BMJ Open

Choosing Wisely: validity assessment of current US top five list recommendations using a pragmatic approach

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012366
Article Type:	Research
Date Submitted by the Author:	21-Apr-2016
Complete List of Authors:	Horvath, Karl; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (2) Department of Internal Medicine, Division of Endocrinology and Diabetology Semlitsch, Thomas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Jeitler, Klaus; Medical University Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (3) Institute for Medical Informatics, Statistics and Documentation Abuzahra, Muna; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Posch, Nicole; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Domke, Andreas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Siebenhofer, Andrea; Institute of General Practice and Evidence-based Health Services Research, Medical University Graz; Institute for General Practice, Goethe University
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Public health, Health services research, Patient-centred medicine
Keywords:	Choosing Wisely, top five lists, guidelines, methodological quality

Choosing Wisely: validity assessment of current US top five list recommendations using a pragmatic approach

Authors: Karl Horvath (1,2) karl.horvath@medunigraz.at, Thomas Semlitsch (1) Thomas.semlitsch@medunigraz.at, Klaus Jeitler (1,3) klaus.jeitler@medunigraz.at, Muna E. Abuzahra (1) muna.abuzahara@medunigraz.at, Nicole Posch (1) nicole.posch@medunigraz.at, Andreas Domke (1) andreas.domke@medunigraz.at, Andrea Siebenhofer (1,4) andrea.siebenhofer@medunigraz.at

(1) Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz, Austria
(2) Department of Internal Medicine, Division of Endocrinology and Diabetology, Medical University of Graz, Austria
(3) Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
(4) Institute of General Practice, Goethe University Frankfurt am Main, Germany

Karl Horvath, MD

Auenbruggerplatz 2/9

8036 Graz, Austria

Tel: +43-316-385-73559

Email: karl.horvath@medunigraz.at

Key words: Choosing Wisely, top five lists, methodological quality, guidelines

Word count: 2593

ABSTRACT

Objectives: Validity assessment of current top five list recommendations from the US Choosing Wisely Initiative.

Setting: Not applicable

Participants: All top five list recommendations available from the American Board of Internal Medicine Foundation website.

Main outcome measures/interventions: Compilation of US top five lists and search for current German high methodological quality (S3) guidelines. Extraction of guideline recommendations, including grade of recommendation (GoR), for suggestions comparable to top five list recommendations. For recommendations without guideline equivalents, the methodological quality was assessed using criteria similar to that used to judge guidelines, and relevant meta-literature was identified in cited references. Classification of top five list recommendations was based either on the GoR of guideline equivalents or on the methodological quality and citation of relevant meta-literature.

Results: 412 top five recommendations were identified. For 75 (18%), equivalents were found in current German S3 guidelines. 44 of these recommendations were associated with an "A" GoR, or a strong recommendation based on good evidence. A further 16 recommendations had a "B" GoR and 10 a "C". No GoR was provided for 5 recommendations. 337 top five list recommendations had no equivalent in the German S3 guidelines. The methodological quality of the development process was high and relevant literature was included in the citations for 87 top five list recommendations. For a further 36, either the methodological quality was high without any meta-literature citations, or meta-literature citations existed but the methodological quality was lacking. For the remaining 214 recommendations, either the methodological quality was lacking and no literature was cited, or the methodological quality was generally unsatisfactory.

Conclusions: 131 of current US top 5 list recommendations were judged to be sufficiently valid. For a substantial number of current US top five list recommendations their validity remains unclear. Methodological requirements for developing top five lists are recommended.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This is a systematic assessment of the validity of all current top 5 recommendations from the US Choosing Wisely Initiative.
- By matching top 5 list recommendations with recommendations from high quality German S3 guidelines or by assessing their methodological quality allying indicators otherwise user for the quality assessment of guidelines together with quoted supporting meta-literature allowed for a save identification of sufficiently valid top 5 list recommendations.
- Only recommendations from the US campaign were considered.
- Using only high quality guidelines might have resulted in an underestimation of the validity of recommendations for which good evidence but no S3 guidelines exist. Also, employing only German guidelines might have led us to underrate recommendations for which there are no equivalents in Germany, although high quality international guidelines exist.
- Underestimation of the validity of some of the recommendations might have occurred because recommendations were actually based on the best current evidence, but either no meta-literature was available or it was not quoted or no meta-literature but sufficient evidence from primary studies was available. Another source of possible misjudgement is that the recommendation was actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient.

INTRODUCTION

The Choosing Wisely initiative (CWI), a campaign led by the American Board of Internal Medicine (ABIM) Foundation, promotes doctor-patient communication and reducing waste in health care.[1] Within the initiative different medical societies develop and publish so called top five lists, naming (at least) five tests, interventions or services which are commonly overused in their respective specialties and should be questioned by doctors and patients. In light of the fact that for years rigorous guidelines have been published and yet they were not widely adopted or implemented in practice, a deliberately pragmatic approach was chosen to engage as many physicians and patients as possible. Because of this, only some loose methodological requirements for the development of top 5 lists were formulated, but among them was the prerequisite that all recommendations had to be evidence based. [1 2]

However, the campaign is currently experiencing some setbacks.[3] There is criticism and questions about the quality and reliability of the top five list recommendations because of the lack of comprehensive methodological requirements for the development of top five lists.[4] It was also noted that some lists might be influenced by financial self-interests.[5] To date only a few and limited attempts have been made to determine how evidence-based the available CWI recommendations are.[6-8] Uncertainty about the reliability of the top five lists can impede the implementation of top five lists in daily practice.[9 10] Also, recommendations lacking a basis in evidence might not only not reduce waste but lead to possible harm. Reliable recommendations are necessary to minimize the chance for error in decisions made by patients, doctors and policymakers. Differentiating between reliable and questionable recommendations is also key since top five lists will have increasing influence, as the Choosing Wisely campaign is being adopted in more countries.[11-13]

Our main aim was to assess the validity of current top five list recommendations from the US Choosing Wisely Initiative and categorize them accordingly.

METHODS

We carried out a search for top five lists on the ABIM website on April 24th, 2015. All identified top five lists were included. From the available lists we extracted all stated recommendations, information on which medical society was responsible for developing the top five list, the methods used for their development, the rationale, and the cited supporting literature. We then examined the recommendations with regard to possible congruence of content. Congruent recommendations were combined and considered as one single recommendation.

Next we conducted a search for all current (as of the year 2015) German S3 guidelines in the web portal of the Association of the Scientific Medical Associations in Germany (AWMF). All German S3 guidelines can be found in this web portal. No restrictions concerning medical specialities were made.

The Association of the Scientific Medical Associations in Germany classifies guidelines into three categories: S1 expert recommendations developed by informal consensus, S2 requiring a formal consensus finding and/or a search for evidence and S3 denoting guidelines of the highest methodological quality. S3 guidelines must contain all elements of the AGREE II instrument, including a multidisciplinary development group, a systematic search for and a systematic appraisal of relevant literature, and a structured process for finding consensus. Also, for every recommendation a justified grade of recommendation (GoR) and the level of evidence (LoE) must be stated.[14,15] By using only S3 guidelines we aimed for the highest validity of guideline recommendations. We matched the top five list recommendations with the identified guidelines based on the guidelines' title and the issuing medical societies. Relevant guideline recommendations and their associated grade of recommendation were extracted. We only considered guideline recommendations as equivalent to top 5 list recommendations if they referred to omitting tests or interventions. We did not consider recommendations for certain services associated with a low GoR and/or insufficient evidence as a top 5 list recommendations equivalent. Matching and extraction was done by two authors independently and any differences were resolved by discussion. Because different guidelines used different terms for their grades of recommendations, a standardised GoR scheme was developed (table 1) and assigned to the respective recommendations.

Table 1: Standardised Grade of Recommendation

Phrasing of Recommendation in Guideline	GoR label used in Guideline	Evidence	Standardised GoR
Strong recommendation („shall")	A, ↑↑, ↓↓	Strong for or against	**A**
Recommendation („should")	B, ↑, ↓	Moderate for or against	**B**
Recommendation based on expert consensus	CCP, EC, GCP	Not possible or sought	**C**
Open („can")	C, 0, ↔	Weak or unclear	**D**

EC: expert consensus; GCP: good clinical practice; GoR: grade of recommendation; CCP: clinical consensus point

In the case of top five list recommendations for which no guideline equivalent could be identified, we assessed the methodological quality using criteria otherwise applied for the evaluation of guideline quality: systematic literature searches, involvement of a multidisciplinary group of experts, patient participation, management of conflicts of interests, method of consensus finding and planned updates.[4] [16] We only considered information reported in the "How the list was developed" sections of the top five lists. Based on these criteria we judged the methodological quality as high (requirements fully or largely met), moderate (requirements partially met) or low (requirements not or mostly not met). Additionally, we searched the references quoted in the top 5 lists for supporting systematic meta-literature (meta-analyses, systematic reviews, health technology reports and evidence based guidelines utilizing systematic searches), because we hypothesised that the availability of such relevant meta-literature would increase the chance of a full consideration of the available evidence with appraisals of

the effect sizes, the chance for bias and the consistency of results by the top five list authors. We evaluated the relevance of the identified meta-literature based on their full text publications. For top five recommendations with insufficient methodological quality, we omitted the meta-literature assessment. Quality assessment was done by two authors independently and discrepancies were resolved by discussion.

Finally, based on the standardised GoR of guideline equivalents or on their methodological quality and the availability of supporting systematic meta-literature, we classified all top five recommendations into eight groups as shown in table 2.

Table 2: Categories of top five list recommendations

Categories	Criteria
1. CWI recommendations with corresponding recommendations from S3 guidelines	
1A	standardised GoR A
1B	standardised GoR B
1C	standardised GoR C
1D	standardised GoR D
1E	no GoR available
2. CWI recommendations without corresponding recommendations from S3 guidelines	
2A	high methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2B	high methodological quality but no supporting systematic meta-literature (SG, SR, MA, HTA) cited or moderate methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2C	moderate methodological quality and no supporting systematic meta-literature (SG, SR, MA, HTA) cited or low methodological quality

CWI: Choosing Wisely initiative; GoR: grade of recommendation; HTA: health technology assessment; MA: meta-analysis; SR: systematic review; SG: systematic guideline

Since patients were not involved in this investigation and no data linked to persons were used, this project was not reviewed by the ethics committee.

Patient involvement

Patients were not involved in formulating the research question, the design or conduct of this study.

RESULTS

From the ABIM website, searched on April 24th 2015,[17] we identified 412 top five list recommendations developed by 66 different medical societies. Of these, 96 (23%) were of congruent content.

Top five list recommendations with S3 guideline equivalents

The search in the web portal of the Association of the Scientific Medical Associations of Germany (search date June 2nd 2015) yielded 139 methodologically high quality German S3 guidelines.[18] We excluded 23 guidelines because they were outdated (expiration dates before January 1st 2015).

For 75 (18%) top five list recommendations we identified guideline equivalents. For 9 recommendations we found equivalents in more than one (up to five) guideline. In these instances, we based our assessments on the guideline with the closest fit of content. 44 (11%) top five list recommendations were equivalent to a standardised "A" GoR, or a strong recommendation based on good evidence. For 16 (4%) and 10 (2%) recommendations, the corresponding standardised GoR was "B" or "C" respectively. There were no recommendations classified as "D" GoR but 5 (1%) could not be classified because no GoR was available for their guideline equivalents (for all see figure 1). We did not find any guideline recommendation contradicting its associated CWI recommendation.

