BMJ Open

Are weak or negative clinical recommendations associated with higher geographical variation in utilisation than strong or positive recommendations? Cross-sectional study of 24 healthcare services

Agne Ulyte ,[1] Wenjia Wei,[1] Oliver Gruebner,[1,2] Caroline Bähler,[1,3] Beat Brüngger ,[1,3] Eva Blozik,[3,4] Viktor von Wyl,[1] M Schwenkglenks,[1] Holger Dressel[5]

To cite: Ulyte A, Wei W, Gruebner O, *et al*. Are weak or negative clinical recommendations associated with higher geographical variation in utilisation than strong or positive recommendations? Cross-sectional study of 24 healthcare services. *BMJ Open* 2021;**11**:e044090. doi:10.1136/bmjopen-2020-044090

► Prepublication history and supplemental material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-044090).

Received 24 August 2020
Revised 19 February 2021
Accepted 20 April 2021

For numbered affiliations see end of article.

Correspondence to
Dr Agne Ulyte;
agne.ulyte@uzh.ch

ABSTRACT

Objectives When research evidence is lacking, patient and provider preferences, expected to vary geographically, might have a stronger role in clinical decisions. We investigated whether the strength or the direction of recommendation is associated with the degree of geographic variation in utilisation.

Design In this cross-sectional study, we selected 24 services following a comprehensive approach. The strength and direction of recommendations were assessed in duplicate. Multilevel models were used to adjust for demographic and clinical characteristics and estimate unwarranted variation.

Setting Observational study of claims to mandatory health insurance in Switzerland in 2014.

Participants Enrolees eligible for the 24 healthcare services.

Primary outcome measures The variances of regional random effects, also expressed as median odds ratios (MOR). Services grouped by strength and direction of recommendations were compared with Welch's t-test.

Results The sizes of the eligible populations ranged from 1992 to 409 960 patients. MOR ranged between 1.13 for aspirin in secondary prevention of myocardial infarction to 1.68 for minor surgical procedures performed in inpatient instead of outpatient settings. Services with weak recommendations had a negligibly higher variance and MOR (difference in means (95% CI) 0.03 (−0.06 to 0.11) and 0.05 (−0.11 to 0.21), respectively) compared with strong recommendations. Services with negative recommendations had a slightly higher variance and MOR (difference in means (95% CI) 0.07 (−0.03 to 0.18) and 0.14 (−0.06 to 0.34), respectively) compared with positive recommendations.

Conclusions In this exploratory study, the geographical variation in the utilisation of services associated with strong vs weak and negative vs positive recommendations was not substantially different, although the difference was somewhat larger for negative vs positive recommendations. The relationships between the strength or direction of recommendations and the variation may be indirect or modified by other characteristics of services. As initiatives discouraging low-value care are gaining attention worldwide, these findings may inform future research in this area.

Strengths and limitations of this study

- ► Although the strength and direction of recommendations is generally expected to influence the variation in clinical decisions, this is the first study to analyse this relationship quantitatively.
- ► The effect of the strength and direction of a recommendation on the geographical variation in healthcare utilisation was assessed within a comprehensive set of 24 healthcare services.
- ► Unwarranted variation of the services utilisation was extracted with a single standard approach.
- ► Indirect relationship and modifiers of the effect could not be studied.

BACKGROUND

According to the evidence-based medicine (EBM) framework, clinical decisions should be guided by research evidence, clinical circumstances, and patient preferences and be integrated with clinical expertise.[1] If evidence is weak or lacking, patient preferences and clinical expertise have a particularly strong role in the decision.[2 3] In a clinical practice guideline, such a situation would be reflected by a weak recommendation.[2] As patient preferences tend to vary geographically,[4] and physician practice styles are also significantly influenced by the region of practice,[5 6] clinical decisions associated with less

conclusive research evidence or weak recommendations may have higher geographical variation.

Surprisingly, there is little direct evidence whether weak recommendations are, in fact, associated with higher variation. The few available studies focused on a single specialty and did not quantify the variation in a uniform way, complicating the comparison of results.[7 8] Therefore, despite many studies highlighting the substantial geographic variation in the utilisation of various healthcare services,[9–11] it is not clear, what role the components of the EBM framework play for this variation.

A second potential contributor to different services having different degrees of geographic variation is the direction of recommendation: positive (prescriptive) or negative (proscriptive), as in Choosing Wisely recommendations.[12] Negative recommendations usually concern long-used low-value practices and are based on the lack of supporting evidence or evidence of harms.[12–14] In contrast to positive recommendations, which often introduce new services or indications, negative recommendations usually challenge existing practices that are justified primarily by clinical expertise and judgement, expected to vary regionally.[4] Positive and negative recommendations have different perceived barriers to their implementation,[15] which could contribute to different variation patterns as well. However, no study has directly compared the geographic variation associated with positive and negative recommendations.

The primary aim of this study was to assess whether healthcare services with weak recommendations are associated with higher geographical variation in utilisation. In addition, a secondary aim was to test the association of geographic variation with the direction of the recommendation.

METHODS

Study hypotheses

Although this is primarily an explorative study, we formulated two specific hypotheses. The primary hypothesis of the study was that healthcare services with weaker evidence, as reflected in weak recommendations in clinical guidelines, would have higher geographical variation in utilisation than those with strong recommendations. The secondary hypothesis was that services with negative (proscriptive) recommendations would have higher geographic variation compared with those with positive (prescriptive) recommendations.

Selection of studied healthcare services

This study was part of a project assessing the geographic variation of the utilisation of a set of healthcare services in Switzerland.[16] Studied healthcare services were translated from selected recommendation statements in clinical practice guidelines, following a systematic approach. We collected clinical practice guidelines of Swiss, European and applicable international medical societies, used in Switzerland and guiding the care for major non-communicable diseases (as defined by the Swiss Federal Office of Public Health[17]). Recommendation statements from selected clinical practice guidelines were considered pragmatically by the authors according to their clinical relevance, the expected frequency of service use, and the size of the eligible population. Identified recommended or discouraged services were then screened for feasibility of measuring the utilisation in eligible populations with Swiss health insurance claims data, based on an approach described earlier.[18]

We aimed for the selected services to reflect both strong and weak, positive and negative recommendations, as well as different healthcare services types. We focused particularly on outpatient primary healthcare services, as they are relevant to the biggest part of the population. However, we also included some discouraged services outside primary healthcare to extend the spectrum of populations investigated.

The final selection comprised 24 services, including services for screening (N=4), diagnosis (N=6), primary prevention (N=1), treatment (N=4) and secondary prevention (N=9). Definitions of the selected services are provided in online supplemental file 1.

Assessment of recommendations: strength and direction

Once the services were selected, their associated recommendations were formally assessed.

For each service, we selected the guideline in which the service was originally identified, and also looked up corresponding guidelines by the relevant European, American and international clinical societies. From this set of guidelines, we selected for assessment the one that was the most applicable to Switzerland in 2014 (see online supplemental file 2 for the prioritisation algorithm). We did not consider the guidelines of Swiss medical societies, as their quality of reporting is partially low,[19] and they tend to be consistent with European and international guidelines. If the service was initially selected based on a guideline published after 2014, and no applicable guideline could be identified for 2014, the recommendation was automatically considered weak.

Thus, a single recommendation statement was assessed for each service. The assessment was done in duplicate by two authors (AU and HD, both medical doctors). Discordant judgements were resolved with mutual agreement in a discussion. Each recommendation was classified as strong or weak (corresponding to Grading of Recommendations Assessment, Development and Evaluation (GRADE) definition[20]), and positive or negative. The algorithm and criteria for the classification are detailed in online supplemental file 2 for the strength, and in online supplemental file 3 for the direction of a recommendation. The list of guidelines containing the recommendation statements that were assessed is provided in online supplemental file 4.

