino**, **Ruud** oftewel mijn Aries. Deze vriendengroep is wat mij betreft het levende bewijs van het bestaan van het natuurfenomeen entropie. Want wie had verwacht dat een groep die zo aan elkaar hangt van willekeur en chaos, toch kon blijven voortbestaan. De willekeurige gesprekken en activiteiten hebben door de jaren heen mijn geest en blik helder gehouden, waarvoor ik jullie zeer dankbaar ben. Hiervoor

ben ik ook mijn dank verschuldigd aan **Maartje** en **Yara**, hoewel jullie wat meer op afstand staan, stonden ook jullie altijd voor mij klaar. Jullie gekke belletjes en vooral willekeurige bezoekjes deden mij altijd goed. En tot slot **Laurie**, aan jou ben ik mogelijk nog de meeste dank verschuldigd. Voor het grootste deel van mijn PhD-avontuur, ben jij mijn rots in de branding geweest waarvoor ik mijn dank niet hard genoeg kan benadrukken. Zonder jouw steun en gezelligheid betwijfel ik of het geheel wel succesvol zou hebben kunnen afronden; iets waarvoor ik je altijd dankbaar blijf.

Research Data Management

Ethics & Privacy

Most of the chapters as described in this thesis used data obtained from human participants (chapters 3, 4, 5, 6, 7). All of these studies were conducted in accordance with the principles of the Declaration of Helsinki and the guidelines of Good Clinical Practice. Informed consent for the use of their personal data of all patients and physicians was previously obtained from the different data suppliers whom supplied us with their data. For all studies, the medical ethical committee CMO Radboudumc, Nijmegen, judged that the studies included in this thesis did all within the Dutch National Law (Medical Research Involving Human Subject Act (WMO)), because no patients were subjected to any actions or interventions. The studies therefore are considered as non-WMO research.

Data collection and storage

For all these studies the analyses were conducted within the (digital) environment of the data suppliers, under supervision of one of the contributing employees. None of the original source data have left the digital environment of each of the data suppliers. Only in case of the study of chapter 7, we received sensitive information regarding participants; e.g. personal information of the general practitioners that signed up for our study. The collected data were pseudonymized and stored in a secured folder at the Radboudumc server of the department of IQ healthcare. Only authorized project team members have access to these data. Identifiable data collected within the Radboudumc were stored in a local secured folder of the department of IQ healthcare, separately from the pseudonymised data.

Availability of data

All studies are published open access (chapter 2, 3, 4) or will be published as such (submitted: chapters 4, 5; in preparation: chapter 6). The data for all chapters will be archived and available upon reasonable request from each of the respective participating data suppliers for 15 years after termination of the study. Reusing the data for future research is only possible after filing a renewed data request at each of the involved data suppliers. The processed anonymous meta-data of the papers included in this thesis are available from the corresponding author upon reasonable request.

About the author

Joris Müskens was born on September 8th, 1992 in Nijmegen, the Netherlands. He completed his secondary school in 2010 at the Canisius College om Nijmegen, over the course of which he developed an interest in research and the workings of the human body. These interests led him starting his bachelor in medical biology in 2011 at the Radboud University Nijmegen. Over the course of his bachelor, he became particularly interested in neurobiology and the associated research. Eventually leading into him following an research masters in medical biology (with a focus on neurobiology) at the Radboud university.

After his graduation in 2017, Joris started his career as a Quality assurance and Compliance officer at "*Interchemie werken 'De Adelaar' B.V.*" in Venray. Over the course of this employment period, he realized that he missed working with data and contributing to improving human wellbeing. By coincidence he came across a vacancy for a PhD position, which combined both of these interests. On the 1st of April in 2019 he started as a PhD candidate at the department of IQ healthcare, examining the volume of low-value care provision in the Netherlands as part of the "*To do or not to do*" team. Following completion of his PhD thesis, Joris will continue to combine his interest in both data and the improvement of human wellbeing by working as a project leader at the Dutch Hospital Data.

PhD Portfolio of J.L.J.M. Müskens

Institute for Health Sciences
Radboudumc

Name PhD candidate:	PhD period:
J.L.J.M. (Joris) Müskens	01/04/2019 – 30/09/2023
Department:	**Promotor(s):**
Scientific Centre for Quality of Healthcare	Prof. dr. G.P. Westert
	Prof. dr. R.B. Kool
Graduate School:	**Co-promotor(s):**
Radboud Institute for Health Sciences	Dr. S.A. (Simone) van Dulmen

TRAINING ACTIVITIES	Hours
Other	
Intervisie - 2020	14
Intervisie - 2021	16
Intervisie - 2022	16
PhD retreat 2019	28
PhD retreat 2020 – Oral presentation	28
PhD retreat 2021 – Oral presentation	28
Courses	
Introduction course for PhD candidates, RIHS (2019)	15
Scientific Writing for PhD candidates (2019)	84
Projectmanagement voor Promovendi (2019)	45
Writing a Review Article (2019)	28
Introductiecursus kwalitatief onderzoek (2020)	42
Introduction into STATA (2020)	12
Radboudumc - eBROK course (2020)	26
Multilevel course (K74), VUMC (2020)	56
Radboudumc - Scientific integrity (2020)	20
Presenting and Poster Pitching (2022)	51
Analytic Storytelling (2022)	20
Career development workshop: The next step in my career (2022)	20
Conferences	
European Implementation Event 2023 (contributed to a symposium session)	17
Congres Resultaten van 8 jaar Doen of laten? 2023 (charing sessions & oral presentation)	6.5
HTAi Annual meeting – Utrecht 2022 (oral presentation)	8
Preventing Overdiagnosis 2022 - Calgary Canada (2x oral presentation)	32
Citrien impact festival 2022 (oral presentation)	6
Skipr Doen of Laten congress 2019	8
TEACHING ACTIVITIES	
Supervision of biomedical science student	120
TOTAL	746.5

Radboudumc
university medical center

Radboud University