Top five list recommendations without S3 guideline equivalents

The majority of the top 5 list recommendations, 337 or 82%, had no equivalent in current German S3 guidelines. For 103 (25%) recommendations we judged the methodological quality as high. Relevant systematic meta-literature was included in the references lists of 87 (21%) of these high quality recommendations. For further 36 (9%) recommendations, either the methodological quality was high without citation of relevant meta-literature, or literature citations existed but the methodological quality was only moderate. For the remaining 214 (52%) top five list recommendations, either the methodological quality was judged as moderate and no relevant meta-literature was cited, or the methodological was generally unsatisfactory (for all see figure 1).

Concerning the quality criteria (table 3) a systematic search was reported for 91 (22%) recommendations. We found indications for patient participation in the development process for 17 (4%) and for the involvement of a multidisciplinary group of experts for 208 (50%) recommendations. An expiration date or information on planned updates was not given for any of the recommendations. Also, information concerning the management of potential conflicts of interests of top five list authors was not available for 16 (4%) recommendations. All remaining recommendations contained references only to the respective general policies as stated on the websites of the different medical societies. While for 328 (80%) recommendations some information on the process for formulating the recommendations was available, a structured, validated process was described only for 98 (24%) recommendations.

Table 3: Top five list recommendations without S3 guideline equivalents, methodological quality

	Systematic search (n)	Multidisciplinary expert team (n)	Patient participation (n)	Structured consensus finding (n)	Management of CoI (n)	Expiration date (n)
yes	91	208	17	98	0	0
no	184	129	320	239	16	337
unclear	62	0	0	0	321	0

CoI: conflict of interest

Validity of top five recommendations

Of all 412 available top five list recommendations, we judged 131 (32%) to be sufficiently valid, 44 (11%) because their S3 guideline equivalents were associated with an "A" GoR indicating a strong recommendation with good supporting evidence, and 87 (21%) because their methodological quality was high and relevant systematic meta-literature was cited in their support (figure 1 and supplementary material table A).

The validity of 281 of the top five list recommendations remains unclear.

DISCUSSION

Principal findings

Our study provides evidence that only about a third of current US top five list recommendations up to April 2015 provide sufficient valid information on tests, interventions or services which are commonly overused. Methodological quality varied considerably, especially with regard to conducting systematic searches for evidence, the methods for achieving a structured consensus, and the involvement of experts from multiple disciplines. Patient participation in the development of lists, and information on the management of potential conflicts of interest were scarce.

While it is likely that the results reflect mainly the lack of adequate methodological requirements on how to develop top 5 lists,[4] other possible causes such as discrepancy of actual methods and their reporting, or financial self-interest[5], cannot be ruled out completely.

Strengths and limitations

All current top five list recommendations were included in our investigation. We systematically assessed the validity and methodological quality of the recommendations. Searching guidelines for equivalents identified recommendations with sufficient importance for daily practice. German S3

guidelines are required to incorporate all aspects of the AGREE II instrument and the given GoR in those guidelines always also reflects the quality and level of the underlying evidence. Thus we were able to judge top five list recommendations for which we identified guideline equivalents associated with the highest GoR (standardised GoR "A") as sufficiently valid with a high level of certainty. A guideline GoR below "A" is an indication of uncertain or insufficient evidence and we thus judged the validity of top 5 list recommendations with equivalents which were associated with a GoR below "A" as unclear. Using only high quality guidelines might also have resulted in an underestimation of the validity of recommendations for which good evidence but no S3 guidelines exist. Also, employing only German guidelines might have led us to underrate recommendations for which there are no equivalents in Germany, although high quality international guidelines exist.

Top 5 list recommendations without S3 guideline equivalents were only judged as sufficiently valid if a methodological quality was found. This was determined by applying indicators such as a transparent and structured development process including multidisciplinary experts and patients, and the quotation of supporting meta-literature. However, since we did not check whether additional meta-literature potentially contradicting the quoted references was available, the validity might have been overestimated in some cases. On the other hand, using this approach, it seems likely that we underestimated the validity of some of the recommendations for which the validity remained unclear because they were either of a lesser methodological quality or no meta-literature was quoted. This might be the case when recommendations which were actually based on the best current evidence, but either no meta-literature was available or it was not quoted. Also the validity of recommendations for which no meta-literature but sufficient evidence from primary studies was available might have been underestimated. Another source of possible misjudgement is that the recommendation was actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient. Also we considered only top five list recommendations from the US while many more countries have now started to produce their own[13].

To assess the validity of CWI recommendations without guideline equivalents with a high level of certainty, it would be necessary to conduct systematic reviews, based on primary or secondary literature, for each of these recommendations. This is the only method to assure that all available evidence will be considered, and the effect sizes and the likelihood of bias are sufficiently assessed.[19] But conducting such systematic reviews is highly time consuming. We thus used a pragmatic approach, based on the hypothesis that developing recommendations according to stringent methodological criteria[16] which are used in developing high-quality guidelines would suffice to assume a low likelihood of error.

Comparison with other studies

To our knowledge, this is the first study to comprehensively assess the methodological quality and reliability of all currently available top five list recommendations. In a somewhat similar attempt Hipkins et al investigated the top five lists in regard to a thorough literature search and an evidence based process used in the development of the lists.[6] They considered the information given by the authors in the "How the list was developed" sections and any additional information from searches in MEDLINE, Google Scholar, relevant websites and publications. They found a description of some review of literature in more than a brief, non-specific way for only 20% to 35% of the lists they examined, and an evidence based process for about 38% of the lists. These results are in good accordance with our own findings. Gliwa and Pearson in their 2014 study did not assess the quality of the development process or reliability, but categorized the reported evidence according to the evidentiary rationales given by the top five list authors.[8] Institute for Clinical and Economical Review (ICER) reports[7] are only available for a small number of lists and the evaluation of the supporting evidence is based on the work by Gliwa and Pearson.

Potential implications for clinicians or policymakers

The lack of stringent standards for developing top five lists should not so much be viewed as a flaw, but rather as a necessary pragmatic approach for the campaign to gain momentum. But from the results of our study, it is clear that methodological requirements for the development of top five lists need to be formulated. An explicit, comprehensive consideration of the current best evidence and a transparent development should be mandatory. Attention should also be given to an adequate management of possible conflicts of interests and to patient participation. While an evidence based development process is imperative, additional criteria such as the extent of potential harm, disease severity and urgency, health resources consumption and others have to be considered when prioritizing recommendations to allow for a substantial impact on the health system. Better reporting is necessary. To keep top five lists concise, a comprehensive description might be given on the medical societies' websites with a link provided in the published lists.

New ways of developing top five lists, for example using big data or utilizing high quality guidelines[20] [21], need to be explored. In the context of overuse, study results showing no differences between interventions are helpful findings in providing a solid evidence base for respective recommendations. Thus it is important that such negative studies are published.

Unanswered questions and future research

The proposed method for assessing the reliability of top five list recommendations still needs to be validated, which we have planned as a follow-up project. The assessment also needs to be expanded to include international top 5 list recommendations and guidelines.

Contributors: KH, TS, KJ and AS designed the study. KH, TS, KJ, AS, MEA, NP and AD were involved in the conduct of the study, data analysis and interpretation. KH drafted the manuscript and TS, KJ, AS, MEA, NP and AD critically revised it for important intellectual content. KH is the guarantor.

Funding: This study was supported by the Techniker Krankenkasse, a German health insurance provider. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) KH, TS, MA, NP, AD, KJ, AS have support from the Techniker Krankenkasse, a German health insurance provider, for the submitted work; (2) KH, TS, MA, NP, AD, KJ, AS have no relationships with companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (4) KH, TS, MA, NP, AD, KJ, AS have no non-financial interests that may be relevant to the submitted work.

Ethical approval: No ethical approval was sought.

Data sharing: No additional data available.

REFERENCES

1. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely. About the Campaign. 2015 [updated 25.11.2015. Available from: http://www.choosingwisely.org/wp-content/uploads/2015/04/About-Choosing-Wisely.pdf.
2. Wolfson D, Santa J, Slass L. Engaging Physicians and Consumers in Conversations About Treatment Overuse and Waste: A Short History of the Choosing Wisely Campaign. Acad Med 2014;**89**(7):990-95.
3. Lenzer J. Choosing Wisely: setbacks and progress. BMJ 2015;**351**:h6760.
4. Strech D, Follmann M, Klemperer D, et al. When Choosing Wisely meets clinical practice guidelines [Wenn "Choosing Wisely" auf Leitlinien trifft]. Z Evid Fortbild Qual Gesundh wesen (ZEFQ) 2014;**108**(10):601-03.
5. Morden NE, Colla CH, Sequist TD, et al. Choosing wisely--the politics and economics of labeling low-value services. N Engl J Med 2014;**370**(7):589-92.
6. Hipkins B, Cuervo C, Barclay C, et al. Choosing Wisely: Progress in use of evidence to develop top 5 lists? [Poster]. Preventing Overdiagnosis, Winding back the harms of too much medicine; September 1-3, 2015; Bethesda, Maryland USA 2015.
7. American Board of Internal Medicine (ABIM) Foundation. ICER Baseline Reports [webpage] 2016 [Available from: http://www.choosingwisely.org/resources/icer-baseline-reports/.
8. Gliwa C, Pearson SD. Evidentiary rationales for the Choosing Wisely Top 5 lists. JAMA 2014;**311**(14):1443-4.

9. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;**282**(15):1458-65.
10. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004;**8**(6):iii-iv, 1-72.
11. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Gemeinsam Klug Entscheiden [webpage] 2016 [Available from: http://www.awmf.org/medizin-versorgung/gemeinsam-klug-entscheiden.html.
12. Malhotra A, Maughan D, Ansell J, et al. Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. BMJ 2015;**350**:h2308.
13. Levinson W, Kallewaard M, Bhatia RS, et al. 'Choosing Wisely': a growing international campaign. Bmj Qual Saf 2015;**24**(2):167-74.
14. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien". 1. Auflage 2012 [updated 2012-11-06. Available from: http://www.awmf.org/leitlinien/awmf-regelwerk.html.
15. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. Klassifikation S3-Leitlinien 2012 [Available from: http://www.awmf.org/leitlinien/awmf-regelwerk/ll-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/klassifikation-s3.html.
16. Semlitsch T, Jeitler K, Kopp IB, et al. [Development of a workable mini checklist to assess guideline quality]. Z Evid Fortbild Qual Gesundhwes 2014;**108**(5-6):299-312.
17. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely - Clinician Lists 2015 [Available from: http://www.choosingwisely.org/wp-content/uploads/2015/01/Choosing-Wisely-Recommendations.pdf.
18. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). AWMF online; Das Portal der wissenschaftlichen Medizin - Aktuelle Leitlinien 2015 [Available from: http://www.awmf.org/leitlinien/aktuelle-leitlinien.html.
19. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [update March 2011]: The Cochrane Collaboration; 2011 [Available from: www.cochrane-handbook.org.
20. Grad R, Pluye P, Tang D, et al. Patient-oriented evidence that matters (POEMs) suggest potential clinical topics for the Choosing Wisely campaign. J Am Board Fam Med 2015;**28**(2):184-9.
21. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Manual Entwicklung von Empfehlungen im Rahmen der Initiative Gemeinsam Klug Entscheiden (GKE) [webpage] 2016 [Available from: http://www.awmf.org/fileadmin/user_upload/Medizinische_Versorgung/GKE/Manual_GKE_AWMF_V1.0-Konsultationsfassung.pdf