Swiss health insurance claims data

The utilisation of the selected healthcare services was evaluated using mandatory health insurance claims data from

the Helsana Group, covering approximately 1.2 million people (15% of the Swiss population). Helsana Group is one of several private companies providing mandatory health insurance in Switzerland. Eligible patient populations were identified from the patients enrolled with Helsana in 2014. Patients with incomplete address information, living in nursing homes and receiving reimbursement via lump-sums (masking some outpatient services), asylum seekers, those living outside Switzerland, and Helsana employees were excluded. The data provided by Helsana were anonymised.

Models of geographic variation

The utilisation of each healthcare service was determined for each member of the eligible population (see online supplemental file 1 for definitions of the populations and services). For each service, the resulting binary outcome variable was modelled with a multilevel logistic regression technique, using 106 Swiss MobSpat regions ('mobilité spatiale'), as defined by the Swiss Federal Statistical Office,[21] as the higher level. MobSpat regions are constructed by combining several neighbouring municipalities based on geographic and population mobility criteria, and are often used as intermediate-size units of analysis for scientific and regional policy purposes. Each study participant's residence was assigned to the corresponding MobSpat region.

Fixed effects were estimated for the following explanatory variables: age, sex, number of comorbidities (0, 1, 2 and 3 or more), and clinical characteristics of relevance for specific indicators (see online supplemental file 1). These variables are often viewed as associated with warranted variation.[22] In contrast, we did not adjust for variables associated with unwarranted variation (eg, insurance characteristics or provider density). From each multilevel model, we extracted the variance of the regional random effects, reflecting the potentially unwarranted geographic variation. We also converted the variance to median odds ratios (MORs) for more convenient interpretation[23 24] and plotting. MOR is interpreted as the median odds of service utilisation by two individuals with identical characteristics in two randomly selected regions. As MOR is directly extrapolated from the variance, the ranking of these two parameters coincides.

Statistical analysis of the hypotheses

Variances of the regional random effects of services utilisation from the models were used as data points in the final analysis of the hypotheses. Variances of services associated with weak and strong recommendations, as well as negative and positive recommendations were compared with Welch's unequal variances t-test. Mean differences and 95% CIs were presented. The same analysis was also performed for MOR, to improve interpretability of detected group differences.

Although the number of the services analysed was rather small (24), the distribution of the analysed variances was deemed sufficiently close to normal to warrant the use of parametric tests. To account for the small and unequal sample sizes, we used Welch's t-test, which is considered more robust in this setting.[25] CIs were not adjusted for multiple testing.

Statistical analyses were performed using R V.3.6.0[26] and MLwiN V.3.01[27] integrated in STATA V.14.2 using the runmlwin package.[28]

Patient and public involvement

This study was performed as part of the National Research Programme 74 'Smarter Healthcare' of the Swiss National Science Foundations. Patients and public, including policy-makers and healthcare services providers, are involved in interpreting, disseminating and translating the overall results of studies conducted under this programme. Representatives of patients, healthcare providers, health insurers and healthcare policy-makers are members of the advisory board of the project. They provided feedback on the planned study design and its preliminary results. Individual patients were not directly involved in the planning and conducting of this study.

RESULTS

Characteristics of the eligible populations and the geographic variation of the services are shown in table 1. Across the services, the sizes of the eligible populations ranged from 1 992 patients with a new disease-modifying antirheumatic drug prescription to 409 960 patients with recommended influenza vaccination. MOR, reflecting potentially unwarranted geographic variation in utilisation of the services, ranged from 1.13 (1.02–1.29) for aspirin in secondary prevention of myocardial infarction (MI) to 1.68 (1.53–1.87) for minor surgical procedures performed in inpatient instead of outpatient settings.

For three services, a major guideline relevant in 2014 in Switzerland could not be identified (long-term use of proton pump inhibitors, minor inpatient surgery procedures, elective Caesarean section). A total of eight services had weak, and six services had negative underlying recommendations. MOR was 1.29 for services with weak and 1.25 for services with strong recommendations; 1.26 for services with positive and 1.46 for services with negative recommendations (figure 1).

Based on Welch's t-test, the difference in mean variances (95% CI) of services with weak and strong recommendations was 0.03 (−0.06 to 0.11), and the difference in mean MOR was 0.05 (−0.11 to 0.21). The difference in mean variances (95% CI) of services with negative and positive recommendations was 0.07 (−0.03 to 0.18) and the difference in mean MOR was 0.14 (−0.06 to 0.34).

DISCUSSION

We did not find a direct association between the strength of clinical recommendation and the geographical variation in the utilisation of 24 healthcare services. The geographical variation in the utilisation of services with underlying negative recommendations was slightly higher

Table 1 Characteristics of the recommended or discouraged healthcare services studied

Category	Healthcare service (abbreviated)	Utilisation in eligible population	Eligible population			Recommendation		Random effects in multilevel model	
			Total N	Mean age (SD)	Female N (%)	Strength	Direction	Variance	Median OR (MOR)
Screening	Colon cancer screening	5.9%	276387	58.6 (5.8)	142675 (51.6)	Strong	Positive	0.04 (0.03–0.06)	1.21 (1.17–1.26)
	Breast cancer screening	20.9%	178145	61.0 (7.2)	178145 (100)	Weak	Positive	0.22 (0.16–0.29)	1.56 (1.47–1.67)
	Prostate cancer screening	28.4%	145874	59.1 (6.2)	0 (0)	Weak	Positive	0.07 (0.05–0.10)	1.29 (1.25–1.35)
	Osteoporosis screening	4.9%	60812	72.6 (8.7)	60812 (100)	Weak	Positive	0.08 (0.04–0.13)	1.31 (1.22–1.41)
Diagnosis	DM: HbA1c test	69.6%	49198	66.6 (13.0)	22138 (45.0)	Strong	Positive	0.17 (0.12–0.23)	1.48 (1.40–1.58)
	DM: renal function test	44.3%	49198	66.6 (13.0)	22138 (45.0)	Strong	Positive	0.06 (0.04–0.09)	1.27 (1.22–1.33)
	DM: LDL test	44.3%	33975	60.1 (11.2)	13977 (41.2)	Strong	Positive	0.13 (0.09–0.19)	1.42 (1.34–1.51)
	DM: eye examination	55.5%	49198	66.6 (13.0)	22138 (45.0)	Weak	Positive	0.07 (0.05–0.10)	1.29 (1.24–1.35)
	TSH screening	76.1%	169232	56.8 (18.5)	111847 (66.1)	Strong	Negative	0.18 (0.13–0.25)	1.50 (1.42–1.61)
	POCR	13.0%	47215	60.3 (17.2)	27086 (57.4)	Strong	Negative	0.18 (0.13–0.26)	1.50 (1.40–1.62)
Primary prevention	Influenza vaccination	20.9%	409960	64.1 (16.3)	230202 (56.2)	Strong	Positive	0.04 (0.03–0.05)	1.20 (1.17–1.24)
Treatment	**Benzodiazepines**	18.6%	243951	75.0 (7.6)	141986 (58.2)	Strong	Negative	0.14 (0.10–0.18)	1.42 (1.36–1.50)
	Proton pump inhibitors	55.5%	153523	55.7 (17.8)	93543 (60.9)	Weak	Negative	0.02 (0.02–0.03)	1.16 (1.13–1.19)
	Inpatient procedures	61.4%	10656	50.5 (13.7)	7719 (72.4)	Weak	Negative	0.30 (0.20–0.43)	1.68 (1.53–1.87)
	Caesarean section	28.5%	9449	31.9 (5.1)	9449 (100)	Weak	Negative	0.05 (0.02–0.09)	1.24 (1.16–1.34)
Secondary prevention	AMI: aspirin	47.0%	2232	72.4 (13.7)	801 (35.9)	Strong	Positive	0.02 (0.00–0.07)	1.13 (1.02–1.29)
	AMI: statin	34.2%	2232	72.4 (13.7)	801 (35.9)	Strong	Positive	0.14 (0.06–0.27)	1.43 (1.25–1.63)
	AMI: beta-blocker	42.1%	2232	72.4 (13.7)	801 (35.9)	Strong	Positive	0.05 (0.00–0.13)	1.25 (1.05–1.40)
	AMI: ACE/ARB	43.8%	2232	72.4 (13.7)	801 (35.9)	Strong	Positive	0.04 (0.00–0.12)	1.21 (1.03–1.39)
	AMI: P2Y12 inhibitors	46.8%	2232	72.4 (13.7)	801 (35.9)	Strong	Positive	0.03 (0.00–0.10)	1.18 (1.04–1.36)
	PPI with NSAID	43.5%	95072	61.0 (16.2)	60804 (64.0)	Strong	Positive	0.02 (0.01–0.03)	1.15 (1.12–1.18)
	PAD: statin	28.5%	23868	63.6 (16.5)	12113 (50.7)	Strong	Positive	0.04 (0.03–0.07)	1.22 (1.17–1.28)
	Afib: anticoagulation	27.5%	8291	80.8 (7.9)	4037 (48.7)	Strong	Positive	0.05 (0.02–0.09)	1.24 (1.16–1.33)
	GCC with new DMARD	58.7%	1992	59.2 (15.3)	1369 (68.7)	Weak	Positive	0.06 (0.01–0.18)	1.27 (1.07–1.49)

