

RESEARCH ARTICLE

Open Access



Are low-value care measures up to the task? A systematic review of the literature

Eline F. de Vries^{1*}, Jeroen N. Struijs², Richard Heijink², Roy J. P. Hendriks¹ and Caroline A. Baan^{1,2}

Abstract

Background: Reducing low-value care is a core component of healthcare reforms in many Western countries. A comprehensive and sound set of low-value care measures is needed in order to monitor low-value care use in general and in provider-payer contracts. Our objective was to review the scientific literature on low-value care measurement, aiming to assess the scope and quality of current measures.

Methods: A systematic review was performed for the period 2010–2015. We assessed the scope of low-value care recommendations and measures by categorizing them according to the Classification of Health Care Functions. Additionally, we assessed the quality of the measures by 1) analysing their development process and the level of evidence underlying the measures, and 2) analysing the evidence regarding the validity of a selected subset of the measures.

Results: Our search yielded 292 potentially relevant articles. After screening, we selected 23 articles eligible for review. We obtained 115 low-value care measures, of which 87 were concentrated in the cure sector, 25 in prevention and 3 in long-term care. No measures were found in rehabilitative care and health promotion. We found 62 measures from articles that translated low-value care recommendations into measures, while 53 measures were previously developed by institutions as the National Quality Forum. Three measures were assigned the highest level of evidence, as they were underpinned by both guidelines and literature evidence. Our search yielded no information on coding/criterion validity and construct validity for the included measures. Despite this, most measures were already used in practice.

Conclusion: This systematic review provides insight into the current state of low-value care measures. It shows that more attention is needed for the evidential underpinning and quality of these measures. Clear information about the level of evidence and validity helps to identify measures that truly represent low-value care and are sufficiently qualified to fulfil their aims through quality monitoring and in innovative payer-provider contracts. This will contribute to creating and maintaining the support of providers, payers, policy makers and citizens, who are all aiming to improve value in health care.

Keywords: Low-value care, Measures, Quality improvement, Performance measures

Abbreviations: AHRQ, Agency of Healthcare Research and Quality; AQC, Alternative quality contract; CMS, Centers of Medicare and Medicaid Services; CPT, Current procedural terminology; CW, Choosing wisely; e.g., *exempli gratia*; i.e., *id est*; ICD, International Classification of Diseases; ICHA-HC, Classification of health care functions; NICE, National Institute of Clinical Excellence; NQF, National Quality Forum; OECD, Organization for Economic Co-operation and Development; U.S., United States; USPSTF, United States Preventive Services Task Force; WHO, World Health Organization

* Correspondence: eline.de.vries@rivm.nl

¹Department Tranzo (Scientific Center for Care and Welfare), Tilburg University, Tilburg School of Social and Behavioral Sciences, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Full list of author information is available at the end of the article



Background

The concept of low-value care, defined as services that provide no benefit to patients or can even cause harm [1], has received much attention in recent years in Western countries. Reducing the use of low-value care is expected to contribute to cost containment and more efficiency in health care [2–4]. It leads to a reduction in medical spending without harming health outcomes and it may stimulate a reallocation of resources to high-value services [3]. In this way, measuring low-value care for which the non-effectiveness is proven provides information on a specific type of inefficiency, i.e. spending with no benefit, which can be used besides other, more indirect, types of efficiency analysis such as traditional cost-effectiveness studies or analyses of practice variation.

Internationally, several initiatives have been launched to reduce low-value service utilization, among which the Choosing Wisely (CW) campaign in the US. Similar initiatives have originated in 12 other countries including the United Kingdom, Canada, Australia and the Netherlands [3, 5]. In the CW campaign, participating specialty societies produce lists of recommendations that are to be discussed in the doctor's office, as for example, 'don't order diagnostic tests at regular intervals (such as every day), but rather in response to specific clinical questions' [6]. Ideally, these lists of recommendations would meet the CW criteria: 1) each of the services is within the specialty's purview, 2) each of the services is frequently used or costly, 3) each recommendation is based on sufficient evidence, and 4) the process for developing the recommendation list is documented and is made available to the public if requested [7]. In general, the recommendations aim to increase awareness among both doctors and patients [4] and subsequently influence the decision whether or not to use a specific service.

Besides these rather generic *recommendations*, studies have tried to assess the prevalence and geographic or practice variation in low-value care utilization (e.g. [8–11]) using direct *measures* of low-value care. The aim of the direct measures differs from the aim of recommendations. Where recommendations aim to create awareness among physicians and patients, low-value care measures may be widely used, for example in payer-provider contracts [12, 13] and for monitoring low-value care initiatives [3, 14].

To meet these aims, low-value care measures need to be methodologically sound [15, 16]. Otherwise, using these measures might create misinterpretation, underuse of indicated services, patient selection or damage the patient-physician relationship [17]. To date, only one study [18] reviewed the state of low-value care measurement by performing a scan of the published and grey literature. They found 37 specified measures and 123 services that may be developed into measures, covering

mainly diagnostic or therapeutic areas. Furthermore, another study [19] identified a set of low-value services and demonstrated significant variance in its utilization between hospital referral regions in the US.

Still, major knowledge gaps exist in the literature on measuring low-value care. First, there is lack of knowledge regarding the validity of current low-value care measures [15, 16]. As Baker et al. [14] pointed out earlier, low-value care measures must at least be rigorously evidence-based. In addition, they must be able to detect variation between providers, regions or countries, reflect actual cases of the concept of interest, be supported by correlations to other measures indicating the same concept, and not be subject to substantive systemic bias (i.e. importance, coding or criterion validity, construct validity and risk adjustment) [20]. Therefore, specific standards for how to develop and assess low-value care measures should be developed [14, 17]. Second, it is unclear whether current low-value care measures cover the whole continuum of care. This is important, because it was argued that low-value care use is present in all sectors along the care continuum [14, 21]. However, the low-value service recommendations from the CW initiative cover mainly specialist care in the cure sector [7].