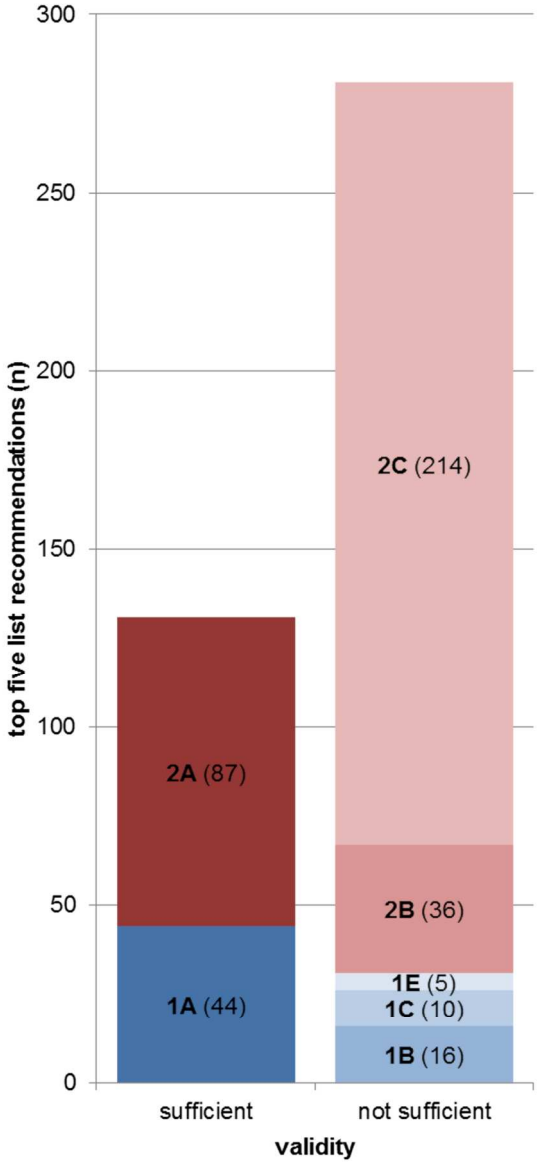


Figure 1: Validity of top five list recommendations. Blue columns represent top five list recommendations with guideline equivalents, red columns top five list recommendations without guideline equivalents. Numbers and letters in brackets denote different categories of top five recommendations (see table 2).

87x178mm (300 x 300 DPI)

Table A: Top five list recommendations with sufficient reliability

Recommendation	Publishing Medical Society
CWI recommendations with S3 guideline equivalents associated with an "A" GoR	
Don't prescribe bed rest for acute localized back pain without completing an evaluation.	American Academy of Physical Medicine and Rehabilitation
Don't order an imaging study for back pain without performing a thorough physical examination.	American Academy of Physical Medicine and Rehabilitation
Avoid lumbar spine imaging in the emergency department for adults with non-traumatic back pain unless the patient has severe or progressive neurologic deficits or is suspected of having a serious underlying condition (such as vertebral infection, cauda equine syndrome, or cancer with bony metastasis).	American College of Emergency Physicians
Don't do imaging for low back pain within the first six weeks, unless red flags are present.	American Academy of Family Physicians
Don't obtain imaging (plain radiographs, magnetic resonance imaging, computed tomography [CT], or other advanced imaging) of the spine in patients with non-specific acute low back pain and without red flags.	American Association of Neurological Surgeons and Congress of Neurological Surgeons
Don't obtain imaging studies in patients with non-specific low back pain.	American College of Physicians
Avoid imaging studies (MRI, CT or X-rays) for acute low back pain without specific indications.	American Society of Anesthesiologists – Pain Medicine
Don't recommend advanced imaging (e.g., MRI) of the spine within the first six weeks in patients with non-specific acute low back pain in the absence of red flags.	North American Spine Society
Avoid prescribing antibiotics in the emergency department for uncomplicated sinusitis.	American College of Emergency Physicians
Don't order sinus computed tomography (CT) or indiscriminately prescribe antibiotics for uncomplicated acute rhinosinusitis.	American Academy of Allergy, Asthma & Immunology
Don't routinely prescribe antibiotics for acute mild-to-moderate sinusitis unless symptoms last for seven or more days, or symptoms worsen after initial clinical improvement.	American Academy of Family Physicians
Antibiotics should not be used for apparent viral respiratory illnesses (sinusitis, pharyngitis, bronchitis).	American Academy of Pediatrics
Avoid prescribing antibiotics for upper respiratory infections.	Infectious Diseases Society of America
Don't perform sentinel lymph node biopsy or other diagnostic tests for the evaluation of early, thin melanoma because they do not improve survival.	American Academy of Dermatology
Don't screen for carotid artery stenosis (CAS) in asymptomatic adult patients.	American Academy of Family Physicians
Don't routinely screen for prostate cancer using a prostate-specific antigen (PSA) test or digital rectal exam.	American Academy of Family Physicians
Don't routinely perform PSA-based screening for prostate cancer.	American College of Preventive Medicine
Don't perform PSA testing for prostate cancer screening in men with no symptoms of the disease when they are expected to live less than 10 years.	American Society of Clinical Oncology
Don't use post-operative splinting of the wrist after carpal tunnel release for long-term relief.	American Academy of Orthopaedic Surgeons
Don't perform annual stress cardiac imaging or advanced non-invasive imaging as part of routine	American College of Cardiology

follow-up in asymptomatic patients.	
Avoid performing routine stress testing after percutaneous coronary intervention (PCI) without specific clinical indications.	Society for Cardiovascular Angiography and Interventions
Don't perform routine annual stress testing after coronary artery revascularization.	Society of Nuclear Medicine and Molecular Imaging
Don't perform stress cardiac imaging or advanced non-invasive imaging in the initial evaluation of patients without cardiac symptoms unless high-risk markers are present.	American College of Cardiology
Don't perform cardiac imaging for patients who are at low risk.	American Society of Nuclear Cardiology
Don't perform stress cardiac imaging or coronary angiography in patients without cardiac symptoms unless high-risk markers are present.	American Society of Nuclear Cardiology
Avoid using stress echocardiograms on asymptomatic patients who meet "low risk" scoring criteria for coronary disease.	American Society of Echocardiography
Don't perform coronary CMR in the initial evaluation of asymptomatic patients.	Society for Cardiovascular Magnetic Resonance
Don't perform stress cardiovascular magnetic resonance (CMR) in the initial evaluation of chest pain patients with low pretest probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Don't screen for ovarian cancer in asymptomatic women at average risk.	American College of Obstetricians and Gynecologists
Don't screen low risk women with CA-125 or ultrasound for ovarian cancer.	Society of Gynecologic Oncology
Don't take a multi-vitamin, vitamin E or beta carotene to prevent cardiovascular disease or cancer.	American College of Preventive Medicine
Don't prescribe biologics for rheumatoid arthritis before a trial of methotrexate (or other conventional non-biologic DMARDs).	American College of Rheumatology
For a patient with functional abdominal pain syndrome (as per ROME III criteria) computed tomography (CT) scans should not be repeated unless there is a major change in clinical findings or symptoms.	American Gastroenterological Association
Don't use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present.	American Geriatrics Society
Don't treat asymptomatic bacteriuria with antibiotics.	Infectious Diseases Society of America
Avoid using PET or PET-CT scanning as part of routine follow-up care to monitor for a cancer recurrence in asymptomatic patients who have finished initial treatment to eliminate the cancer unless there is high-level evidence that such imaging will change the outcome.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early breast cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early prostate cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't initiate management of low-risk prostate cancer without discussing active surveillance.	American Society for Radiation Oncology
Don't recommend bed rest for more than 48 hours when treating low back pain.	North American Spine Society
Avoid coronary angiography in post-coronary artery bypass graft (CABG) and post-PCI patients who are asymptomatic, or who have normal or mildly abnormal	Society for Cardiovascular Angiography and Interventions

stress tests and stable symptoms not limiting quality of life.	
Don't perform stress CMR in patients with acute chest pain and high probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Avoid routine imaging for cancer surveillance in women with gynecologic cancer, specifically ovarian, endometrial, cervical, vulvar and vaginal cancer.	Society of Gynecologic Oncology
Patients with suspected or biopsy proven Stage I NSCLC do not require brain imaging prior to definitive care in the absence of neurologic symptoms.	The Society of Thoracic Surgeons
CWI recommendations without S3-guideline equivalents associated with good methodological quality and relevant meta-literature	
Avoid CT pulmonary angiography in emergency department patients with a low-pretest probability of pulmonary embolism and either a negative Pulmonary Embolism Rule-Out Criteria (PERC) or a negative D-dimer.	American College of Emergency Physicians
Don't perform chest computed tomography (CT angiography) to evaluate for possible pulmonary embolism in patients with a low clinical probability and negative results of a highly sensitive D-dimer assay.	American College of Chest Physicians and American Thoracic Society
Don't place an indwelling urinary catheter to manage urinary incontinence.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't place or maintain a urinary catheter in a patient unless there is a specific indication to do so.	American Academy of Nursing
Avoid placing indwelling urinary catheters in the emergency department for either urine output monitoring in stable patients who can void, or for patient or staff convenience.	American College of Emergency Physicians
Don't place, or leave in place, urinary catheters for incontinence or convenience or monitoring of output for non-critically ill patients (acceptable indications: critical illness, obstruction, hospice, preoperatively for <2 days for urologic procedures; use weights instead to monitor diuresis).	Society of Hospital Medicine – Adult Hospital Medicine
Don't initiate antihypertensive treatment in individuals ≥60 years of age for systolic blood pressure (SBP) <150 mm Hg or diastolic blood pressure (DBP) <90 mm Hg.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't recommend screening for breast, colorectal or prostate cancer if life expectancy is estimated to be less than 10 years.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid colorectal cancer screening tests on asymptomatic patients with a life expectancy of less than 10 years and no family or personal history of colorectal neoplasia.	American College of Surgeons
Don't recommend screening for breast, colorectal, prostate or lung cancer without considering life expectancy and the risks of testing, overdiagnosis and overtreatment.	American Geriatrics Society
Don't perform routine cancer screening for dialysis patients with limited life expectancies without signs or symptoms.	American Society of Nephrology
Don't recommend cancer screening in adults with life expectancy of less than 10 years.	Society of General Internal Medicine
Don't obtain a C. difficile toxin test to confirm "cure" if symptoms have resolved.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't insert percutaneous feeding tubes in individuals with advanced dementia. Instead, offer oral assisted	AMDA – The Society for Post-Acute and Long-Term Care Medicine