Healthcare services, highlighted in bold, are associated with a negative recommendation. Utilisation was assessed within 1 year, 2014, including for services that are recommended less frequently (eg, colon cancer screening).
Afib, atrial fibrillation; AMI, acute myocardial infarction; ARB, angiotensin II receptor blockers; DM, diabetes mellitus; DMARD, disease-modifying antirheumatic drug; GCC, glucocorticosteroid drugs; HbA1c, glycated haemoglobin; LDL, low density lipid; NSAID, non-steroidal anti-inflammatory drugs; PAD, peripheral artery disease; POCR, preoperative chest radiography; PPI, proton pump inhibitors; TSH, thyroid-stimulating hormone.

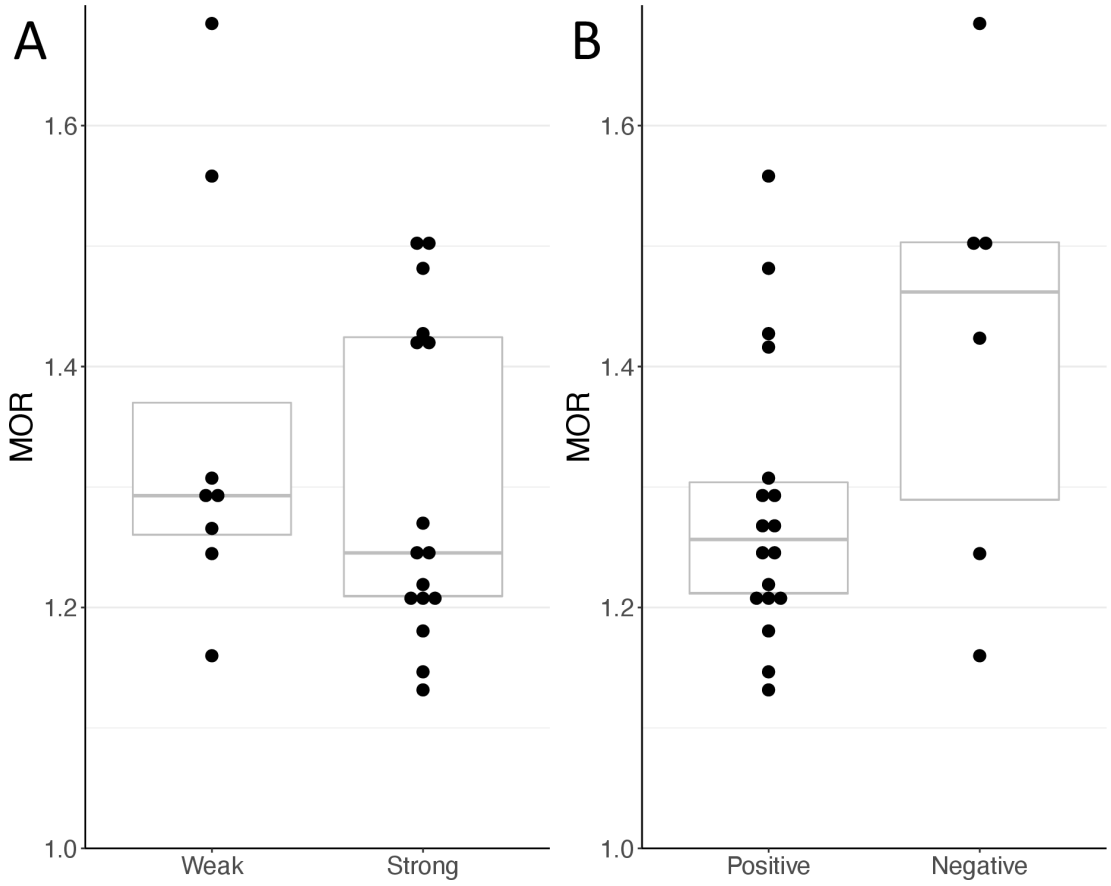


Figure 1 Geographical variation of the healthcare services grouped by strength (A) and direction (B) of recommendations. (A) weak and strong recommendations; (B) positive and negative recommendations. Boxplots depict the interquartile range values (upper and lower hinges) and the median value. MOR, median OR.

than for those with positive recommendations, and for services with underlying weak recommendations than for those with strong recommendations. The difference was larger for negative versus positive recommendations; however, both differences were not statistically significant. In general, moderate potentially unwarranted geographical variation was observed, with MOR smaller than 1.50 for all but one service.

At least two other studies have to some extent examined the association between the strength of recommendations and the variation in adherence, each focusing on a single clinical specialty. In *et al*[7], examining a set of recommendations in oncology, found higher variation in the utilisation of services associated with a lower level of evidence. However, this study focused not on regional but on interinstitutional variation, comparing two groups of providers. In contrast to this study and in agreement with our results, Mayer *et al*[8] found that surgeon practices in knee and hip arthroplasty in Australia varied regardless of the strength of evidence available. Potentially, different clinical areas could be associated with different barriers to guideline implementation, modifying the relationship between recommendations and variation.

To better understand why a direct association of recommendation strength and variation in adherence was not observed, it may be useful to revisit the EBM framework.[1] The EBM framework is normative and defines how clinical decisions *should* be made.[1 29] However, this may not always coincide with how decisions *are* made—a process analysed by descriptive theories.[29] In fact, the EBM model has been developed as conceptual rather than practical guidance of evidence implementation,[1] and has not yet generated a coherent theory of clinical decision making, and in particular, of how evidence is incorporated.[30] Thus, although a direct relationship between the strength of recommendation and the geographical variation of service utilisation would be encouraged by the normative EBM framework, it may not always be observed.