In this study, we aimed to start filling these gaps by performing a systematic review of the recent scientific literature on low value care measurement. Our objective was twofold. Firstly, to assess the scope of low-value care *recommendations* and *measures* in the literature by categorizing them according to health care function (such as curative care, long-term care and rehabilitation). Secondly, to assess the quality of the *measures* by 1) analysing their development process and the evidence that underlies the measures and 2) analysing the evidence regarding the validity of a selection of the included measures.

Methods

Study design and search strategy

A systematic review of the literature was performed, focusing on English-language articles published between January 2010 and January 2015. As recommended by Cochrane [22], we performed our search in multiple databases including EMBASE, Medline, SciSearch, BIOSIS Previews and GLOBAL Health. We developed a search strategy to identify articles matching a variation of the following search terms: 1) initiatives, design, measuring, indicators, instrument, identifying, index; 2) waste, overuse, overutilization, misuse, low-value; and 3) health care, cure, care, prevention. Additional file 1 gives a detailed description of the search strategy.

Article selection

Two researchers (EFdV & RJPH) independently reviewed the relevance of the articles by screening titles and

abstracts. As recommended by Cochrane [22], we included articles from peer-reviewed journals only. The full-text was retrieved when both researchers considered the paper relevant. Articles were eligible for review when they met the following predefined criteria: 1) the low-value service recommendation or measure in the paper matched the definition ‘services that provide no benefit to patients or may even cause harm [1]’; 2) the low-value service recommendation or measure was described using clinical details such as diagnosis, patient population and treatment. We removed duplicate articles and replies or commentaries and theoretical or discussion articles that did not present any low-value service recommendations or measures. Any disagreement between the reviewers was resolved by discussion and consensus.

Data extraction

We extracted general characteristics of the articles (i.e. name of first author, year of publication, country, aim of the paper, methods) and the measures (i.e. the name of the measure, the numerator, the denominator, exclusion criteria and direction). In addition, we retrieved the original source or reference of the measure.

Recommendations versus measures

The literature search yielded both recommendations and measures for low-value care. We considered a description of low-value care as ‘measure’ when at least a numerator and denominator were specified as such. We identified the scope of both recommendations and measures, while the quality assessment was performed for the measures only.

Categorizing low-value care recommendations and measures by function in health care

All recommendations and measures were categorized using the Classification of Health Care Functions (ICHA-HC) as defined by the Organization for Economic Co-operation and Development (OECD), the World Health Organization (WHO) and Eurostat [23]. The ICHA-HC provides a framework to classify services according to their purpose or function and is commonly used to compare medical services internationally. It covers the entire continuum of the health system, i.e. curative care, rehabilitative care, long-term care and preventive care. We subcategorized curative care into general (i.e. primary) care and specialized care. General care involves basic care such as routine examinations, basic maternity care, routine diagnosis and follow-up, prescriptions and vaccinations (unless they are covered under a preventive program) [23]. Specialized care involves more complex technology and is often a breakdown from the basic fields (e.g. neurosurgery or allergology) [23]. In addition, the measures were

categorized according to the non-functional categories ancillary services (i.e. laboratory, imaging, transport), and medical goods (i.e. pharmaceutical and therapeutic appliances).

Assessing the quality of low-value care measures

We assessed the quality of the measures by 1) analysing their development process and the level of evidence underlying the measures, and 2) analyse the validity of a selection of the measures.

Development process and level of evidence

We distinguished two groups: A) articles that translated low-value service recommendations into low-value care measures, and B) articles that used measures previously developed by institutions. For both groups we reviewed how the measures were developed.

For group A, we searched for evidence underlying the recommendations. We categorized each measure based on the evidence, distinguishing three levels of evidence: 1) a combination of evidence from the literature (trial or review), guidelines and from CW, United States Preventive Services Task Force (USPSTF) or National Institute of Clinical Excellence (NICE) recommendations, 2) evidence from the literature (trial or review) or guidelines, and 3) evidence not found. As criteria for developing CW recommendations do not prescribe the level of evidence required [7] we labelled measures with CW, USPSTF or NICE evidence only, as ‘unknown’. We valued the first level highest, and the third level lowest.

For group B, we distinguished the same levels of evidence. However, here we specifically searched for elements of a quality label indicating the soundness of the measure. A National Quality Forum (NQF) endorsement corresponds with the qualification of ‘minor or no evidence gaps’ [20]. Measures with such qualification have the strongest evidence base regarding importance, face validity, criterion validity, construct validity and risk adjustment [20]. Therefore, NQF endorsed measures were valued highest. The Agency of Healthcare Research and Quality (AHRQ) and the Centers of Medicare and Medicaid Services’ (CMS) Quality provide information on the level of evidence by specifying the literature underpinning the measure. Therefore, measures from these sources were valued second best.

For both groups, our assessment was limited to the evidence provided in the reviewed article and the first document retrieved by reference tracking.

Validity

We selected a subset of five unique measures in order to gain insight in the quality of the low value care measures. Ideally, we would extensively assess each measure

regarding their validity. However, for 115 measures this was beyond the scope of this review. Therefore, we chose five unique measures that appeared most frequently in the reviewed articles, assuming more information on validity to be available for these measures. For these five measures, we searched for evidence regarding the measures' validity by reviewing the original source and reference tracking. In addition, we performed a PubMed search using key words from the name of the measures (i.e. diagnosis and procedure) and "low-value" or "overuse", augmented with "validity". Specifically, we searched for studies that aimed to assess the validity of the selected low-value care measures. Hereby, we distinguished between the most commonly used types of validity (as seen in e.g. [20, 24, 25]): face validity, coding/criterion validity (i.e. reflect actual cases low-value care) and construct validity (i.e. supported by correlations to other measures indicating low-value care) [20]. Face validity refers to the empirical or clinical rationale of the measure, and therefore we used the information from Table 2 for this criterion.

Results

Article retrieval

Our literature search yielded 292 potentially relevant articles (Fig. 1). Based on titles and abstracts, 108 articles were selected for full-text retrieval and thorough screening. This screening process generated 23 articles that were eligible for review. Main reasons for exclusion were using a different definition of low-value care ($n = 138$), for example articles on garbage, patient safety or drug abuse, or not providing clinical details ($n = 49$). Figure 1 shows all reasons for exclusion.