feedings.	
Don't use sliding scale insulin (SSI) for long-term diabetes management for individuals residing in the nursing home.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't obtain a urine culture unless there are clear signs and symptoms that localize to the urinary tract.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid the use of surveillance cultures for the screening and treatment of asymptomatic bacteruria.	American Academy of Pediatrics
Don't order annual electrocardiograms (EKGs) or any other cardiac screening for low-risk patients without symptoms.	American Academy of Family Physicians
Don't prescribe antibiotics for otitis media in children aged 2-12 years with non-severe symptoms where the observation option is reasonable.	American Academy of Family Physicians
Don't screen women older than 65 years of age for cervical cancer who have had adequate prior screening and are not otherwise at high risk for cervical cancer.	American Academy of Family Physicians
Don't perform screening for cervical cancer in low-risk women aged 65 years or older and in women who have had a total hysterectomy for benign disease.	American College of Preventive Medicine
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks, 0 days gestational age.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks 0 days gestational age.	American College of Obstetricians and Gynecologists
Don't screen women younger than 30 years of age for cervical cancer with HPV testing, alone or in combination with cytology.	American Academy of Family Physicians
Avoid elective, non-medically indicated inductions of labor between 39 weeks, 0 days and 41 weeks, 0 days unless the cervix is deemed favorable.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor between 39 weeks 0 days and 41 weeks 0 days unless the cervix is deemed favorable.	American College of Obstetricians and Gynecologists
Don't perform Pap smears on women younger than 21 or who have had a hysterectomy for non-cancer disease.	American Academy of Family Physicians
Don't screen adolescents for scoliosis.	American Academy of Family Physicians
Don't perform voiding cystourethrogram (VCUG) routinely in first febrile urinary tract infection (UTI) in children aged 2 -24 months.	American Academy of Family Physicians
Don't perform imaging of the carotid arteries for simple syncope without other neurologic symptoms.	American Academy of Neurology
Don't recommend CEA for asymptomatic carotid stenosis unless the complication rate is low (<3%).	American Academy of Neurology
Don't perform electroencephalography (EEG) for headaches.	American Academy of Neurology
Don't prescribe interferon-beta or glatiramer acetate to patients with disability from progressive, non-relapsing forms of multiple sclerosis.	American Academy of Neurology
Don't automatically initiate continuous electronic fetal heart rate (FHR) monitoring during labor for women without risk factors; consider intermittent auscultation (IA) first.	American Academy of Nursing
Don't routinely use blood products to reverse warfarin.	American Association of Blood Banks
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated	American Society of Hematology

emergent surgery).	
Don't transfuse more units of blood than absolutely necessary.	American Association of Blood Banks
Don't transfuse more than the minimum number of red blood cell (RBC) units necessary to relieve symptoms of anemia or to return a patient to a safe hemoglobin range (7 to 8 g/dL in stable, non-cardiac in-patients).	American Society of Hematology
Don't perform stress cardiac imaging or advanced non-invasive imaging as a pre-operative assessment in patients scheduled to undergo low-risk non cardiac surgery.	American College of Cardiology
Don't obtain baseline diagnostic cardiac testing (trans-thoracic/esophageal echocardiography – TTE/TEE) or cardiac stress testing in asymptomatic stable patients with known cardiac disease (e.g., CAD, valvular disease) undergoing low or moderate risk non-cardiac surgery.	American Society of Anesthesiologists
Don't perform cardiac imaging as a pre-operative assessment in patients scheduled to undergo low- or intermediate- risk non-cardiac surgery.	American Society of Nuclear Cardiology
Don't perform stress CMR as a pre-operative assessment in patients scheduled to undergo low-risk, non-cardiac surgery.	Society for Cardiovascular Magnetic Resonance
Patients who have no cardiac history and good functional status do not require preoperative stress testing prior to non-cardiac thoracic surgery.	The Society of Thoracic Surgeons
Avoid cardiovascular testing for patients undergoing low-risk surgery.	Society for Vascular Medicine
Avoid computed tomography (CT) scans of the head in emergency department patients with minor head injury who are at low risk based on validated decision rules.	American College of Emergency Physicians
Avoid ordering a brain CT or brain MRI to evaluate an acute concussion unless there are progressive neurological symptoms, focal neurological findings on exam or there is concern for a skull fracture.	American Medical Society for Sports Medicine
Avoid instituting intravenous (IV) fluids before doing a trial or oral rehydration therapy in uncomplicated emergency department cases of mild to moderate dehydration in children.	American College of Emergency Physicians
Don't order low back X-rays as part of a routine preplacement medical examination.	American College of Occupational and Environmental Medicine
Don't prescribe opioids for treatment of chronic or acute pain for workers who perform safety-sensitive jobs such as operating motor vehicles, forklifts, cranes or other heavy equipment.	American College of Occupational and Environmental Medicine
Don't routinely order sleep studies (polysomnogram) to screen for/diagnose sleep disorders in workers suffering from chronic fatigue/insomnia.	American College of Occupational and Environmental Medicine
Don't routinely order X-ray for diagnosis of plantar fascitis/heel pain in employees who stand or walk at work.	American College of Occupational and Environmental Medicine
Don't initially obtain X-rays for injured workers with acute non-specific low back pain.	American College of Occupational and Environmental Medicine
Don't test ANA sub-serologies without a positive ANA and clinical suspicion of immune-mediated disease.	American College of Rheumatology
Don't order autoantibody panels unless positive antinuclear antibodies (ANA) and evidence of rheumatic disease.	American College of Rheumatology – Pediatric Rheumatology
Don't perform methotrexate toxicity labs more often	American College of Rheumatology – Pediatric

than every 12 weeks on stable doses.	Rheumatology
Don't perform MRI of the peripheral joints to routinely monitor inflammatory arthritis.	American College of Rheumatology
Don't routinely repeat DXA scans more often than once every two years.	American College of Rheumatology
Don't routinely perform surveillance joint radiographs to monitor juvenile idiopathic arthritis (JIA) disease activity.	American College of Rheumatology – Pediatric Rheumatology
Don't test for Lyme disease as a cause of musculoskeletal symptoms without an exposure history and appropriate exam findings.	American College of Rheumatology
Don't use antipsychotics as the first choice to treat behavioral and psychological symptoms of dementia.	American Geriatrics Society
Don't routinely use antipsychotics as first choice to treat behavioral and psychological symptoms of dementia.	American Psychiatric Association
Don't prescribe antipsychotic medications for behavioral and psychological symptoms of dementia (BPSD) in individuals with dementia without an assessment for an underlying cause of the behavior.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't treat with an anticoagulant for more than three months in a patient with a first venous thromboembolism (VTE) occurring in the setting of a major transient risk factor.	American Society of Hematology
Don't perform baseline or routine surveillance computed tomography (CT) scans in patients with asymptomatic, early-stage chronic lymphocytic leukemia (CLL).	American Society of Hematology
Don't use inferior vena cava (IVC) filters routinely in patients with acute VTE.	American Society of Hematology
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated emergent surgery).	American Society of Hematology
Don't routinely transfuse patients with sickle cell disease (SCD) for chronic anemia or uncomplicated pain crisis without an appropriate clinical indication.	American Society of Hematology
Don't test for thrombophilia in adult patients with venous thromboembolism (VTE) occurring in the setting of major transient risk factors (surgery, trauma or prolonged immobility).	American Society of Hematology
Don't test or treat for suspected heparin-induced thrombocytopenia (HIT) in patients with a low pre-test probability of HIT.	American Society of Hematology
Don't treat patients with immune thrombocytopenic purpura (ITP) in the absence of bleeding or a very low platelet count.	American Society of Hematology
Avoid using drains in breast reduction mammaplasty.	American Society of Plastic Surgeons
Avoid continuing prophylactic antibiotics for greater than 24 hours after a surgical procedure.	American Society of Plastic Surgeons
Avoid performing routine and follow-up mammograms of reconstructed breasts after mastectomies.	American Society of Plastic Surgeons
Avoid performing routine mammagrams before breast surgery.	American Society of Plastic Surgeons
Don't routinely use extended fractionation schemes (>10 fractions) for palliation of bone metastases.	American Society for Radiation Oncology
Don't initiate non-curative radiation therapy without defining the goals of treatment with the patient and	American Society for Radiation Oncology

considering palliative care referral.	
Don't recommend radiation following hysterectomy for endometrial cancer patients with low-risk disease.	American Society for Radiation Oncology
Don't use aloe vera on skin to prevent or treat radiodermatitis.	American Academy of Nursing
Don't use mixed medication mouthwash, commonly termed "magic mouthwash," to prevent or manage cancer treatment-induced oral mucositis.	American Academy of Nursing
Don't use L-carnitine/acetyl-L-carnitine supplements to prevent or treat symptoms of peripheral neuropathy in patients receiving chemotherapy for treatment of cancer.	American Academy of Nursing
Don't treat gastroesophageal reflux in infants routinely with acid suppression therapy.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't routinely use bronchodilators in children with bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't order chest radiographs in children with uncomplicated asthma or bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use continuous pulse oximetry routinely in children with acute respiratory illness unless they are on supplemental oxygen.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use systemic corticosteroids in children under 2 years of age with an uncomplicated lower respiratory tract infection.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't initiate routine evaluation of carotid artery disease prior to cardiac surgery in the absence of symptoms or other high-risk criteria.	The Society of Thoracic Surgeons
Don't perform a routine pre-discharge echocardiogram after cardiac valve replacement surgery.	The Society of Thoracic Surgeons

CWI: Choosing Wisely Initiative; GoR: Grad of recommendation

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	n.a.
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	n.a.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	n.a.
Bias	9	Describe any efforts to address potential sources of bias	8-9
Study size	10	Explain how the study size was arrived at	n.a.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	n.a.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n.a.
		(*b*) Describe any methods used to examine subgroups and interactions	n.a.
		(*c*) Explain how missing data were addressed	n.a.
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n.a.
		(*e*) Describe any sensitivity analyses	n.a.
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	7
		(b) Give reasons for non-participation at each stage	n.a.
		(c) Consider use of a flow diagram	n.a.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n.a.
		(b) Indicate number of participants with missing data for each variable of interest	n.a.
Outcome data	15*	Report numbers of outcome events or summary measures	7-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n.a.

		(*b*) Report category boundaries when continuous variables were categorized	n.a.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n.a.
Discussion			
Key results	18	Summarise key results with reference to study objectives	8
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	8-9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8, 10
Generalisability	21	Discuss the generalisability (external validity) of the study results	n.a.
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	11

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Choosing Wisely: assessment of current US top five list recommendations`trustworthiness using a pragmatic approach

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012366.R1
Article Type:	Research
Date Submitted by the Author:	23-Jun-2016
Complete List of Authors:	Horvath, Karl; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (2) Department of Internal Medicine, Division of Endocrinology and Diabetology Semlitsch, Thomas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Jeitler, Klaus; Medical University Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (3) Institute for Medical Informatics, Statistics and Documentation Abuzahra, Muna; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Posch, Nicole; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Domke, Andreas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Siebenhofer, Andrea; Institute of General Practice and Evidence-based Health Services Research, Medical University Graz; Institute for General Practice, Goethe University
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Public health, Health services research, Patient-centred medicine
Keywords:	Choosing Wisely, top five lists, trustworthiness, guidelines

Choosing Wisely: assessment of current US top five list recommendations` trustworthiness using a pragmatic approach

Authors: Karl Horvath (1,2) karl.horvath@medunigraz.at, Thomas Semlitsch (1) Thomas.semlitsch@medunigraz.at, Klaus Jeitler (1,3) klaus.jeitler@medunigraz.at, Muna E. Abuzahra (1) muna.abuzahara@medunigraz.at, Nicole Posch (1) nicole.posch@medunigraz.at, Andreas Domke (1) andreas.domke@medunigraz.at, Andrea Siebenhofer (1,4) andrea.siebenhofer@medunigraz.at

(1) Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz, Austria
(2) Department of Internal Medicine, Division of Endocrinology and Diabetology, Medical University of Graz, Austria
(3) Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
(4) Institute of General Practice, Goethe University Frankfurt am Main, Germany

Karl Horvath, MD

Auenbruggerplatz 2/9

8036 Graz, Austria

Tel: +43-316-385-73559

Email: karl.horvath@medunigraz.at

Key words: Choosing Wisely, top five lists, methodological quality, guidelines

Word count: 3521

ABSTRACT

Objectives: Identification of sufficiently trustworthy top five list recommendations from the US choosing wisely campaign..