There are numerous reasons why even strong recommendations,[31 32] or conclusive research evidence more broadly,[33] may not directly translate into clinical practice. Research on knowledge translation has identified multiple barriers at different levels of the healthcare system, including structural, organisational, peer-group and professional factors[34] – many of which depend on the specific context where a service is provided, and thus may vary geographically. Knowledge transfer processes are highly non-linear and rely on triggering mechanisms.[35] Factors external to research evidence significantly affect translation—potentially creating large geographic heterogeneity even within services with strong recommendations. Finally, strong recommendations sometimes

describe care with varying patient preferences. For example, although colon cancer screening is strongly recommended, patient preferences for test attributes and modalities vary significantly.[36 37]

An influential framework, explaining different degrees of variation between healthcare services, has been proposed by Wennberg *et al*.[38] According to this framework, services are classified into effective, preference-sensitive, and supply-sensitive care. Effective care (services based on solid evidence, so that virtually all patients would choose them) largely corresponds to services with strong recommendations, as defined by GRADE and applied in this study.[2] Preference-sensitive care partly corresponds to services with weak recommendations, as they both imply trade-offs of risks and benefits of multiple options of care.[38] The utilisation of preference-sensitive surgical procedures usually has higher variation than of those associated with effective care.[39] In contrast, supply-sensitive care defines the frequency, setting and intensity of care provision rather than specific types of healthcare services. It is associated with high, supply-related variation, but is rarely discussed in guidelines,[39] and therefore, could not be included in our study. However, the service of minor surgeries performed as inpatient instead of outpatient procedures could be considered close to the supply-sensitive category. In fact, it had the highest MOR (1.68) in our study.

Regarding the secondary hypothesis, we found that services associated with negative recommendations had slightly higher geographic variation. We did not find other studies directly comparing the regional variation of services with the direction of recommendations. Few studies, focusing mostly on low-value care, have reported MOR as an expression of geographical variation, further limiting the comparison. For example, in a study by Badgery-Parker *et al*.[40] services discouraged by Choosing Wisely were shown to have regional MOR from 1.1 to 2.6—a range that includes all of our observed MORs.

Negative recommendations usually address a widespread service that lacks supporting evidence of benefit or the benefit is outweighed by harms.[2] In contrast to services with positive recommendations, which are introduced after supporting evidence is produced, services with negative recommendations typically become part of the clinical practice before evidence is sufficient to rule out their overall benefit. Therefore, their use could be related more to clinical expertise and practice, and could be expected to vary locally. Indeed, the barriers to implementing positive and negative recommendations seem to be different[15]—signalling that the pathways how they are interpreted and integrated into clinical decisions might also be different. As Choosing Wisely and similar initiatives are increasingly gaining attention,[41] our finding of higher geographic variation associated with negative recommendations may inform future research and implementation strategies.

This exploratory study has several limitations. First, although we aimed at a balanced selection of clinical fields and service types, the number (24) and range of studied services was limited by the data source, leading to somewhat unbalanced groups of strong and weak, positive and negative recommendations. Swiss claims data lack information on outpatient diagnoses, inpatient treatment details, and clinical information such as test results.[18] Lack of clinical information also meant that some populations were not as specific as defined by the recommendation. For example, beta-blockers and ACE inhibitors are recommended after an MI contingent on heart failure and left ventricular dysfunction. As such clinical details were unavailable, we had to rely on them being present in the majority of the hospitalised MI cases and distributed equally geographically. However, we believe that estimates of variation are accurate, as each of the 24 data points was generated by multilevel modelling of utilisation in populations of 85 000 patients on average, including all major explanatory variables such as age, sex and indicators of morbidity. Second, the services studied were unavoidably different by characteristics other than the strength or direction of recommendation, such as service type or clinical area, potentially resulting in confounding. Indeed, although distributed among all recommendation types, diagnostic services had somewhat higher regional variation in utilisation compared with treatment services (see online supplemental file 5). Although most of the selected services are delivered by primary care providers, their varied nature also means that the applied MobSpat regional units might not capture the regional variation equally well. Third, both the observed utilisation and its geographical variation depend on the definition of the service and population.[42] We aimed to measure the unwarranted variation in utilisation by using service-specific denominators (eligible populations) and adjusting for relevant clinical characteristics. How exactly unwarranted and warranted variation should be defined and measured, and what adjustments are necessary to differentiate them, is debated.[22 43] Fourth, the grouping of recommendations by strength and direction was partly subjective, although we tried to make it reproducible with a clear algorithm, implemented in duplicate. Unfortunately, many different systems for evaluating the strength of recommendations exist,[44] which cannot be easily reconciled, and the most prominent, GRADE approach, is not always explicitly used.

To explore the studied questions further, the sample of services could be expanded to inpatient and specialist care. Further, a meta-study of the numerous individual studies of geographic variation in healthcare services could be undertaken. However, this would currently be challenging, as studies choose different adjustment variables and specificity of studied populations, and report the variation in different quantitative forms (eg, MOR, systematic component of variation, range). Furthermore, there is a need for qualitative studies of the reasons for the variability of clinical decisions and how clinical expertise in these decisions interacts with evidence, clinical circumstances and patient preferences. Qualitative

evidence could help to generate more complex hypotheses for further quantitative studies, built on a finer understanding of all the factors influencing the variability of clinical decisions. Specialty-specific sets of services could also be further investigated.

CONCLUSIONS

In this exploratory study of 24 healthcare services mostly in the outpatient primary care setting, we did not observe a significant difference in the degree of geographic variation in utilisation of services associated with strong versus weak recommendations. Services associated with negative recommendations had slightly, although also not statistically significantly, higher geographical variation. The relationship between the strength of recommendations and the variation may be indirect or modified by other characteristics of services, such as service type or clinical area. As initiatives discouraging low-value care are gaining attention worldwide, these findings may inform future research in this area.

Author affiliations
[1]Department of Epidemiology, Epidemiology, Biostatistics & Prevention Institute, University of Zurich, Zurich, Switzerland
[2]Department of Geography, University of Zurich, Zurich, Switzerland
[3]Department of Health Sciences, Helsana Group, Zurich, Switzerland
[4]Institute of Primary Care, University of Zurich and University Hospital Zurich, Zurich, Switzerland
[5]Division of Occupational and Environmental Medicine, Department of Epidemiology, Epidemiology, Biostatistics & Prevention Institute, University of Zurich and University Hospital Zurich, Zurich, Switzerland

Contributors MS, VvW and HD developed the underlying study program. AU and HD developed study design, with support from all authors. AU, WW, CB, EB, BB did data extraction, preparation and management. WW, MS and OG developed multilevel models applied in this study. AU and WW analysed the data. AU drafted the manuscript, with major inputs from HD and MS, and contributions from all authors. All authors read and approved the final manuscript.

Funding This work was supported by the Swiss National Science Foundation (SNSF) National Research Program 'Smarter Health Care' (NRP 74), as part of project number 26, grant number 407440_167349.

Competing interests MS declares a grant from Helsana Insurance Group, outside the submitted work. Helsana Group provided support in the form of salaries for authors BB, EB and CB, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The other authors declare no competing interests.

Patient consent for publication Not required.

Ethics approval According to the national ethical and legal regulations, ethical approval was not needed for this analysis of anonymised data. This was confirmed by a waiver of the competent ethics committee (Kantonale Ethikkommission Zürich, dated January 11, 2017, BASEC-Nr. Req-2017–00011).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data underlying this study cannot be shared publicly because they are the property of Helsana (https://www.helsana.ch/en/helsana-group), and have restricted public access on grounds of patient privacy. The data are managed by Helsana and subsets of the database are available for researchers after request and under specific conditions. Data are available from Helsana (gesundheitskompetenz@helsana.ch) for researchers who meet the criteria for access to confidential data. Helsana will consider the possibilities of the research proposal and decide to grant access if the research questions can be answered with use of the Helsana data. When requests are granted, data are accessible only in a secure environment.