Article characteristics

All articles were published after 2011 and the vast majority of the 23 included articles originated from the United States ($n = 22$) (Table 1). Seven articles explicitly focused on low-value care measures. One of these reviewed the literature on low-value care measurement [18], and six were empirical studies measuring low-value care utilization [2, 8, 10, 11, 19, 26]. Low-value care

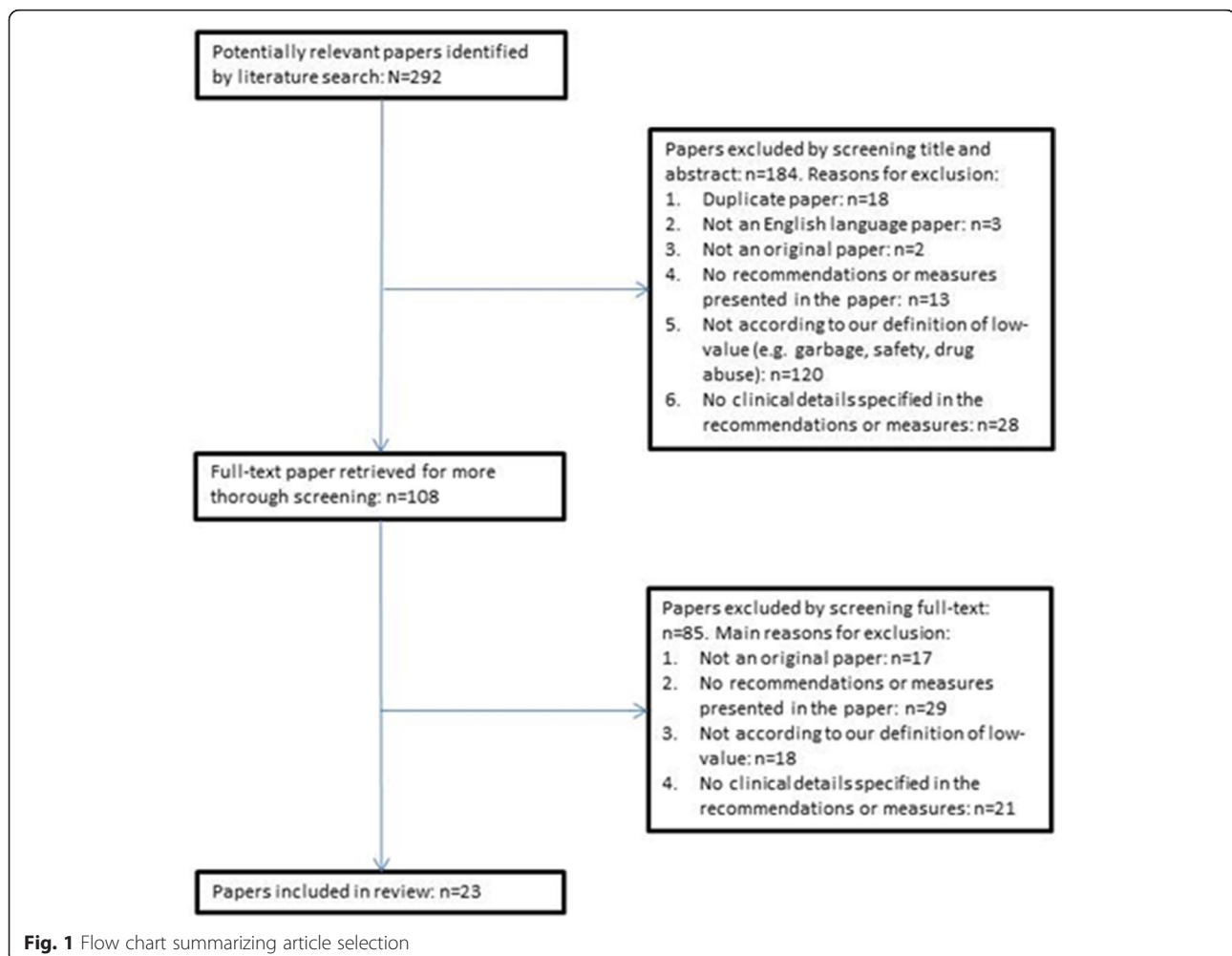


Table 1 General characteristics of the included articles ($n = 23$)

First author	Year of publication	Country	Aim	Method	Number retrieved		Recommendation initiative
					Measures ^a	Recommendations	
AGS Choosing Wisely AGSCW Workgroup [7]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi/consensus	0	5	CW
AGS Choosing Wisely Workgroup [34]	2014	US	To identify another five services that physicians and patients should question.	Delphi/consensus	0	5	CW
Amos [35]	2015	US	To determine the prevalence of PIMs for older adults in Elimia-Romagna, Italy, using updated Maio criteria.	Empirical analysis	0	16	Other
Bulger [36]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi/consensus	0	5	CW
Chan [18]	2013	US	To describe and critique the current state of overuse measurement.	Review	37	122	Other
Colla [8]	2015	US	To develop claims-based algorithms to estimate the prevalence of Choosing Wisely services and to examine the demographic, health and health care system correlates of low-value care at a regional level.	Empirical analysis	11	0	N.A.
Elshaug [37]	2012	AUS	To develop and apply a novel method for scanning a range of sources to identify existing health care services (excluding pharmaceuticals) that have questionable benefit, and produce a list that warrant further investigation.	Review	0	174	Other
Halpern [38]	2014	US	To present the Critical Care Societies Collaborative top 5 list in Critical Care Medicine and describe its development.	Review + Delphi/consensus	0	5	CW
Hicks [39]	2013	US	To identify five services that physicians and patients should question.	Review + Delphi/consensus	0	5	CW
Kale [26]	2013	US	The objective of this study was to determine whether the overuse and misuse of health care services in the ambulatory setting has decreased in the past decade.	Empirical analysis	13	0	N.A.
Keyhani [40]	2013	US	To compare rates of overuse in different health care systems and examine whether certain systems of care or insurers have lower rates of overuse of health care services.	Systematic review	0	7	Other
Korenstein [41]	2012	US	To perform an extensive search for studies of overuse of therapeutic procedures, diagnostic tests, and medications in the United States and describe the state of the literature.	Extensive search	0	33	Other
Mathias [10]	2012	US	To characterize performance on imaging-use measures, determine whether performance was consistent across measures, and identify hospital characteristics associated with highest-decile imaging use.	Empirical analysis	4	0	N.A.
Morden [11]	2014	US	To measure the prevalence and describe the geographic variation of short-interval (repeated in under 2 years)	Empirical analysis	4	0	N.A.