Setting: Not applicable

Participants: All top five list recommendations available from the American Board of Internal Medicine Foundation website.

Main outcome measures/interventions: Compilation of US top five lists and search for current German highly trustworthy (S3) guidelines. Extraction of guideline recommendations, including grade of recommendation (GoR), for suggestions comparable to top five list recommendations. For recommendations without guideline equivalents, the methodological quality of the top five list development process was assessed using criteria similar to that used to judge guidelines, and relevant meta-literature was identified in cited references. Judgement of sufficient trustworthiness of top five list recommendations was based either on an "A" GoR of guideline equivalents or on high methodological quality and citation of relevant meta-literature.

Results: 412 top five list recommendations were identified. For 75 (18%), equivalents were found in current German S3 guidelines. 44 of these recommendations were associated with an "A" GoR, or a strong recommendation based on strong evidence, 26 had a "B" or a "C" GoR. No GoR was provided for 5 recommendations. 337 recommendations had no equivalent in the German S3 guidelines. The methodological quality of the development process was high and relevant meta-literature was cited for 87 top five list recommendations. For a further 36, either the methodological quality was high without any meta-literature citations, or meta-literature citations existed but the methodological quality was lacking. For the remaining 214 recommendations, either the methodological quality was lacking and no literature was cited, or the methodological quality was generally unsatisfactory.

Conclusions: 131 of current US top 5 list recommendations were found to be sufficiently trustworthy. For a substantial number of current US top five list recommendations, their trustworthiness remains unclear. Methodological requirements for developing top five lists are recommended.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This is a systematic assessment of the trustworthiness of all current top 5 recommendations from the US Choosing Wisely Initiative.
- Matching top 5 list recommendations with recommendations from trustworthy German S3 guidelines or assessing the methodological quality of the lists' development process together with quoted supporting meta-literature allowed for a safe identification of sufficiently trustworthy top 5 list recommendations.
- Only recommendations from the US campaign were considered.
- Underestimation of the trustworthiness of some recommendations might have occurred because recommendations were actually based on the best current evidence, but either no meta-literature was available or it was not quoted or no meta-literature but sufficient evidence from primary studies was available. Another source of possible misjudgement is that recommendations were actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient.

INTRODUCTION

The Choosing Wisely Initiative (CWI), a campaign led by the American Board of Internal Medicine (ABIM) Foundation, promotes doctor-patient communication and reducing waste in health care.[1] Within the initiative different medical societies develop and publish so called top five lists, naming (at least) five tests, interventions or services which are commonly overused in their respective specialties and should be questioned by doctors and patients. In light of the fact that for years rigorous guidelines have been published and yet they were not widely adopted or implemented in practice, a deliberately pragmatic approach was chosen to engage as many physicians and patients as possible. Because of this, only some loose methodological requirements for the development of top 5 lists were formulated, but among them was the prerequisite that all recommendations had to be evidence based. [1 2]

However, the campaign is currently experiencing some setbacks.[3] There is criticism and questions about the trustworthiness of the top five list recommendations because of the lack of comprehensive methodological requirements for the development of top five lists.[4] It was also noted that some lists might be influenced by financial self-interests.[5] To date only a few and limited attempts have been made to determine how evidence-based the available CWI recommendations are.[6-8] Uncertainty about the trustworthiness of the top five lists can impede the implementation of top five lists in daily practice.[9 10] Also, recommendations lacking a basis in evidence might not only not reduce waste but lead to possible harm. Trustworthy recommendations are necessary to minimize the chance for error in decisions made by patients, doctors and policymakers. Differentiating between sufficiently trustworthy recommendations and recommendations for which trustworthiness is unclear is also a key issue since top five lists will have increasing influence, as the Choosing Wisely campaign is being adopted in more countries.[11-13]

The aim of this study was to identify top five list recommendations from the US choosing wisely campaign which can be regarded as sufficiently trustworthy based on a pragmatic assessment approach

METHODS

We carried out a search for top five lists on the ABIM website on April 24th, 2015. All identified top five lists were included. From the available lists we extracted all stated recommendations, information on which medical society was responsible for developing the top five list, the methods used for their development, the rationale, and the cited supporting literature. Multiple items from different lists with nearly identical recommendations were combined and considered as one single item.

To assess the trustworthiness of top five list items, we aimed to identify equivalent recommendations in German S3 guidelines. We used German S3 guidelines with the following rationale: To be considered trustworthy, guidelines must meet certain quality criteria specified in the AGREE II instrument [14] or in the paper by Quaseem et al [15]. The Association of the Scientific Medical Societies

in Germany (AWMF) classifies guidelines into three categories: S1 expert recommendations developed by informal consensus, S2 guidelines requiring a formal consensus finding and/or a search for evidence and S3 denoting guidelines of the highest methodological quality. S3 guidelines must contain all elements of the AGREE II instrument, including a multidisciplinary development group, a systematic search for and a systematic appraisal of relevant literature, and a structured process for finding consensus. Thus all German S3 guidelines can a priori be considered trustworthy without further assessment. Also, in these guidelines, a sufficiently solid evidence base is a prerequisite for the highest "A" GoR. In the web portal of the AWMF all available German S 3 guidelines from many different medical specialist societies can be found. It thus allows for an efficient way of identifying highly trustworthy guidelines on a wide variety of medical topics. Also, a justified grade of recommendation (GoR) and the level of evidence (LoE) must be stated for every recommendation.[16,17] A high level of evidence is a prerequisite for the highest GoR. Thus recommendations from German S3 guidelines with such a high GoR can safely be regarded as evidence based. Top five list items for which such equivalent guideline recommendations exist would then be classified as trustworthy themselves. Guidelines will most likely differ regionally in regard to prioritization and importance of guideline topics and recommendations, because of differences in the health care system, ethnicities, local practice and so on. But as long as they have been developed in a way that assured a comprehensive structured consideration of the available evidence, all guidelines should agree on the evidence for or against a test or intervention. Thus while it might not be adequate to judge a US recommendation`s importance, with respect to its overuse, based on German guidelines, its evidence base can very well be judged using highly trustworthy German guidelines.

We conducted a search for all available German S3 guidelines in the web portal of the AWMF without restrictions concerning medical specialities or topics. We then matched the top five list recommendations with the identified current (as of the year 2015) guidelines based on the guidelines' title and the issuing medical societies. We only considered guideline items as equivalent to top 5 list recommendations if they referred directly to omitting tests or interventions, that is if they recommended against them. If a recommendation with a low GoR or insufficient evidence did not specifically state that a service should be avoided, we did not consider it to be equivalent to a top 5 list recommendation. Relevant guideline recommendations and their associated grade of recommendation were extracted. Because different guidelines used different terms for their grades of recommendations, a standardised GoR scheme was developed (table 1) and assigned to the respective recommendations. Matching and extraction was done by two authors independently and any differences were resolved by discussion.

Table 1: Standardised Grade of Recommendation

Standardised GoR	Strength of Recommendation in Guideline	Level of Evidence
A	**Strong recommendation** against a test, medical intervention or health care service based on strong solid evidence.	**Strong evidence** (e.g. systematic reviews of RCTs or level 1 diagnostic studies, individual RCTs)
B	**Recommendation** against a test, medical intervention or health care service based on moderate evidence.	**Moderate evidence** (e.g. systematic reviews of cohort studies or level >2 diagnostic studies, individual cohort studies, ecological studies)
C	**Recommendation** against a test, medical intervention or health care service based on expert consensus.	**No evidence possible** or sought
D	**No recommendation** for or against a test, medical intervention or health care service because of unclear or conflicting evidence.	**Weak evidence** (e.g. systematic reviews of case control studies or level 3b diagnostic studies, individual case control studies, case series, poor or non-independent reference standard, expert opinion)

A standardized GoR was then assigned to all top five list recommendations with guideline equivalents resulting in five categories (table 2). Top five list recommendations for which the equivalent in German S3 guidelines was a standardized "A" GoR were considered as trustworthy (category 1A in table 2, figure 1), because within the S3 guidelines a high GoR always reflects a high level of evidence (table 1). Top five list items with guideline equivalents associated with a lesser GoR were classified as being of unclear trustworthiness (figure 1).

Table 2: Categories of top five list recommendations

Categories	Criteria
1. CWI recommendations with corresponding recommendations from S3 guidelines	
1A	standardised GoR A
1B	standardised GoR B
1C	standardised GoR C
1D	standardised GoR D
1E	no GoR available
2. CWI recommendations without corresponding recommendations from S3 guidelines	
2A	high methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2B	high methodological quality but no supporting systematic meta-literature (SG, SR, MA, HTA) cited or moderate methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2C	moderate methodological quality and no supporting systematic meta-literature (SG, SR, MA, HTA) cited or low methodological quality

CWI: Choosing Wisely initiative; GoR: grade of recommendation; HTA: health technology assessment; MA: meta-analysis; SR: systematic review; SG: systematic guideline

In the case of top five list recommendations for which no guideline equivalent could be identified, we assessed the trustworthiness of the respective top five lists.For this, in a first step, we appraised the methodological quality of the development process of these lists using a validated rapid-assessment tool [4 18 19] based on criteria otherwise applied for the evaluation of guideline trustworthiness: systematic literature searches, involvement of a multidisciplinary group of experts, patient participation, management of conflicts of interests, method of consensus finding and planned updates.[4 18] We only considered information reported in the "How the list was developed" sections of the top five lists without additional searches for further information. Based on these criteria we judged the methodological quality of the development process as high (requirements fully or largely met), moderate (requirements partially met) or low (requirements not or mostly not met). In a second step, we searched the references quoted in the top 5 lists for supporting systematic meta-literature (meta-analyses, systematic reviews, health technology reports and evidence based guidelines utilizing systematic searches), because we hypothesised that the citation of such relevant meta-literature would increase the chance of a full consideration of the available evidence with appraisals of the effect sizes, the chance for bias and the consistency of results by the top five list authors. We evaluated the relevance of the identified meta-literature based on their full text publications. For top five list recommendations with a low quality development process, we omitted the meta-literature assessment. Quality assessment and assessment of the meta-literature was done by two authors independently and discrepancies were resolved by discussion. The resulting categories of top five list recommendations are shown in table 2.

Top five list recommendations were considered as sufficiently trustworthy if they came from a top five list with a high quality development process and supporting meta-literature was included in the lists' references (category 2A table 2, figure 1). Top five lists recommendations for which the top five list development process was judged to be of lesser quality and/or for which no supporting meta-literature was available from the reference lists were categorized to be of unclear trustworthiness. The classification process is summarized in figure 1.

Patient involvement

Patients were not involved in formulating the research question, the design or conduct of this study. Since patients were not involved in this investigation and no data linked to persons were used, this project was not reviewed by the ethics committee.

RESULTS

From the ABIM website, searched on April 24th 2015,[20] we identified 412 top five list recommendations developed by 66 different medical societies. Of these, 96 (23%) items represented nearly identical recommendations.

Top five list recommendations with S3 guideline equivalents

The search in the web portal of the AWMF (search date June 2nd 2015) yielded 139 methodologically high quality German S3 guidelines.[21] We excluded 23 guidelines because they were outdated (expiration dates before January 1st 2015).