ORCID iDs
Agne Ulyte http://orcid.org/0000-0001-7419-9778
Beat Brüngger http://orcid.org/0000-0001-6173-5375

REFERENCES

1 Haynes RB, Devereaux PJ, Guyatt GH. Clinical expertise in the era of evidence-based medicine and patient choice. *Evid Based Med* 2002;7:36–8.
2 Andrews J, Guyatt G, Oxman AD, *et al*. Grade guidelines: 14. going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66:719–25.
3 Andrews JC, Schünemann HJ, Oxman AD, *et al*. Grade guidelines: 15. going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726–35.
4 Cutler D, Skinner JS, Stern AD, *et al*. Physician beliefs and patient preferences: a new look at regional variation in health care Spendingf. *Am Econ J Econ Policy* 2019;11:192–221.
5 Finkelstein A, Gentzkow M, Williams H. Sources of geographic variation in health care: evidence from patient migration. *Q J Econ* 2016;131:1681–726.
6 Molitor D. The evolution of physician practice styles: evidence from cardiologist migration. *Am Econ J Econ Policy* 2018;10:326–56.
7 In H, Neville BA, Lipsitz SR, *et al*. The role of national cancer Institute-designated cancer center status: observed variation in surgical care depends on the level of evidence. *Ann Surg* 2012;255:890–5.
8 Mayer M, Naylor J, Harris I, *et al*. Evidence base and practice variation in acute care processes for knee and hip arthroplasty surgeries. *PLoS One* 2017;12:e0180090.
9 Corallo AN, Croxford R, Goodman DC, *et al*. A systematic review of medical practice variation in OECD countries. *Health Policy* 2014;114:5–14.
10 Dartmouth Atlas Project. The Dartmouth atlas of health care, 2020. Available: https://www.dartmouthatlas.org/
11 Public Health England. Atlas of variation, 2020. Available: https://fingertips.phe.org.uk/profile/atlas-of-variation
12 Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA* 2012;307:1801–2.
13 Elshaug AG, Watt AM, Mundy L, *et al*. Over 150 potentially low-value health care practices: an Australian study. *Med J Aust* 2012;197:556–60.
14 Prasad V, Vandross A, Toomey C, *et al*. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc* 2013;88:790–8.
15 Carlsen B, Glenton C, Pope C. Thou shalt versus thou shalt not: a meta-synthesis of GPs' attitudes to clinical practice guidelines. *Br J Gen Pract* 2007;57:971–8.
16 National Research Programme 74. Project 26: how do guidelines and recommendations influence medical treatment? 2020. Available: http://www.nfp74.ch/en/projects/healthcare-across-sectors/project-schwenkglenks
17 Swiss Federal Office of Public Health and Swiss Conference of Cantonal Health Directors. Nationale Strategie Prävention nichtübertragbarer Krankheiten (NCD-Strategie) 2017-2024 [National Strategy of Non-communicable Diseases] 2016.
18 Ulyte A, Bähler C, Schwenkglenks M, *et al*. Measuring diabetes guideline adherence with claims data: systematic construction of indicators and related challenges. *BMJ Open* 2019;9:e027138.

19 Bachmann L, Ulytė A, Dressel H. Clinical practice guidelines of medical societies in Switzerland: analysis of the current state. *Swiss Med Wkly* 2019;149:w20134.
20 Guyatt GH, Oxman AD, Kunz R, *et al*. Going from evidence to recommendations. *BMJ* 2008;336:1049–51.
21 Bundesamt für Statistik. MS-Regionen, 2020. Available: https://www.bfs.admin.ch/bfs/de/home/statistiken/raum-umwelt/nomenklaturen/msreg.assetdetail.415729.html
22 Sutherland K, Levesque J-F. Unwarranted clinical variation in health care: definitions and proposal of an analytic framework. *J Eval Clin Pract* 2020;26:687–96.
23 Larsen K, Petersen JH, Budtz-Jørgensen E, *et al*. Interpreting parameters in the logistic regression model with random effects. *Biometrics* 2000;56:909–14.
24 Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *Am J Epidemiol* 2005;161:81–8.
25 Ruxton GD. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behav Ecol* 2006;17:688–90.
26 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. https://www.r-project.org/
27 Charlton C, Rasbash J, Browne WJ. *MLwiN version 3.01. centre for multilevel modelling*. Bristol: University of Bristol, 2017.
28 Leckie G, Charlton C. Runmlwin: a program to run the MLwiN multilevel modeling software from within Stata. *J Stat Softw* 2013;52:1–40.
29 Djulbegovic B, Elqayam S, Dale W. Rational decision making in medicine: implications for overuse and underuse. *J Eval Clin Pract* 2018;24:655–65.
30 Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet* 2017;390:415–23.
31 Francke AL, Smit MC, de Veer AJE, *et al*. Factors influencing the implementation of clinical guidelines for health care professionals: a systematic meta-review. *BMC Med Inform Decis Mak* 2008;8:38.
32 Lugtenberg M, Zegers-van Schaick JM, Westert GP, *et al*. Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners. *Implementation Sci* 2009;4:1–9.
33 Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. *BMJ Open* 2012;2:e001220.
34 Grimshaw JM, Eccles MP, Lavis JN, *et al*. Knowledge translation of research findings. *Implement Sci* 2012;7:50.
35 Braithwaite J, Churruca K, Long JC, *et al*. When complexity science meets implementation science: a theoretical and empirical analysis of systems change. *BMC Med* 2018;16:63.
36 Marshall DA, Johnson FR, Phillips KA, *et al*. Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value Health* 2007;10:415–30.
37 Mansfield C, Tangka FKL, Ekwueme DU, *et al*. Stated preference for cancer screening: a systematic review of the literature, 1990-2013. *Prev Chronic Dis* 2016;13:E27.
38 Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. *Health Affairs* 2003;22:W96–114.
39 Wennberg JE. Tracking medicine : a researcher's quest to understand health care. Oxford University Press 2010.
40 Badgery-Parker T, Feng Y, Pearson S-A, *et al*. Exploring variation in low-value care: a multilevel modelling study. *BMC Health Serv Res* 2019;19:345.
41 Levinson W, Kallewaard M, Bhatia RS, *et al*. 'Choosing wisely': a growing international campaign. *BMJ Qual Saf* 2015;24:167–74.
42 Schwartz AL, Landon BE, Elshaug AG, *et al*. Measuring low-value care in Medicare. *JAMA Intern Med* 2014;174:1067–76.
43 Harrison R, Manias E, Mears S, *et al*. Addressing unwarranted clinical variation: a rapid review of current evidence. *J Eval Clin Pract* 2019;25:53–65.
44 Atkins D, Eccles M, Flottorp S, *et al*. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the grade Working group. *BMC Health Serv Res* 2004;4:1–7.