Table 1 General characteristics of the included articles (*n* = 23) (Continued)

			DXAs among Medicare beneficiaries and estimated the cost of this testing and its responsiveness to payment change.					
Onuoha [42]	2014	US	To develop a top 5 list of unnecessary medical services in anesthesiology.	Review + Delphi/consensus	0	5		CW
Quinonez [43]	2013	US	To produce top 5 lists.	Review + Delphi/consensus	0	5		CW
Rouster-Stevens [44]	2014	US	To create a pediatric rheumatology Top 5 list as part of the American Board of Internal Medicine Foundation's Choosing Wisely campaign.	Review + Delphi/consensus	0	5		CW
Schuur [45]	2014	US	To create a top-five list of tests, treatments, and disposition decisions that are of little value, are amenable to standardization, and are actionable by emergency medicine clinicians.	Delphi/consensus	0	5		CW
Schwartz [2]	2014	US	To develop claims-based measures of low-value services, examine service use (and associated spending) detected by these measures in Medicare, and determine whether patterns of use are related across different types of low-value services.	Empirical analysis	26	0		N.A.
Segal [19]	2014	US	To identify a set of possible indicators of overuse that can be operationalized with claims data and to describe variation in these indicators across the hospital referral regions (HRRs).	Empirical analysis	20	0		N.A.
Wiener [46]	2014	US	To create a top 5 list.	Review + Delphi/consensus	0	5		CW
Williams [47]	2012	US	To present the final five Choosing Wisely Don't do recommendations, the rationale for these specific recommendations, and two other recommendations.	Delphi/consensus	0	5		CW
Wood [48]	2013	US	To report on the CW top 5 list.	Review + Delphi/consensus	0	5		CW

AGS American Geriatrics Society, AUS Australia, CW Choosing Wisely, N.A. Not Applicable, PIM Potentially Inappropriate Medications, US United States

*at least a numerator and denominator was specified

recommendations were presented in 17 articles of which most were related to the CW campaign ($n = 12$).

Our search yielded 115 low-value care measures and 412 low-value care recommendations. Additional file 2 shows the characteristics of the 115 low-value care measures (i.e. containing a numerator and denominator). Out of these 115 measures, 42 contained exclusion criteria. For one of these measures (measure no. 72, Additional file 2), the direction of the measure was specified. Additional file 3 lists all recommendations (i.e. not containing numerator and/or denominator).

Low-value care recommendations and measures by function in health care

Figure 2 displays an overview of low-value care recommendations and measures categorized by health care function [23]. Here, we combined recommendations and measures covering the same combination of diagnosis and procedure. For instance, we found 8 measures for imaging in low back pain (measure no. 2–9, Additional file 2) using slightly varying exclusion criteria regarding e.g. age category (18–50 years versus 18–55 years) or intervention (imaging in general versus specific MRI). These eight measures were combined into a single group. In this manner, we found that 115 measures and 101 low-value care recommendations corresponded with 65 measure groups. The remaining recommendations ($n = 412 - 101 = 311$) were aggregated into 241 new recommendation groups.

In the cure dimension we found 87 measures, which we further subdivided in general care ($n = 85$) and specialized care ($n = 2$). Most measures in the cure dimension were in imaging ($n = 50$) or pharmaceutical goods ($n = 15$). The remaining measures were categorized in long-term care ($n = 3$) and secondary prevention ($n = 25$).

Quality of low-value care measures

Development process

Approximately half of the measures ($n = 62$) originated from low-value care recommendations (group A). Although the authors of the articles [2, 8, 11, 19, 26] described the methods to operationalize the low-value care measures, it was not specifically described how each specific low-value care recommendation was translated into a measure, i.e. how the denominator, numerator, exclusion and direction were determined for the purpose of the study. We did find that the measures developed by clinicians ($n = 18$) [8, 19] used (a combination of) International Classification of Diseases (ICD-9) and/or current procedural terminology (CPT) codes to construct the denominator [8, 19, 26].

The other half of the measures ($n = 54$) were developed by institutions (group B), including the NQF ($n = 25$) [8, 18, 19], the AHRQ ($n = 10$) [18, 19, 26], CMS QualityNet ($n = 16$) [10, 18, 19] and Blue Cross Blue Shield ($n = 2$) [19].