For 75 (18%) top five list recommendations we identified guideline equivalents. For 9 recommendations we found equivalents in more than one (up to five) guideline. In these instances, we based our assessments on the guideline with the closest fit of content. 44 (11%) top five list recommendations were equivalent to a standardised "A" GoR, or a strong recommendation based on strong evidence. For 16 (4%) and 10 (2%) recommendations, the corresponding standardised GoR was "B" or "C" respectively. There were no recommendations classified as "D" GoR but 5 (1%) could not be classified because no GoR was available for their guideline equivalents (for all see figure 2).

We did not find any guideline recommendation contradicting its associated CWI recommendation.

Top five list recommendations without S3 guideline equivalents

The majority of the top 5 list recommendations, 337 or 82%, had no equivalent in current German S3 guidelines. For 103 (25%) recommendations we judged the methodological quality of the respective top five list`s development process as high. Relevant systematic meta-literature was included in the references lists of 87 (21%) of these recommendations. For further 36 (9%) recommendations, either the methodological quality of the top five list development process was high without citation of relevant meta-literature, or literature citations existed but the quality of the development process was only moderate. For the remaining 214 (52%) top five list recommendations, either the methodological quality of the respective top five lists was judged as moderate and no relevant meta-literature was cited, or the methodological quality was generally unsatisfactory (for all see figure 2).

Concerning the quality criteria (table 3), a systematic search was reported for 91 (22%) top five list recommendations. We found indications for patient participation in the development process for 17 (4%) and for the involvement of a multidisciplinary group of experts for 208 (50%) recommendations. An expiration date or information on planned updates was not given for any of the recommendations. Also, information concerning the management of potential conflicts of interests of top five list authors was not available for 16 (4%) recommendations. All remaining recommendations contained references only to the respective very general policies as stated on the websites of the different medical societies

but no specific information on potential conflicts of interests of the development group members. While for 328 (80%) recommendations some information on the process for formulating the recommendations was available, a structured, validated process was described only for 98 (24%) recommendations.

Table 3: Top five list recommendations without S3 guideline equivalents, methodological quality

	Systematic search (n)	**Multidisciplinary expert team** (n)	**Patient participation** (n)	**Structured consensus finding** (n)	**Management of CoI** (n)	**Expiration date** (n)
yes	91	208	17	98	0	0
no	184	129	320	239	16	337
unclear	62	0	0	0	321	0

CoI: conflict of interest

Trustworthiness of top five recommendations

Of all 412 available top five list recommendations, we judged 131 (32%) to be sufficiently trustworthy, 44 (11%) because their S3 guideline equivalents were associated with an "A" GoR indicating a strong recommendation with strong supporting evidence, and 87 (21%) because their methodological quality of the respective top five lists was high and relevant systematic meta-literature was cited in their support of the recommendation (figure 2 and supplementary material table A).

The trustworthiness of 281 top five list recommendations remained unclear.

DISCUSSION

Principal findings

Our study provides evidence that about a third of current US top five list recommendations up to April 2015 provide sufficiently trustworthy information on tests, interventions or services which are commonly overused. Methodological quality of the top five lists` development process varied considerably, especially with regard to conducting systematic searches for evidence, the methods for achieving a structured consensus, and the involvement of experts from multiple disciplines. Patient participation in the development of of top five lists, and information on the management of potential conflicts of interest were scarce.

While it is likely that the results reflect mainly the lack of adequate methodological requirements on how to develop top 5 lists,[4] other possible causes such as discrepancy of actual methods and their reporting, or financial self-interest[5], cannot be ruled out completely.

Strengths and limitations

All current top five list recommendations were included in our investigation. We systematically assessed the trustworthiness of the recommendations. Searching guidelines for equivalents identified recommendations with sufficient importance for daily practice. German S3 guidelines are required to incorporate all aspects of the AGREE II instrument and the given GoR in those guidelines always also reflects the quality and level of the underlying evidence. Thus we were able to judge top five list recommendations for which we identified guideline equivalents associated with the highest GoR (standardised GoR "A") as sufficiently trustworthy with a high level of certainty. A guideline GoR below "A" is an indication of uncertain or insufficient evidence and we thus judged the trustworthiness of top 5 list recommendations with equivalents which were associated with a GoR below "A" as unclear. Using only high quality S3 guidelines might also have resulted in an underestimation of the trustworthiness of recommendations for which good evidence but no S3 guidelines exist. Also, employing only German guidelines might have led us to underrate recommendations for which there are no equivalents in Germany, but would be available from highly trustworthy international guidelines. But since we did not a priori judge the trustworthiness of recommendations without guideline equivalents as unclear, but assessed them using a different method, this should not have resulted in misjudgement of many recommendations.

Top five list recommendations without S3 guideline equivalents were only judged as sufficiently trustworthy if a methodological quality of the top five lists` development process was found to be high. This was determined by applying indicators such as a transparent and structured development process including multidisciplinary experts and patients, and the quotation of supporting meta-literature. However, since we did not check whether additional meta-literature potentially contradicting the quoted references was available, the trustworthiness might have been overestimated in some cases. On the other hand, using this approach, it seems likely that we underestimated some of the recommendations for which the trustworthiness remained unclear because the respective top five lists were either of a lesser methodological quality or no meta-literature was quoted. This might be the case when recommendations which were actually based on the best current evidence, but either no meta-literature was available or it was not quoted. Also the trustworthiness of recommendations for which no meta-literature but sufficient evidence from primary studies was available might have been underestimated. Another source of possible misjudgement is that top five lists were was actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient. Also we considered only top five list recommendations from the US while many more countries have now started to produce their own[13].

To assess the trustworthiness of CWI recommendations without guideline equivalents with the highest level of certainty, it would be necessary to conduct systematic reviews, based on primary or secondary literature, for each of these recommendations. This is the only method to assure that all available

evidence will be considered, and the effect sizes and the likelihood of bias are sufficiently assessed.[22] But conducting such systematic reviews is highly time consuming. We thus used a pragmatic approach, based on the hypothesis that developing recommendations according to stringent methodological criteria[18] which are used in developing high-quality guidelines would suffice to assume a low likelihood of error.

In conclusion we think that our proposed method identifies trustworthy recommendations with a high specificity but a lesser sensitivity. Because of this, it was not possible to use the category "not trustworthy". Thus in the end we distinguished only between two categories, that is top five list recommendations with sufficient or unclear trustworthiness.

Comparison with other studies

To our knowledge, this is the first study to comprehensively assess the trustworthiness of all currently available US top five list recommendations. In a somewhat similar attempt Hipkins et al investigated the top five lists in regard to a thorough literature search and an evidence based process used in the development of the lists.[6] They considered the information given by the authors in the "How the list was developed" sections and any additional information from searches in MEDLINE, Google Scholar, relevant websites and publications. They found a description of some review of literature in more than a brief, non-specific way for only 20% to 35% of the lists they examined, and an evidence based process for about 38% of the lists. These results are in good accordance with our own findings. Gliwa and Pearson in their 2014 study did not assess the quality of the development process or reliability, but categorized the reported evidence according to the evidentiary rationales given by the top five list authors.[8] Institute for Clinical and Economical Review (ICER) reports[7] are only available for a small number of lists and the evaluation of the supporting evidence is based on the work by Gliwa and Pearson.

Potential implications for clinicians or policymakers

The lack of stringent standards for developing top five lists should not so much be viewed as a flaw, but rather as a necessary pragmatic approach for the campaign to gain momentum. But from the results of our study, it is clear that methodological requirements for the development of top five lists need to be formulated. An explicit, comprehensive consideration of the current best evidence and a transparent development should be mandatory. Attention should also be given to an adequate management of possible conflicts of interests and to patient participation. While an evidence based development process is imperative, additional criteria such as the extent of potential harm, disease severity and urgency, health resources consumption and others have to be considered when prioritizing recommendations to allow for a substantial impact on the health system. Better reporting is necessary. To keep top five lists concise, a comprehensive description might be given on the medical societies' websites with a link provided in the published lists.

New ways of developing top five lists, for example using big data or utilizing high quality guidelines[23] [24], need to be explored. Different groups have already developed new top five lists emphasising a solid evidence base, consideration of the potential impact and a structured transparent development process as important criteria. [25-27] While such an approach strengthens the trustworthiness of recommendations, the higher effort needed in their development will perhaps raise the barrier for creating and implementing top five lists. In the context of overuse, study results showing no differences between interventions are helpful findings in providing a solid evidence base for respective recommendations. Thus it is important that such negative studies are published.

Unanswered questions and future research

The proposed method for assessing the trustworthiness of top five list recommendations still needs to be validated, which we have planned as a follow-up project. The assessment also needs to be expanded to include international top 5 list recommendations and guidelines.

Contributors: KH, TS, KJ and AS designed the study. KH, TS, KJ, AS, MEA, NP and AD were involved in the conduct of the study, data analysis and interpretation. KH drafted the manuscript and TS, KJ, AS, MEA, NP and AD critically revised it for important intellectual content. KH is the guarantor.

Funding: This study was supported by the Techniker Krankenkasse, a German health insurance provider. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) KH, TS, MA, NP, AD, KJ, AS have support from the Techniker Krankenkasse, a German health insurance provider, for the submitted work; (2) KH, TS, MA, NP, AD, KJ, AS have no relationships with companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (4) KH, TS, MA, NP, AD, KJ, AS have no non-financial interests that may be relevant to the submitted work.

Ethical approval: No ethical approval was sought.

Data sharing: No additional data available.

REFERENCES

1. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely. About the Campaign. 2015 [updated 25.11.2015. Available from: http://www.choosingwisely.org/wp-content/uploads/2015/04/About-Choosing-Wisely.pdf.
2. Wolfson D, Santa J, Slass L. Engaging Physicians and Consumers in Conversations About Treatment Overuse and Waste: A Short History of the Choosing Wisely Campaign. Acad Med 2014;**89**(7):990-95.
3. Lenzer J. Choosing Wisely: setbacks and progress. BMJ 2015;**351**:h6760.
4. Strech D, Follmann M, Klemperer D, et al. When Choosing Wisely meets clinical practice guidelines [Wenn "Choosing Wisely" auf Leitlinien trifft]. Z Evid Fortbild Qual Gesundh wesen (ZEFQ) 2014;**108**(10):601-03.
5. Morden NE, Colla CH, Sequist TD, et al. Choosing wisely--the politics and economics of labeling low-value services. N Engl J Med 2014;**370**(7):589-92.
6. Hipkins B, Cuervo C, Barclay C, et al. Choosing Wisely: Progress in use of evidence to develop top 5 lists? [Poster]. Preventing Overdiagnosis, Winding back the harms of too much medicine; September 1-3, 2015; Bethesda, Maryland USA 2015.
7. American Board of Internal Medicine (ABIM) Foundation. ICER Baseline Reports [webpage] 2016 [Available from: http://www.choosingwisely.org/resources/icer-baseline-reports/.
8. Gliwa C, Pearson SD. Evidentiary rationales for the Choosing Wisely Top 5 lists. JAMA 2014;**311**(14):1443-4.
9. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;**282**(15):1458-65.
10. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004;**8**(6):iii-iv, 1-72.
11. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Gemeinsam Klug Entscheiden [webpage] 2016 [Available from: http://www.awmf.org/medizin-versorgung/gemeinsam-klug-entscheiden.html.
12. Malhotra A, Maughan D, Ansell J, et al. Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. BMJ 2015;**350**:h2308.
13. Levinson W, Kallewaard M, Bhatia RS, et al. 'Choosing Wisely': a growing international campaign. Bmj Qual Saf 2015;**24**(2):167-74.
14. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)-Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien" - 1. Auflage 2012 [updated 06.11.2012. Available from: http://www.awmf.org/leitlinien/awmf-regelwerk.html
15. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012;**156**(7):525-31.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien". 1. Auflage 2012 [updated 2012-11-06. Available from: http://www.awmf.org/leitlinien/awmf-regelwerk.html.
17. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. Klassifikation S3-Leitlinien 2012 [Available from: http://www.awmf.org/leitlinien/awmf-regelwerk/ll-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/klassifikation-s3.html.
18. Semlitsch T, Jeitler K, Kopp IB, et al. [Development of a workable mini checklist to assess guideline quality]. Z Evid Fortbild Qual Gesundhwes 2014;**108**(5-6):299-312.
19. Siebenhofer A, Semlitsch T, Herborn T, et al. Validation and reliability of a guideline appraisal mini-checklist for daily practice use. BMC Med Res Methodol 2016;**16**:39.
20. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely - Clinician Lists 2015 [Available from: http://www.choosingwisely.org/wp-content/uploads/2015/01/Choosing-Wisely-Recommendations.pdf.

21. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). AWMF online; Das Portal der wissenschaftlichen Medizin - Aktuelle Leitlinien 2015 [Available from: http://www.awmf.org/leitlinien/aktuelle-leitlinien.html.
22. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [update March 2011]: The Cochrane Collaboration; 2011 [Available from: www.cochrane-handbook.org.
23. Grad R, Pluye P, Tang D, et al. Patient-oriented evidence that matters (POEMs) suggest potential clinical topics for the Choosing Wisely campaign. J Am Board Fam Med 2015;**28**(2):184-9.
24. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Manual Entwicklung von Empfehlungen im Rahmen der Initiative Gemeinsam Klug Entscheiden (GKE) [webpage] 2016 [Available from: http://www.awmf.org/fileadmin/user_upload/Medizinische_Versorgung/GKE/Manual_GKE_AWMF_V1.0-Konsultationsfassung.pdf
25. Hicks LK, Bering H, Carson KR, et al. Five hematologic tests and treatments to question. Hematology Am Soc Hematol Educ Program 2014;**2014**(1):599-603.
26. McMahon LF, Jr., Beyth RJ, Burger A, et al. Enhancing patient-centered care: SGIM and choosing wisely. J Gen Intern Med 2014;**29**(3):432-3.
27. Schuur JD, Carney DP, Lyn ET, et al. 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.

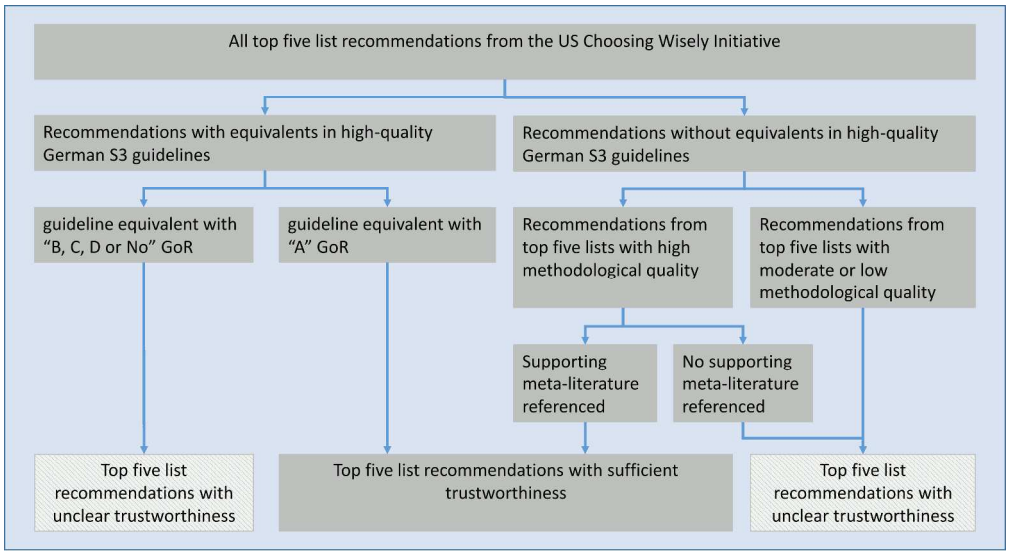


Figure 1: Is this top five list recommendation sufficiently trustworthy?

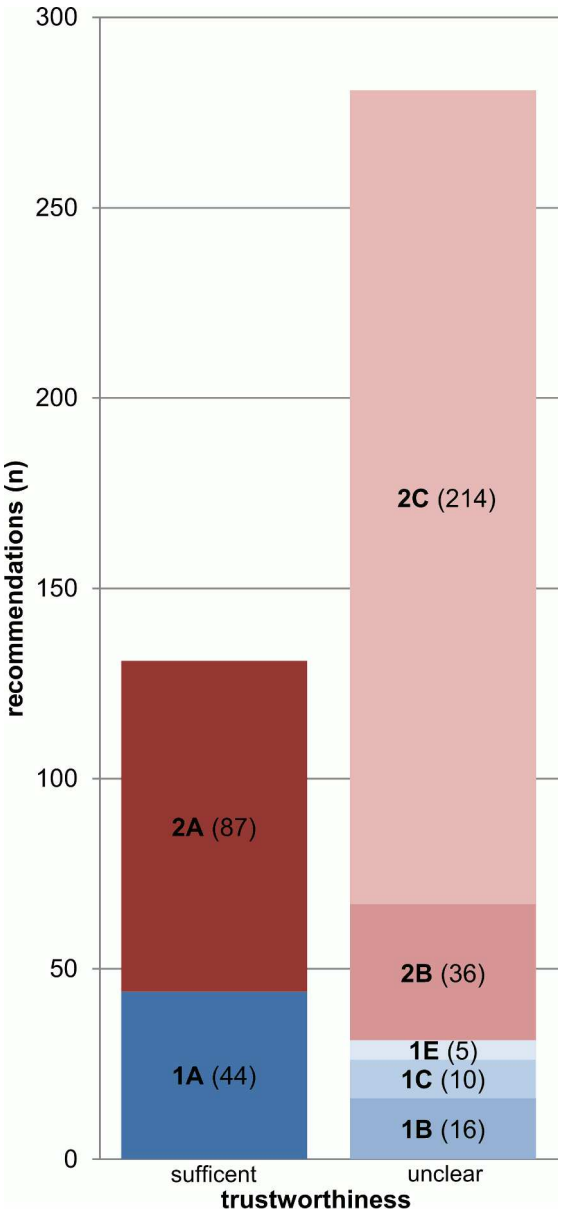


Figure 2: Trustworthiness of top five list recommendations. Blue columns represent top five list recommendations with guideline equivalents, red columns top five list recommendations without guideline equivalents. Numbers and letters in brackets denote different categories of top five recommendations (see table 2).

Table A: Top five list recommendations with sufficient reliability

Recommendation	Publishing Medical Society
CWI recommendations with S3 guideline equivalents associated with an "A" GoR	
Don't prescribe bed rest for acute localized back pain without completing an evaluation.	American Academy of Physical Medicine and Rehabilitation
Don't order an imaging study for back pain without performing a thorough physical examination.	American Academy of Physical Medicine and Rehabilitation
Avoid lumbar spine imaging in the emergency department for adults with non-traumatic back pain unless the patient has severe or progressive neurologic deficits or is suspected of having a serious underlying condition (such as vertebral infection, cauda equine syndrome, or cancer with bony metastasis).	American College of Emergency Physicians
Don't do imaging for low back pain within the first six weeks, unless red flags are present.	American Academy of Family Physicians
Don't obtain imaging (plain radiographs, magnetic resonance imaging, computed tomography [CT], or other advanced imaging) of the spine in patients with non-specific acute low back pain and without red flags.	American Association of Neurological Surgeons and Congress of Neurological Surgeons
Don't obtain imaging studies in patients with non-specific low back pain.	American College of Physicians
Avoid imaging studies (MRI, CT or X-rays) for acute low back pain without specific indications.	American Society of Anesthesiologists – Pain Medicine
Don't recommend advanced imaging (e.g., MRI) of the spine within the first six weeks in patients with non-specific acute low back pain in the absence of red flags.	North American Spine Society
Avoid prescribing antibiotics in the emergency department for uncomplicated sinusitis.	American College of Emergency Physicians
Don't order sinus computed tomography (CT) or indiscriminately prescribe antibiotics for uncomplicated acute rhinosinusitis.	American Academy of Allergy, Asthma & Immunology
Don't routinely prescribe antibiotics for acute mild-to-moderate sinusitis unless symptoms last for seven or more days, or symptoms worsen after initial clinical improvement.	American Academy of Family Physicians
Antibiotics should not be used for apparent viral respiratory illnesses (sinusitis, pharyngitis, bronchitis).	American Academy of Pediatrics
Avoid prescribing antibiotics for upper respiratory infections.	Infectious Diseases Society of America
Don't perform sentinel lymph node biopsy or other diagnostic tests for the evaluation of early, thin melanoma because they do not improve survival.	American Academy of Dermatology
Don't screen for carotid artery stenosis (CAS) in asymptomatic adult patients.	American Academy of Family Physicians
Don't routinely screen for prostate cancer using a prostate-specific antigen (PSA) test or digital rectal exam.	American Academy of Family Physicians
Don't routinely perform PSA-based screening for prostate cancer.	American College of Preventive Medicine
Don't perform PSA testing for prostate cancer screening in men with no symptoms of the disease when they are expected to live less than 10 years.	American Society of Clinical Oncology
Don't use post-operative splinting of the wrist after carpal tunnel release for long-term relief.	American Academy of Orthopaedic Surgeons
Don't perform annual stress cardiac imaging or advanced non-invasive imaging as part of routine	American College of Cardiology

follow-up in asymptomatic patients.	
Avoid performing routine stress testing after percutaneous coronary intervention (PCI) without specific clinical indications.	Society for Cardiovascular Angiography and Interventions
Don't perform routine annual stress testing after coronary artery revascularization.	Society of Nuclear Medicine and Molecular Imaging
Don't perform stress cardiac imaging or advanced non-invasive imaging in the initial evaluation of patients without cardiac symptoms unless high-risk markers are present.	American College of Cardiology
Don't perform cardiac imaging for patients who are at low risk.	American Society of Nuclear Cardiology
Don't perform stress cardiac imaging or coronary angiography in patients without cardiac symptoms unless high-risk markers are present.	American Society of Nuclear Cardiology
Avoid using stress echocardiograms on asymptomatic patients who meet "low risk" scoring criteria for coronary disease.	American Society of Echocardiography
Don't perform coronary CMR in the initial evaluation of asymptomatic patients.	Society for Cardiovascular Magnetic Resonance
Don't perform stress cardiovascular magnetic resonance (CMR) in the initial evaluation of chest pain patients with low pretest probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Don't screen for ovarian cancer in asymptomatic women at average risk.	American College of Obstetricians and Gynecologists
Don't screen low risk women with CA-125 or ultrasound for ovarian cancer.	Society of Gynecologic Oncology
Don't take a multi-vitamin, vitamin E or beta carotene to prevent cardiovascular disease or cancer.	American College of Preventive Medicine
Don't prescribe biologics for rheumatoid arthritis before a trial of methotrexate (or other conventional non-biologic DMARDs).	American College of Rheumatology
For a patient with functional abdominal pain syndrome (as per ROME III criteria) computed tomography (CT) scans should not be repeated unless there is a major change in clinical findings or symptoms.	American Gastroenterological Association
Don't use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present.	American Geriatrics Society
Don't treat asymptomatic bacteriuria with antibiotics.	Infectious Diseases Society of America
Avoid using PET or PET-CT scanning as part of routine follow-up care to monitor for a cancer recurrence in asymptomatic patients who have finished initial treatment to eliminate the cancer unless there is high-level evidence that such imaging will change the outcome.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early breast cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early prostate cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't initiate management of low-risk prostate cancer without discussing active surveillance.	American Society for Radiation Oncology
Don't recommend bed rest for more than 48 hours when treating low back pain.	North American Spine Society
Avoid coronary angiography in post-coronary artery bypass graft (CABG) and post-PCI patients who are asymptomatic, or who have normal or mildly abnormal	Society for Cardiovascular Angiography and Interventions