Additional file 1 Definitions and descriptions of the studied health care services and eligible populations

Category	Health care service	Service description and frequency	Eligible population	Recommendation for the health care service	Specific clinical explanatory variables	Clinical codes used for identification of the health care service
Screening	Colon cancer screening	Colonoscopy/ year	Anyone 50-69 years old	Colonoscopy should be done every 10 years for people 50-69 years old.	Previous treatment of cancer or inflammatory bowel disease, hospitalization with colon disease in the last year	Colonoscopy: 19.06 (TM Kapitel); G48% (DRG); 45.23, 45.25, 48.29.1%, 48.29.2% (CHOP)
	Breast cancer screening	Mammography/ year	50-74 years old women	Mammography should be done every 2 years for 50-74 years old women.	Previous treatment of breast or other cancer	Mamography: 39.1310, 39.1320, 39.1307, 39.1308, 39.1300, 39.1305, 39.1306 (TM); TZ
	Prostate cancer screening	Prostate-specific antigen (PSA) testing/ year	50-70 years old men	Early detection of prostate cancer (opportunistic screening) should be offered to the well-informed man.	Previous treatment of cancer, hospitalization with prostate disease in the last year	PSA testing: 1626.00 (Ana)
	Osteoporosis screening	Dual-energy x-ray absorptiometry (DXA)/ year	Women over 60 and with risk factors[a] of spontaneous fractures	DXA densitometry is recommended for postmenopausal women with spontaneous fractures or increased risk of them.	Presence of more than one risk factor	DXA densitometry: 39.1950, 39.2140, 39.2150, 39.2160 (TM)

Diagnosis	DM: HbA1c test	Glycated haemoglobin (HbA1c) test twice/ year	>18-year-old drug-treated diabetes patients	HbA1c test should be done for diabetes patients at least twice a year.	Oral diabetes medication or insulin	HbA1c test: 1363.00, 1363.01 (Ana)
	DM: renal function test	Albuminuria and serum creatinine tests/ year	>18-year-old drug-treated diabetes patients	Albuminuria and serum creatinine tests should be done for diabetes patients at least once a year.	Oral diabetes medication or insulin	Albuminuria: 1023.00, 1023.01, 1739.00, 1739.01, 1740.00, 1740.01 (Ana) Serum creatinine: 1509.00, 1509.01 (Ana)
	DM: LDL test	Low-density lipoprotein (LDL) test/ year	19-75-year-old drug-treated diabetes patients	LDL test should be done for diabetes patients at least once a year.	Oral diabetes medication or insulin	LDL test: 1521.00 (Ana) Total cholesterol test: 1230.00, 1230.01 (Ana) HDL test: 1410.01, 1410.10 (Ana) Triglycerides test: 1731.01, 1731.00 (Ana)
	DM: eye examination	Ophthalmologist visit/ year	>18-year-old drug-treated diabetes patients	Eye exam should be performed for diabetes patients at least once a year.	Oral diabetes medication or insulin	Outpatient visit with ophthalmologist: (sub group “Ophthalmologie” in Swiss care provider registry sasis.ch)
	TSH screening	Thyroid-stimulating hormone (TSH) test without T3 and T4 tests on the same day	>18-year-old persons without thyroid disease[b] and receiving TSH test	TSH should be measured as an initial screening test for hypo/hyperthyroidism, while T3 and T4 test should follow if TSH is abnormal.	-	TSH test: 1718.10 (Ana) T3 or T4 test: 1732.00, 1720.00, 733.00, 1721.00 (Ana)

	POCR	Outpatient preoperative chest radiography (POCR) up to 2 months before surgery	>18-year-old patients with inpatient surgical procedures	Routine chest radiography is not recommended before surgery.	-	Chest radiography: 39.0190 (TM)
Primary prevention	Influenza vaccination	Influenza outpatient vaccination/ year	People over 65 years old or with a specified chronic condition[c]	People over 65 years old and patients with chronic conditions, specified by Federal Office of Public Health, should be vaccinated against influenza every year.	Hospitalization with pneumonia in the last year	Influenza vaccination: J07BB02 (ATC)
Treatment	Benzodiazepines	Cumulative prescription of benzodiazepines (BZD) for >8 weeks/ year	Anyone over 65 years old	Long-term use of benzodiazepines and other hypnotics is discouraged for old patients.	Treated epilepsy, stay in a nursing home in the last year, hospitalization in the last year with a diagnosis indicative of justified benzodiazepine use	Benzodiazepines and other hypnotics: N03AE01, N05BA%, N05CD%, N05BB%, N05BE%, N05CA%, N05CB%, N05CC%, N05CF%, N05CH%, N05CM%, N05CX% (ATC)
	Proton pump inhibitors	Cumulative prescription of proton pump inhibitors (PPI) or H2 histamine receptor antagonists (H2) for >8 weeks/ year	>18-year-old persons receiving PPI or H2 drugs	PPI should not be used at maximal dose for prolonged periods of time.	-	PPI or H2: A02BC%, A02BD%, M01AE52, A02BA% (ATC)
	Inpatient procedures	Specified surgical procedures[d] done in	>18-year-old patients with	If none of the special conditions apply, certain surgical	-	

		the outpatient setting	specified surgical procedures (either as in- or outpatient)	procedures should be done in the outpatient setting.		
	Caesarean section	Caesarean section (C-section)	>18-year-old women giving birth without absolute indications[e] for C-section	C-section should not be performed unless absolute or relative indications are present.	-	C-section: 74.0%, 74.1%, 74.2%, 74.4%, 74.99 (CHOP); O01A, O01B, O01C, O01D, O01E, O01F (DRG); 22.2120, 22.2130, 22.2410, 22.2420 (TM)
Secondary prevention	AMI: aspirin	Aspirin prescription within 2 weeks after acute myocardial infarction (AMI)	>18-year-old patients with AMI[f]	All myocardial infarction patients should take aspirin long-term.	Hospitalization for stroke or bleeding event or prescribed anticoagulation in the last year	Aspirin: B01AC06 (ATC)
	AMI: statin	High-dose statin prescription within 2 weeks after AMI	>18-year-old patients with AMI[f]	All myocardial infarction patients should get statins long-term.	Hospitalization for stroke in the last year	High-dose statins: C10AA05, C10AA07 (ATC)
	AMI: beta-blocker	Beta-blocker prescription within 2 weeks after AMI	>18-year-old patients with AMI[f]	All myocardial infarction patients with heart failure or impaired function should get beta-blockers long-term.	Hospitalization with heart failure diagnosis in the last year	Beta-blockers: C07% (ATC)
	AMI: ACE/ARB	Angiotensin-converting enzyme (ACE) or angiotensin	>18-year-old patients with AMI[f]	All myocardial infarction patients with heart failure or impaired function should get	-	ACE or ARB medication: C09% (ATC)

		receptor blocker (ARB) antihypertensive medication prescription within 2 weeks after AMI		ACE or ARB antihypertensive medication long-term.		
	AMI: P2Y12 inhibitors	P2Y12 antiplatelet drug[g] prescription within 2 weeks after AMI	>18-year-old patients with AMI[f]	All myocardial infarction patients should get P2Y12 antiplatelet drugs for at least 1-12 months according to the bleeding risk profile and AMI treatment.	Hospitalization for a bleeding event or prescribed anticoagulation in the last year	P2Y12 drugs: B01AC04, B01AC22, B01AC24 (ATC)
	PPI with NSAID	PPI prescription within 1 month or up to 3 months before initial long-term nonsteroidal anti-inflammatory drug (NSAID) prescription	>18-year-old patients with a cumulative NSAID prescription of >8 weeks at maximal dose	Patients taking long-term NSAID and with risk factors for gastric ulcer[h] should also take PPI.	Concurrent use of antiplatelet, anticoagulation drugs or oral glucocorticoids, hospitalization for bleeding event in the last year.	NSAID: M01A% (ATC) PPI: A02BC%, A02BD%, M01AE52 (ATC)
	PAD: statin	Prescription of statins within 3 months after peripheral artery disease (PAD) identification	>18-year-old patients undergoing diagnostic or treatment	Statins are recommended for all patients with PAD.	-	Statins: C10AA%, C10B% (ATC)