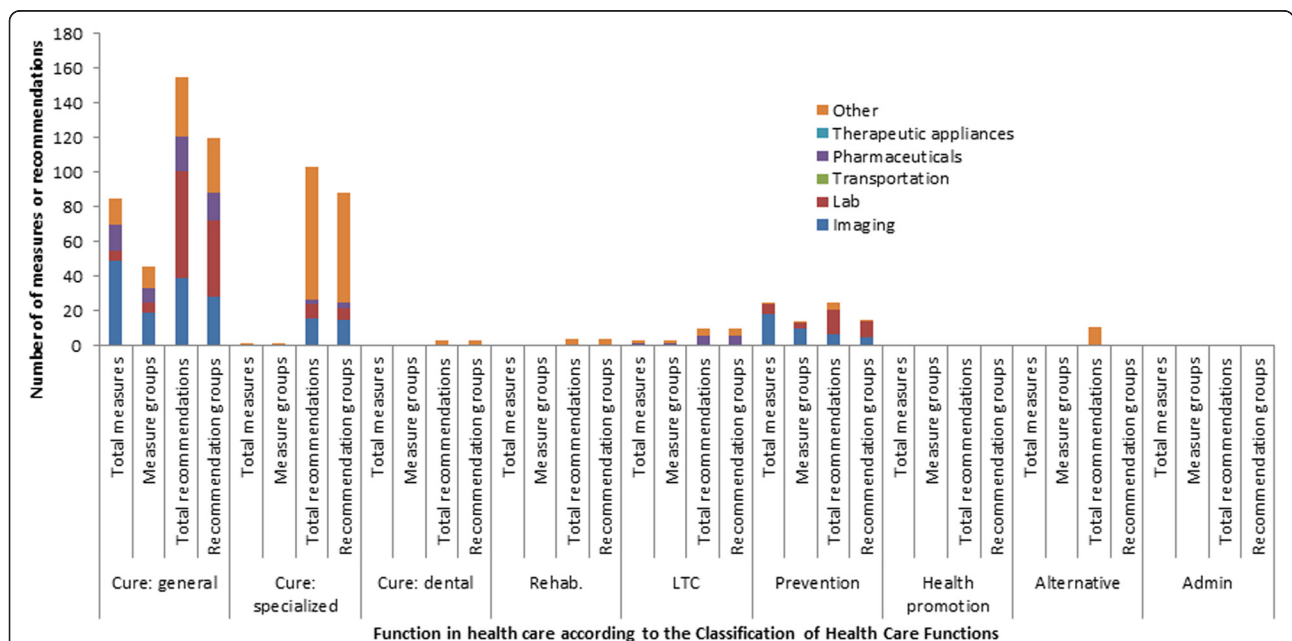


Fig. 2 Number of low-value care recommendations and measures categorized by the OECD/WHO/Eurostat Classification of Health Care Functions ($n = 426$)*. Admin.: Administrative; Alternative: Traditional, Complementary and Alternative Medicine; LTC: Long Term Care; Rehab.: Rehabilitative care; *We yielded 115 low-value care measures and 412 recommendations from the literature. Since 101 recommendations had a similar subjects as the measures, we subtracted these from 412 recommendations. That leaves 311 recommendations. Therefore, the total recommendations and measures in figure is $311 + 101 = 426$

Level of evidence

Table 2 shows the level of evidence provided in the referenced sources for each measure. In group A, the recommendations were mainly derived from CW, USPSTF, and NICE ($n = 45$). Other group A measures originated from guidelines, peer-reviewed literature or sources that summarized low-value services [27].

Three measures (measure no. 39, 40, 46; Additional file 2) were assigned the highest level of evidence (1), as they were underpinned by guidelines and literature (trial or review) and recommendations. For most measures ($n = 33$), however, we found guideline or literature evidence solely. For one measure (measure no. 101) the USPSTF considered the evidence for the underlying recommendation insufficient to assess the benefits and harms of the procedure, which we therefore assigned with the

lowest level of evidence. At the time of our review, for 24 measures, we considered the level of evidence to be unknown.

In group B, we found 19 measures [8, 18] supported by a quality label (NQF). For six measures the NQF endorsement was removed ($n = 4$) or not found ($n = 2$). Although the AHRQ website provides detailed information on the measures, we found no quality label, such as the NQF endorsement. We found seven measures (measure no. 1, 4, 55, 60, 69, 70, 86) displaying measurement characteristics (e.g. domain (process/outcome), description of denominator and numerator and target population) and evidence supporting the measure. The measures derived from QualityNet [28] were described in detail, however, no evidence supporting the description was provided.

Table 2 Level of evidence of low-value care measures

Level of evidence	Group A: Recommendation source	Measure numbers ^a	Count
1	CW, NICE or USPSTF recommendations; Guideline; Literature evidence (review or clinical trial)	39, 40, 46	3
2	CW, NICE or USPSTF recommendations; Literature evidence (review or clinical trial)	13, 14, 19, 20, 22, 23, 24, 25, 26, 44, 48, 50, 55, 77, 80, 90, 95, 103, 112, 115	20
2	CW, NICE or USPSTF recommendations; Guideline	33, 53	2
2	Literature evidence (reviews or clinical trial)	3, 21, 58, 76, 78, 81, 82, 83, 85, 89	10
2	Guideline	54, 57	2
3	USPSTF concludes that evidence is insufficient	101	1
Unknown	Literature: other compiled low-value service lists	47, 49, 113	3
Unknown	USPSTF recommendation not found	104, 107	2
Unknown	CW, NICE or USPSTF recommendations	34, 38, 43, 45, 51, 52, 59, 61, 84, 92, 98, 102, 105, 106, 108, 109, 110, 111, 114	19
Level of evidence	Group B: Institutional measure status	Measure numbers ^a	
1	NQF endorsed	5, 11, 16, 18, 41, 56, 62–67, 72, 73, 91, 93, 94, 96, 97	19
2	AHRQ measure supported by a clinical practice guideline or other peer-reviewed synthesis of clinical research evidence and one or more research studies published in a National Library of Medicine (NLM) indexed, peer-reviewed journal	60, 69, 70	3
2	AHRQ measure supported by a clinical practice guideline or other peer-reviewed synthesis of the clinical research evidence	1, 4, 55, 86	4
2	CMS QualityNet	2, 7, 8, 9, 27, 28, 29, 30, 31, 32, 37, 42, 99, 100	14
3	NQF endorsement removed since April 2014	6, 10, 74, 75	4
Unknown	NQF endorsement not found	17, 71	2
Unknown	AHRQ measure/guideline not found	68, 79, 87	3
Unknown	CMS QualityNet under revision	15	1
Unknown	CMS not found	12	1
Unknown	BCBS AQC measures not found	35, 36	2

AHRQ Agency for Healthcare Research and Quality, BCBS AQC Blue Cross Blue Shield, The Alternative Quality Contract, CMS Centers for Medicare & Medicaid Services, CW Choosing Wisely, IOM Institute of Medicine, NICE National Institute for Clinical Excellence (UK): do not do recommendations, NQF National Quality Forum, USPSTF United States Preventive Services Task Force

^a: measure numbers are in correspondence with Additional file 2

Validity

Table 3 shows the validity of the five measures that were found most frequently ($n = 26$). Two measures had the highest level of evidence. Our search yielded no information on coding/criterion validity and construct validity for the included measures, while four out of five measures are currently used in practice.

Discussion

To the best of our knowledge, this is the first systematic literature review identifying, categorizing and assessing the scope and quality of low-value care measures. We obtained 115 low-value care measures from the literature. Out of these 115 measures, 87 focused on the cure sector (primary and specialized care), 25 on secondary prevention and 3 on long-term care. Most measures ($n = 62$) originated from low-value care recommendations, while 53 were previously developed by institutions as the National Quality Forum. Three measures were assigned the highest level of evidence, as they were underpinned by both guidelines and literature evidence. For other measures, such a level of evidence was not transparently apparent. We do not conclude that these measures are invalid, because validity tests may not have been performed at all. Nevertheless, a lack of evidence is present at least. Our search yielded no information on coding/criterion validity and construct validity for the included subset of measures in this emerging field. Despite this, most measures are currently used in practice.

Low-value care measures have received increased attention and are now used for monitoring purposes, alignment of financial incentives [13, 29] and, in the foreseeable future, in shared saving programs such as the Alternative Quality Contract (AQC) [30]. In this manner, low-value care measurement may incentivize

providers and insurers to shift resources from low-value services to high-value services [31]. Our findings show that more attention is needed for the evidential underpinning and quality of these measures. Otherwise, the lack of transparency and evidence will reduce acceptance of low-value care measures by its users. Additionally, using measures of low quality, might lead to negative consequences including underuse of indicated services, cost-shifting, damages to the patient-physician relationship, provider dissatisfaction, adverse health effects, or patient selection [17].

Our review showed that more than half of the low-value care measures originated from low-value service recommendations (i.e. CW, NICE, USPSTF). This implies that the empirical evidence of many low-value care measures is based on the evidence supporting the underlying low-value service recommendations. However, criteria for the development of recommendation lists remains rather vague in the CW initiative, as well as in other similar campaigns [7]. Therefore, more transparency regarding the evidential underpinning of the recommendations is needed. Next to the importance of evidence underlying both low-value service recommendations and measures, one should be aware that the aim of low-value service recommendations differs from the aim of low-value care measures. The aim of CW recommendations is patient and physician awareness, while the aim of low-value care measures in turn may be to inform decisions on several levels. Consequently, requirements for the quality and development of recommendations and measures approaches vary accordingly.

We found that most current low-value care measures are concentrated in the cure sector even though it was argued that low-value services are provided and used along the entire continuum of care [21]. For example,

Table 3 Validity of the top five published low-value care measures

	Preoperative cardiac tests for non-cardiac low-risk surgery	Antibiotics for upper respiratory tract infections	Imaging for low-back pain	Cervical cancer screening	Imaging for sinusitis diagnosis
Number of measures included in review ^a	4 (measure no.: 42–44, 48)	7 (measure no.: 57–59, 62, 63, 65, 66)	8 (measure no.: 2–9)	3 (measure no.: 110–112)	4 (measure no.: 33, 35, 36, 59)
Measure criteria ^b					
Face validity ^c	Yes: level of evidence is 2	Yes, level of evidence is 1	Yes, level of evidence is 1	Yes, level of evidence is 2	Yes, level of evidence is 2
Coding/criterion validity	Not found	Not found	Not found	Not found	Not found
Construct validity	Not found	Not found	Not found	Not found	Not found
Used in practice	Yes, for payment determination (Hospital Outpatient Quality Reporting) [49]	Yes, in Physician Quality Reporting System [49]	Yes, for payment determination (Hospital Outpatient Quality Reporting) [49]	Yes, in Physician Quality Reporting System [49]	Not found

^a: Measure numbers corresponding with Additional file 2 between brackets

^b: Criteria for quality measures (AHRQ)

^c: For level of evidence also see Table 2

we only found four low-value care recommendations (that could possibly be transformed into low-value care measures) in rehabilitative care and none in the health promotion domain. This is probably the result of most measures originating from the CW initiative, which has its origin in the cure sector. While we acknowledge the emerging state of the field of research, we emphasize that similar consensus-based efforts are needed to stimulate the development of measures in other settings to broaden the scope and impact of the low-value care concept.

Given the potential impact of using low-value care measures, it is essential that guidelines for developing them be created by combined efforts of the involved parties: physicians, citizens, government and insurers [17, 32]. We do not suggest creating an evidence base for *each* health care intervention demonstrating all circumstances in which it is *not* effective. This will prove an un-doable exercise. Expert judgement by the clinician will always remain necessary to some degree. Therefore, other types of information, e.g. from studies on practice variation in procedure rates or cost-effectiveness studies, will remain necessary to identify inefficiencies in health-care, especially when high quality low-value care measures are not available. We do propose using expert opinion from initiatives such as Choosing Wisely as a starting point for monitoring low-value care. These qualitative information sources can be complemented with new scientific insights. For example, the insight that certain genes predict the development of breast cancer, must be used to prevent a considerable amount of low-value care utilization. Still, as soon as we start measuring and monitoring low-value care in such areas, it will be of particular interest to fully specify and define all measurement information, such as exclusion criteria, direction and evidence supporting the measure, and to make this publicly available. Furthermore, low-value care measures should be extensively tested regarding their level of evidence and validity before implementing them for use in practice, and specifically for the measures that are already in use. Recently, articles started studying aspects that are closely related to validity. As for example, Schwartz et al. [2] who found that the sensitivity and specificity strongly depends on the definition of the measures. Notwithstanding the efforts already been made, we stress the importance of the validity of the measures specifically being studied. Another area of research would be to further standardize low-value care measures, which ideally would result in alignment of the low-value care metrics and determining specifically for what subgroup or population a service is of low-value [2, 33]. Moreover, the guidelines should take into account any differences between countries in terms of the availability and provision of healthcare services that are likely to occur due to cultural or economic differences.