			procedures for PAD[i]			
	Afib: anticoagulation	Oral anticoagulation prescription within 2 weeks after atrial fibrillation (Afib) identification	>18-year-old patients with atrial fibrillation diagnosis and additional risk factors[j]	All patients with atrial fibrillation should be prescribed oral anticoagulation for embolic events prevention according to the CHA_2DS_2-VASc score.	-	Oral anticoagulation: B01AE07, B01AF01, B01AF02, B01AF03, B01AA04, B01AA07 (ATC)
	GCC with new DMARD	Glucocorticoid (GCC) prescription within 1 month or up to 3 months before disease-modifying antirheumatic drug (DMARD) prescription	>18-year-old patients with a new prescription of DMARD by a rheumatologist	Short-term glucocorticoids should be taken with newly prescribed DMARD.	-	Glucocorticoids: H02% (ATC) DMARD: L01BA01, L04AX03, M01CX01, L04AA13, M01CX02, P1BA02, P01BA01, M01CC01, L01AA01, M01CB01, L04AX01 (ATC)

a. Recent distal radius, proximal humerus, vertebral or femoral fracture, use of drugs increasing the risk of osteoporosis, use of oral glucocorticoids, diabetes, ankylosing spondylitis, osteogenesis imperfecta, rheumatoid arthritis, inflammatory bowel disease, Cushing's disease, alcohol or nicotine abuse, chronic liver disease, gastrectomy, malnutrition, hypogonadism, hyper- or hypothyroidism, and hyperparathyroidism. Patients currently treated or diagnosed with osteoporosis were excluded.

b. Hyperthyroidism, hypothyroidism, goitre or thyroiditis.

c. Cardiovascular disease, chronic pulmonary disease, diabetes, chronic liver disease, renal failure, immune deficiency, systemic neurologic disorders.

d. Varicose veins ligation and stripping, surgical procedures of haemorrhoids, inguinal hernia and cervix, knee arthroscopy and meniscectomy, tonsillectomy.

e. Placental, umbilical cord or fetal pathology, HIV or genital HSV infection, or multiple pregnancy.

f. Inpatient treatment with a diagnosis of acute myocardial infarction (AMI).

g. Clopidogrel, prasugrel or ticagrelor.

h. Concurrent use of antiplatelet, anticoagulant drugs, oral glucocorticoids or recent hospitalization with any major bleeding.

i. Peripheral artery disease (PAD) or carotid stenosis diagnosed during an inpatient stay, amputation of lower or upper extremity, thrombectomy, stenting or other procedures in peripheral arteries, specialized diagnostic ultrasound, magnetic resonance tomography (MRI) angiography, computer tomography (CT) angiography or angiography of peripheral arteries.

j. Risk factors (congestive heart failure, hypertension, age 65-74 or ≥75 years old, diabetes, previous stroke, transient ischemic attack, or thromboembolism, cardiovascular disease, female sex) were extracted from available claims data and summed according to CHA2DS2-VASc score. Patients with CHA2DS2-VASc score of ≥2 for males and ≥3 for females were included.

DM – diabetes mellitus, HbA1c – Glycated haemoglobin, LDL – low density lipid, TSH – thyroid-stimulating hormone, T3 and T4 – triiodothyronine and thyroxine, POCR – preoperative chest radiography, BZD – benzodiazepines, PPI – proton pump inhibitors, H2 – H2 histamine receptor antagonists, C-section – Caesarean section, AMI – acute myocardial infarction, ACE/ARB – angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers, NSAID – nonsteroidal anti-inflammatory drugs, PAD – peripheral artery disease, Afib – atrial fibrillation, GCC – glucocorticosteroid drugs, DMARD – disease-modifying antirheumatic drug.

Ana – Analysenliste, Swiss outpatient laboratory test codes; ATC - Anatomical Therapeutic Chemical Classification System, code and quantity of a prescription drug; CHOP - Schweizerische Operationsklassifikation, a classification of inpatient procedures; DRG - Swiss Diagnosis Related Groups, a classification of inpatient cases, based on diagnoses, procedures and other clinical information; ICD - International Classification of Diseases, 10th revision, German Modification, codes for primary and secondary diagnoses for each hospitalization episode of an inpatient; TM – Tarmed, Swiss classification of outpatient procedures and services; TM Kapitel – Tarmed chapter codes; TZ – Tarifziffer, further codes representing reimbursement of screening services within cantonal breast cancer screening programs.

Additional file 2 Algorithm and criteria for the assessment of the strength of recommendation

N of authors	Steps	
Single	1.	Identify the relevant medical societies for each selected health care service.
	2.	Look up the European or international medical societies' websites and journals, and identify relevant guidelines published before 2014. In addition, look up if Swiss federal legislation guidelines exist by 2014.
	3.	If none found, look up American medical society and identify relevant guidelines published before 2014.
	4.	If none found, consider the recommendation **weak**.
In duplicate	5.	Once the guideline and the recommendation statement are located, classify the recommendation into strong or weak[a].
	-	**Strong recommendation** implies that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
	-	That means that most informed patients would choose the recommended management and that clinicians can structure their interactions with patients accordingly.
	-	For clinicians, that would mean that most patients should receive the recommended course of action. For patients, that would mean that most people in such a situation would want the recommended course of action and only a small proportion would not; patients should request discussion if the intervention is not offered.
	-	**Weak recommendation** implies that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but the guideline panel is less confident.
	-	Thus, a weak recommendation is conditional or optional, and means that patients' choices will vary according to their values and preferences, and clinicians must ensure that patients' care is in keeping with their values and preferences.
	-	For clinicians, that would mean that they should recognize that different choices will be appropriate for different patients and that they must help each patient to arrive at a management decision consistent with her or his values and preferences. For patients, that would mean that most people in such situation would want the recommended course of action, but many would not.

a adapted from: Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. GRADE: Going from Evidence to Recommendations. BMJ 2008;336:1048–51.

Additional file 3 Algorithm and criteria for the assessment of the direction of recommendation

N of authors	Steps	
In duplicate	1.	Once the guideline and the recommendation statement are located (see Additional file 2), classify the recommendation into positive and negative.
	-	**Positive** recommendation encourages the use of a health care service in a given population.
	-	**Negative** recommendation discourages the use of a health care service in a given population (e.g., contains negative indicatory words, such as *not, no, never*)

Additional file 4 List of guidelines selected for the study, describing the services analysed

Recommendation	Reference	Comment
Colon cancer screening	[1]	
Breast cancer screening	[2]	
Prostate cancer screening	[3]	
Osteoporosis screening	[4]	
DM: HbA1c test	[5]	
DM: renal function test	[5]	
DM: LDL test	[5]	
DM: eye examination	[5]	
TSH screening	[6,7]	
POCR	[8]	
Influenza vaccination	[9]	
Benzodiazepines	[10]	
Proton pump inhibitors	-	Swiss national guideline since 2016 [11]
Inpatient procedures	-	Swiss federal regulation exists from 2019 [12]
Caesarean section	-	Swiss national guideline since 2015 [13]
AMI: aspirin	[14,15]	
AMI: statin	[14,15]	
AMI: beta-blocker	[14,15]	
AMI: ACE/ARB	[14,15]	
AMI: P2Y12 inhibitors	[14,15]	
PPI with NSAID	[16]	
PAD: statin	[17]	
Afib: anticoagulation	[18]	
GCC with new DMARD	[19]	