Another important issue to pay further attention to is the data requirements. Measuring low-value care utilization requires information on services provided to patients in combination with diagnosis and possibly additional patient characteristics. It is not clear to which extent current data sources can provide this information [2, 3], since rather detailed data need to be registered and data sources, such as claims data and detailed (hospital) registration data need to be connected in order to retrieve the necessary information.

Limitations

Our study has two main limitations. First, we did not evaluate the quality of each individual measure. Ideally, we would extensively assess each measure regarding their validity. To perform this task for 115 measures was, however, beyond the scope of this review. Nonetheless, we performed a first attempt in assessing the validity for the five measures that appeared most often in the literature and highlight several important general quality issues. Second, we did not include grey literature in our search. Therefore, we may have missed relevant measures. Nevertheless, for the purpose of our review, namely to systematically map the state of affairs of low-value care measurement, we are confident that the publications we did use provided sufficient evidence.

Conclusions

To conclude, our systematic review provides insight in the current state of low-value care measures. It shows that current low-value care measures only cover a selective part of the health care system. To achieve their full potential, future research should be focused on generating clear information about the level of evidence and validity to identify measures that truly represent low-value care in this emerging field of research. This will contribute to creating and maintaining the support of stakeholders who will use these measures for monitoring purposes and innovative insurer-provider contracts, all aiming to improve efficiency in health care with better health outcomes.

Additional files

Additional file 1: Search strategy. (DOCX 16 kb)

Additional file 2: Low-value care measures including numerator, denominator, exclusion criteria, direction and measure source and reference specified by function according to the OECD/WHO/Eurostat Classification of Health Care Functions ($n = 115$). (DOCX 182 kb)

Additional file 3: Low-value care recommendations. (DOCX 78 kb)

Acknowledgments

We are grateful to Margje H. Haverkamp, PhD (M.H.Haverkamp@lumc.nl) for helpful comments on an earlier version of this manuscript.

Funding

This study was funded under SPR project S/133002 of The National Institute of Public Health and the Environment in the Netherlands. The funder had no role in the design of the study, collection, analysis and interpretation of the data and writing of the manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Authors' contributions

Study concept and design: all authors; acquisition, analysis, or interpretation of data: all authors; drafting of the manuscript: drs. De Vries; critical revision of the manuscript for important intellectual content: all authors; study supervision: dr. Struijs, dr. Heijink and prof. dr. Baan. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹Department Tranzo (Scientific Center for Care and Welfare), Tilburg University, Tilburg School of Social and Behavioral Sciences, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. ²Department of Quality of Care and Health Economics, National Institute of Public Health and the Environment (RIVM), Center for Nutrition, Prevention and Health Services, P.O. Box 13720, BA, Bilthoven, The Netherlands.

Received: 6 February 2016 Accepted: 10 August 2016

Published online: 18 August 2016

References

- IOM (Institute of Medicine). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington: National Academy Press; 2001.
- Schwartz AL, Landon BE, Elshaug AG, Chernew ME, McWilliams JM. Measuring Low-Value Care in Medicare. *JAMA Intern Med*. 2014;174(7):1067–76.
- Bhatia RS, Levinson W, Shortt S, Pendrith C, Fric-Shamji E, Kallewaard M, Peul W, Veillard J, Elshaug A, Forde I, 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.
- Levinson W, Huynh T. Engaging physicians and patients in conversations about unnecessary tests and procedures: Choosing Wisely Canada. *CMAJ*. 2014;186(5):325–6.
- Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf*. 2015;24(2):167–74.
- Angus DC, Deutschman CS, Hall JB, Wilson KC, Munro CL, Hill NS. Choosing wisely in critical care: maximizing value in the intensive care unit. *Crit Care Med*. 2014;42(11):2437–8.
- Workgroup AGSCW. American Geriatrics Society identifies five things that healthcare providers and patients should question. *J Am Geriatr Soc*. 2013; 61(4):622–31.
- 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.
- Keyhani S, Falk R, Bishop T, Howell E, Korenstein D. The relationship between geographic variations and overuse of healthcare services: a systematic review. *Med Care*. 2012;50(3):257–61.
- Mathias JS, Feinglass J, Baker DW. Variations in US hospital performance on imaging-use measures. *Med Care*. 2012;50(9):808–14.
- Morden NE, Schpero WL, Zaha R, Sequist TD, Colla CH. Overuse of short-interval bone densitometry: assessing rates of low-value care. *Osteoporos Int*. 2014;25(9):2307–11.
- Morden NE, Colla CH, Sequist TD, Rosenthal MB. Choosing wisely—the politics and economics of labeling low-value services. *N Engl J Med*. 2014; 370(7):589–92.
- Shaw D, Melton P. Should GPs be paid to reduce unnecessary referrals? *BMJ* (Clinical research ed). 2015;351. doi:10.1136/bmj.h6148.
- Baker DW, Qaseem A, Reynolds PP, Gardner LA, Schneider EC. Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med*. 2013;158(1):55–9.
- Rosenberg A, Agiro A, Gottlieb M, Barron J, Brady P, Liu Y, Li C, DeVries A. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med* 2015;175(12):1–9. doi:10.1001/jamainternmed.2015.5441.
- Struijs JN, Drewes HW, Stein KV. Beyond integrated care: challenges on the way towards population health management. *Int J Integr Care*. 2015;15: e043.
- Mathias JS, Baker DW. Developing quality measures to address overuse. *JAMA*. 2013;309(18):1897–8.
- Chan KS, Chang E, Nassery N, Chang HY, Segal JB. The state of overuse measurement: a critical review. *Med Care Res Rev*. 2013;70(5):473–96.
- Segal JB, Bridges JF, Chang HY, Chang E, Nassery N, Weiner J, Chan KS. Identifying possible indicators of systematic overuse of health care procedures with claims data. *Med Care*. 2014;52(2):157–63.
- Guidance on Using the AHRQ QI for Hospital-Level Comparative Reporting. In: Agency for Healthcare Research and Quality, 2009.
- Fund TK. *Better value in the NHS: The role of changes in clinical practice*. London: The King's Fund; 2015.
- Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>.
- OECD/WHO/Eurostat. *A System of Health Accounts*. OECD Publishing; 2011. doi:10.1787/9789264116016-en.
- Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res*. 1976;11(4):478–507.
- Winters BD, Bharmal A, Wilson RF, Zhang A, Engineer L, Defoe D, Bass EB, Dy S, Pronovost PJ. Validity of the Agency for Health Care Research and Quality Patient Safety Indicators and the Centers for Medicare and Medicaid Hospital-acquired Conditions: A Systematic Review and Meta-Analysis. *Med Care*. 2016; doi:10.1097/MLR.0000000000000550.
- Kale MS, Bishop TF, Federman AD, Keyhani S. Trends in the overuse of ambulatory health care services in the United States. *JAMA Intern Med*. 2013;173(2):142–8.
- Qaseem A, Alguire P, Dallas P, Feinberg LE, Fitzgerald FT, Horwitch C, Humphrey L, LeBlond R, Moyer D, Wiese JG, et al. Appropriate use of screening and diagnostic tests to foster high-value, cost-conscious care. *Ann Intern Med*. 2012;156(2):147–9.
- Imaging Efficiency Measures [<https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1228695266120>]
- McCarthy M. US Choosing Wisely campaign has had only modest success, study finds. *BMJ* (Clinical research ed). 2015;351:h5437.
- Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-payer innovation in Massachusetts: the 'alternative quality contract'. *Health Affairs (Project Hope)*. 2011;30(1):51–61.
- Fendrick AM, Smith DG, Chernew ME. Applying value-based insurance design to low-value health services. *Health Aff*. 2010;29(11):2017–21.
- Willson A. The problem with eliminating 'low-value care'. *BMJ Qual Saf*. 2015;24(10):611–4.
- Elshaug AG, McWilliams J, Landon BE. The value of low-value lists. *JAMA*. 2013;309(8):775–6.
- Workgroup AGSCW. American Geriatrics Society identifies another five things that healthcare providers and patients should question, 5. 2014. p. 950–60.
- Amos TB, Keith SW, Del Canale S, Orsi P, Maggio M, Baccarini S, Gonzi G, Liu M, Maio V. Inappropriate prescribing in a large community-dwelling older population: a focus on prevalence and how it relates to patient and physician characteristics. *J Clin Pharm Ther*. 2015;40(1):7–13.
- Bulger J, Nickel W, Messler J, Goldstein J, O'Callaghan J, Auron M, Gulati M. Choosing wisely in adult hospital medicine: five opportunities for improved healthcare value. *J Hosp Med*. 2013;8(9):486–92.
- Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. *Med J Aust*. 2012;197(10):556–60.
- Halpern SD, Becker D, Curtis JR, Fowler R, Hyzy R, Kaplan LJ, Rawat N, Sessler CN, Wunsch H, Kahn JM, et al. An official American Thoracic Society/