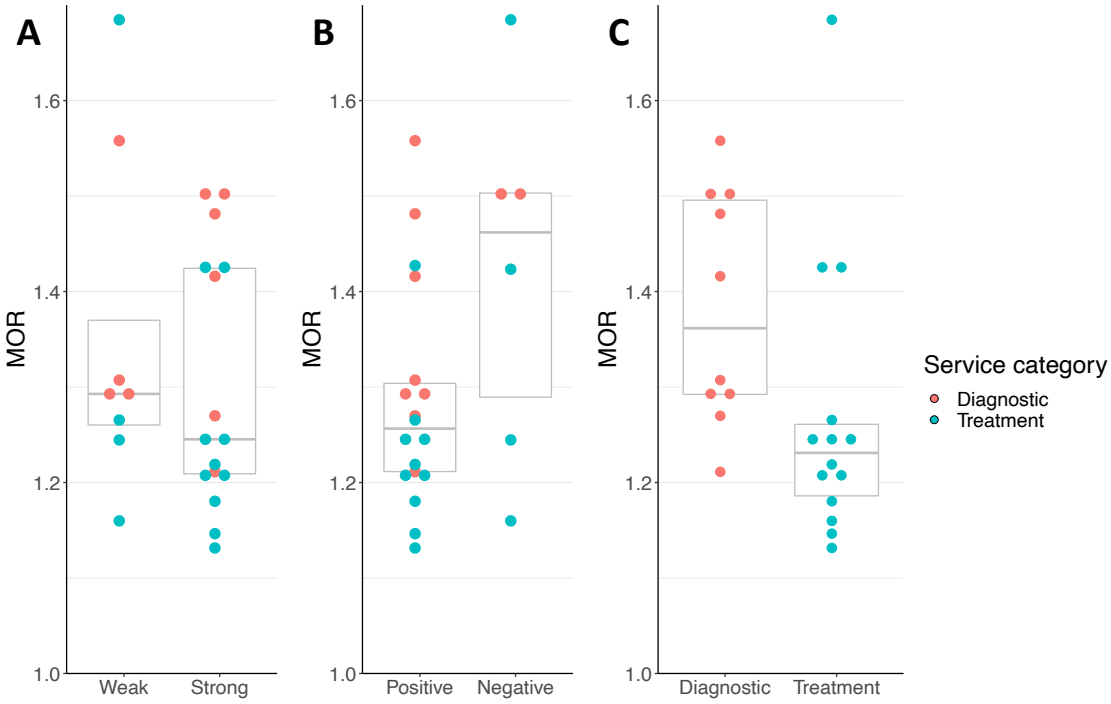
DM – diabetes mellitus, HbA1c – Glycated haemoglobin, LDL – low density lipid, TSH – thyroid-stimulating hormone, POCR – preoperative chest radiography, AMI – acute myocardial infarction, ACE/ARB – angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers, PPI – proton pump inhibitors, NSAID – nonsteroidal anti-inflammatory drugs, PAD – peripheral artery disease, Afib – atrial fibrillation, GCC – glucocorticosteroid drugs, DMARD – disease-modifying antirheumatic drug.

1. Lansdorp-Vogelaar I, Karsa L. European guidelines for quality assurance in colorectal cancer screening and diagnosis. First Edition – Introduction. Endoscopy. 2012 Sep 25;44(S 03):SE15–30.
2. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med. 2009 Nov 17;151(10):716.
3. Heidenreich A, Bellmunt J, Bolla M, Joniau S, Mason M, Matveev V, et al. EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Treatment of Clinically Localised Disease. Eur Urol. 2011 Jan;59(1):61–71.

4. Kanis JA, McCloskey E V., Johansson H, Cooper C, Rizzoli R, Reginster J-Y. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int. 2013 Jan 19;24(1):23–57.
5. Rydén L, Grant PJ, Anker SD, Berne C, Cosentino F, Danchin N, et al. ESC Guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J. 2013 Oct 14;34(39):3035–87.
6. Bahn RS, Burch HB, Cooper DS, Garber JR, Greenlee MC, Klein I, et al. Hyperthyroidism and Other Causes of Thyrotoxicosis: Management Guidelines of the American Thyroid Association and American Association of Clinical Endocrinologists. Thyroid. 2011 Jun;21(6):593–646.
7. Garber J, Cobin R, Gharib H, Hennessey J, Klein I, Mechanick J, et al. Clinical Practice Guidelines for Hypothyroidism in Adults: Cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract. 2012 Nov;18(6):988–1028.
8. Choosing Wisely. American College of Surgeons. Admission pre-op chest x-rays. September 4, 2013. Available from: https://www.choosingwisely.org/clinician-lists/american-college-surgeons-admission-or-preop-chest-x-ray-on-ambulatory-patients/ Accessed on May 11, 2020
9. Swiss Federal Office of Public Health. Recommendations for Influenza Vaccination (Empfehlungen zur Grippeimpfung). 2011.
10. Choosing Wisely. American Geriatrics Society. Benzodiazepines sedative hypnotics for insomnia in older adults. February 21, 2013. Available from: https://www.choosingwisely.org/clinician-lists/american-geriatrics-society-benzodiazepines-sedative-hypnotics-for-insomnia-in-older-adults/ Accessed on May 11, 2020
11. Swiss Society of General Internal Medicine. Smarter Medicine. Top-5 List for Ambulatory Care. 2016. Available from: https://www.smartermedicine.ch/de/top-5-listen/ambulante-allgemeine-innere-medizin.html Accessed on May 11, 2020
12. Swiss Federal Office of Public Health. “Outpatient instead of Inpatient” [Änderung der Krankenpflege-Leistungsverordnung (KLV) betreffend «Ambulant vor Stationär»]. Available from: https://www.bag.admin.ch/bag/de/home/versicherungen/krankenversicherung/krankenversicherung-revisionsprojekte/konsultation-ambulant-vor-stationaer.html Accessed on May 11, 2020
13. Hoesli I, Alma-Stucki S El, Drack G, Girard T, Irion O, Schulzke S, et al. Guideline Sectio Caesarea. 2015;1–20. Available from: https://www.sggg.ch/fachthemen/guidelines/ Accessed on May 11, 2020
14. Hamm CW, Bassand J-P, Agewall S, Bax J, Boersma E, Bueno H, et al. ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevatio. Eur Heart J. 2011 Dec 1;32(23):2999–3054.
15. Steg PG, James SK, Atar D, Badano LP, Lundqvist CB, Borger MA, et al. ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. Eur Heart J. 2012 Oct 1;33(20):2569–619.
16. Recommendations for use of selective and nonselective nonsteroidal antiinflammatory drugs: An American College of Rheumatology white paper. Arthritis Rheum. 2008 Aug 15;59(8):1058–73.

17. Tendera M, Aboyans V, Bartelink M-L, Baumgartner I, Clement D, Collet J-P, et al. ESC Guidelines on the diagnosis and treatment of peripheral artery diseases: Document covering atherosclerotic disease of extracranial carotid and vertebral, mesenteric, renal, upper and lower extremity arteries * The Task Force on the Diagnosis and Treat. Eur Heart J. 2011 Nov 2;32(22):2851–906.
18. Camm AJ, Lip GYH, De Caterina R, Savelieva I, Atar D, Hohnloser SH, et al. 2012 focused update of the ESC Guidelines for the management of atrial fibrillation. Eur Heart J. 2012 Nov 1;33(21):2719–47.
19. Smolen JS, Landewé R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis. 2014 Mar;73(3):492–509.

Additional file 5 Geographic variation of the health care services grouped by strength and direction of recommendations, and service category



A Weak and strong recommendations; B Positive and negative recommendations; C Diagnostic and treatment services. MOR – median odds ratio. Boxplots depict the interquartile range of values (upper and lower hinges), and the median value.

Based on Welch's t-test, the difference in mean variances [95CI%] of diagnostic and treatment services was 0.04 [-0.01, 0.11], and the difference in mean MOR was 0.11 [-0.01, 0.23].