- American Association of Critical-Care Nurses/American College of Chest Physicians/Society of Critical Care Medicine policy statement: the Choosing Wisely Top 5 list in Critical Care Medicine. *Am J Respir Crit Care Med.* 2014; 190(7):818–26.
- 39 Hicks LK, Bering H, Carson KR, Kleinerman J, Kukreti V, Ma A, Mueller BU, O'Brien SH, Pasquini M, Sarode R, et al. The ASH choosing wisely® campaign: Five hematologic tests and treatments to question. *Blood.* 2013;122(24): 3879–83.
- 40 Keyhani S, Falk R, Howell EA, Bishop T, Korenstein D. Overuse and systems of care: a systematic review. *Med Care.* 2013;51(6):503–8.
- 41 Korenstein D, Falk R, Howell AE, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. *Arch Intern Med.* 2012;172(2):171–8.
- 42 Onuoha OC, Arkoosh VA, Fleisher LA. Choosing wisely in anesthesiology: The gap between evidence and practice. *JAMA Internal Medicine.* 2014; 174(8):1391–5.
- 43 Quinonez RA, Garber MD, Schroeder AR, Alverson BK, Nickel W, Goldstein J, Bennett JS, Fine BR, Hartzog TH, McLean HS, et al. Choosing wisely in pediatric hospital medicine: five opportunities for improved healthcare value. *J Hosp Med.* 2013;8(9):479–85.
- 44 Rouser-Stevens KA, Ardoin SP, Cooper AM, Becker ML, Dragone LL, Huttenlocher A, Jones KB, Kolba KS, Moorthy LN, Nigrovic PA, et al. Choosing wisely: The american college of rheumatology's top 5 for pediatric rheumatology. *Arthritis Care Res.* 2014;66(5):649–57.
- 45 Schuur JD, Carney DP, Lyn ET, Raja AS, Michael JA, Ross NG, Venkatesh AK. A top-five list for emergency medicine: a pilot project to improve the value of emergency care. *JAMA Intern Med.* 2014;174(4):509–15.
- 46 Wiener RS, Ouellette DR, Diamond E, Fan VS, Maurer JR, Mularski RA, Peters JI, Halpern SD, American Thoracic S, American College of Chest P. An official American Thoracic Society/American College of Chest Physicians policy statement: the Choosing Wisely top five list in adult pulmonary medicine. *Chest.* 2014;145(6):1383–91.
- 47 Williams AW, Dwyer AC, Eddy AA, Fink JC, Jaber BL, Linas SL, Michael B, O'Hare AM, Schaefer HM, Shaffer RN, et al. Critical and honest conversations: the evidence behind the "Choosing Wisely" campaign recommendations by the American Society of Nephrology. *Clin J Am Soc Nephrol.* 2012;7(10): 1664–72.
- 48 Wood DE, Mitchell JD, Schmitz DS, Grondin SC, Ikonomidis JS, Bakaeen FG, Merritt RE, Meyer DM, Moffatt-Bruce SD, Reece TB, et al. Choosing wisely: cardiothoracic surgeons partnering with patients to make good health care decisions. *Ann Thorac Surg.* 2013;95(3):1130–5.
- 49 QPS Tool [<http://www.qualityforum.org/QPS/>]

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