stress tests and stable symptoms not limiting quality of life.	
Don't perform stress CMR in patients with acute chest pain and high probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Avoid routine imaging for cancer surveillance in women with gynecologic cancer, specifically ovarian, endometrial, cervical, vulvar and vaginal cancer.	Society of Gynecologic Oncology
Patients with suspected or biopsy proven Stage I NSCLC do not require brain imaging prior to definitive care in the absence of neurologic symptoms.	The Society of Thoracic Surgeons
CWI recommendations without S3-guideline equivalents associated with good methodological quality and relevant meta-literature	
Avoid CT pulmonary angiography in emergency department patients with a low-pretest probability of pulmonary embolism and either a negative Pulmonary Embolism Rule-Out Criteria (PERC) or a negative D-dimer.	American College of Emergency Physicians
Don't perform chest computed tomography (CT angiography) to evaluate for possible pulmonary embolism in patients with a low clinical probability and negative results of a highly sensitive D-dimer assay.	American College of Chest Physicians and American Thoracic Society
Don't place an indwelling urinary catheter to manage urinary incontinence.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't place or maintain a urinary catheter in a patient unless there is a specific indication to do so.	American Academy of Nursing
Avoid placing indwelling urinary catheters in the emergency department for either urine output monitoring in stable patients who can void, or for patient or staff convenience.	American College of Emergency Physicians
Don't place, or leave in place, urinary catheters for incontinence or convenience or monitoring of output for non-critically ill patients (acceptable indications: critical illness, obstruction, hospice, preoperatively for <2 days for urologic procedures; use weights instead to monitor diuresis).	Society of Hospital Medicine – Adult Hospital Medicine
Don't initiate antihypertensive treatment in individuals ≥60 years of age for systolic blood pressure (SBP) <150 mm Hg or diastolic blood pressure (DBP) <90 mm Hg.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't recommend screening for breast, colorectal or prostate cancer if life expectancy is estimated to be less than 10 years.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid colorectal cancer screening tests on asymptomatic patients with a life expectancy of less than 10 years and no family or personal history of colorectal neoplasia.	American College of Surgeons
Don't recommend screening for breast, colorectal, prostate or lung cancer without considering life expectancy and the risks of testing, overdiagnosis and overtreatment.	American Geriatrics Society
Don't perform routine cancer screening for dialysis patients with limited life expectancies without signs or symptoms.	American Society of Nephrology
Don't recommend cancer screening in adults with life expectancy of less than 10 years.	Society of General Internal Medicine
Don't obtain a C. difficile toxin test to confirm "cure" if symptoms have resolved.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't insert percutaneous feeding tubes in individuals with advanced dementia. Instead, offer oral assisted	AMDA – The Society for Post-Acute and Long-Term Care Medicine

feedings.	
Don't use sliding scale insulin (SSI) for long-term diabetes management for individuals residing in the nursing home.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't obtain a urine culture unless there are clear signs and symptoms that localize to the urinary tract.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid the use of surveillance cultures for the screening and treatment of asymptomatic bacteruria.	American Academy of Pediatrics
Don't order annual electrocardiograms (EKGs) or any other cardiac screening for low-risk patients without symptoms.	American Academy of Family Physicians
Don't prescribe antibiotics for otitis media in children aged 2-12 years with non-severe symptoms where the observation option is reasonable.	American Academy of Family Physicians
Don't screen women older than 65 years of age for cervical cancer who have had adequate prior screening and are not otherwise at high risk for cervical cancer.	American Academy of Family Physicians
Don't perform screening for cervical cancer in low-risk women aged 65 years or older and in women who have had a total hysterectomy for benign disease.	American College of Preventive Medicine
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks, 0 days gestational age.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks 0 days gestational age.	American College of Obstetricians and Gynecologists
Don't screen women younger than 30 years of age for cervical cancer with HPV testing, alone or in combination with cytology.	American Academy of Family Physicians
Avoid elective, non-medically indicated inductions of labor between 39 weeks, 0 days and 41 weeks, 0 days unless the cervix is deemed favorable.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor between 39 weeks 0 days and 41 weeks 0 days unless the cervix is deemed favorable.	American College of Obstetricians and Gynecologists
Don't perform Pap smears on women younger than 21 or who have had a hysterectomy for non-cancer disease.	American Academy of Family Physicians
Don't screen adolescents for scoliosis.	American Academy of Family Physicians
Don't perform voiding cystourethrogram (VCUG) routinely in first febrile urinary tract infection (UTI) in children aged 2 -24 months.	American Academy of Family Physicians
Don't perform imaging of the carotid arteries for simple syncope without other neurologic symptoms.	American Academy of Neurology
Don't recommend CEA for asymptomatic carotid stenosis unless the complication rate is low (<3%).	American Academy of Neurology
Don't perform electroencephalography (EEG) for headaches.	American Academy of Neurology
Don't prescribe interferon-beta or glatiramer acetate to patients with disability from progressive, non-relapsing forms of multiple sclerosis.	American Academy of Neurology
Don't automatically initiate continuous electronic fetal heart rate (FHR) monitoring during labor for women without risk factors; consider intermittent auscultation (IA) first.	American Academy of Nursing
Don't routinely use blood products to reverse warfarin.	American Association of Blood Banks
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated	American Society of Hematology

emergent surgery).	
Don't transfuse more units of blood than absolutely necessary.	American Association of Blood Banks
Don't transfuse more than the minimum number of red blood cell (RBC) units necessary to relieve symptoms of anemia or to return a patient to a safe hemoglobin range (7 to 8 g/dL in stable, non-cardiac in-patients).	American Society of Hematology
Don't perform stress cardiac imaging or advanced non-invasive imaging as a pre-operative assessment in patients scheduled to undergo low-risk non cardiac surgery.	American College of Cardiology
Don't obtain baseline diagnostic cardiac testing (trans-thoracic/esophageal echocardiography – TTE/TEE) or cardiac stress testing in asymptomatic stable patients with known cardiac disease (e.g., CAD, valvular disease) undergoing low or moderate risk non-cardiac surgery.	American Society of Anesthesiologists
Don't perform cardiac imaging as a pre-operative assessment in patients scheduled to undergo low- or intermediate- risk non-cardiac surgery.	American Society of Nuclear Cardiology
Don't perform stress CMR as a pre-operative assessment in patients scheduled to undergo low-risk, non-cardiac surgery.	Society for Cardiovascular Magnetic Resonance
Patients who have no cardiac history and good functional status do not require preoperative stress testing prior to non-cardiac thoracic surgery.	The Society of Thoracic Surgeons
Avoid cardiovascular testing for patients undergoing low-risk surgery.	Society for Vascular Medicine
Avoid computed tomography (CT) scans of the head in emergency department patients with minor head injury who are at low risk based on validated decision rules.	American College of Emergency Physicians
Avoid ordering a brain CT or brain MRI to evaluate an acute concussion unless there are progressive neurological symptoms, focal neurological findings on exam or there is concern for a skull fracture.	American Medical Society for Sports Medicine
Avoid instituting intravenous (IV) fluids before doing a trial or oral rehydration therapy in uncomplicated emergency department cases of mild to moderate dehydration in children.	American College of Emergency Physicians
Don't order low back X-rays as part of a routine preplacement medical examination.	American College of Occupational and Environmental Medicine
Don't prescribe opioids for treatment of chronic or acute pain for workers who perform safety-sensitive jobs such as operating motor vehicles, forklifts, cranes or other heavy equipment.	American College of Occupational and Environmental Medicine
Don't routinely order sleep studies (polysomnogram) to screen for/diagnose sleep disorders in workers suffering from chronic fatigue/insomnia.	American College of Occupational and Environmental Medicine
Don't routinely order X-ray for diagnosis of plantar fascitis/heel pain in employees who stand or walk at work.	American College of Occupational and Environmental Medicine
Don't initially obtain X-rays for injured workers with acute non-specific low back pain.	American College of Occupational and Environmental Medicine
Don't test ANA sub-serologies without a positive ANA and clinical suspicion of immune-mediated disease.	American College of Rheumatology
Don't order autoantibody panels unless positive antinuclear antibodies (ANA) and evidence of rheumatic disease.	American College of Rheumatology – Pediatric Rheumatology
Don't perform methotrexate toxicity labs more often	American College of Rheumatology – Pediatric

than every 12 weeks on stable doses.	Rheumatology
Don't perform MRI of the peripheral joints to routinely monitor inflammatory arthritis.	American College of Rheumatology
Don't routinely repeat DXA scans more often than once every two years.	American College of Rheumatology
Don't routinely perform surveillance joint radiographs to monitor juvenile idiopathic arthritis (JIA) disease activity.	American College of Rheumatology – Pediatric Rheumatology
Don't test for Lyme disease as a cause of musculoskeletal symptoms without an exposure history and appropriate exam findings.	American College of Rheumatology
Don't use antipsychotics as the first choice to treat behavioral and psychological symptoms of dementia.	American Geriatrics Society
Don't routinely use antipsychotics as first choice to treat behavioral and psychological symptoms of dementia.	American Psychiatric Association
Don't prescribe antipsychotic medications for behavioral and psychological symptoms of dementia (BPSD) in individuals with dementia without an assessment for an underlying cause of the behavior.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't treat with an anticoagulant for more than three months in a patient with a first venous thromboembolism (VTE) occurring in the setting of a major transient risk factor.	American Society of Hematology
Don't perform baseline or routine surveillance computed tomography (CT) scans in patients with asymptomatic, early-stage chronic lymphocytic leukemia (CLL).	American Society of Hematology
Don't use inferior vena cava (IVC) filters routinely in patients with acute VTE.	American Society of Hematology
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated emergent surgery).	American Society of Hematology
Don't routinely transfuse patients with sickle cell disease (SCD) for chronic anemia or uncomplicated pain crisis without an appropriate clinical indication.	American Society of Hematology
Don't test for thrombophilia in adult patients with venous thromboembolism (VTE) occurring in the setting of major transient risk factors (surgery, trauma or prolonged immobility).	American Society of Hematology
Don't test or treat for suspected heparin-induced thrombocytopenia (HIT) in patients with a low pre-test probability of HIT.	American Society of Hematology
Don't treat patients with immune thrombocytopenic purpura (ITP) in the absence of bleeding or a very low platelet count.	American Society of Hematology
Avoid using drains in breast reduction mammaplasty.	American Society of Plastic Surgeons
Avoid continuing prophylactic antibiotics for greater than 24 hours after a surgical procedure.	American Society of Plastic Surgeons
Avoid performing routine and follow-up mammograms of reconstructed breasts after mastectomies.	American Society of Plastic Surgeons
Avoid performing routine mammagrams before breast surgery.	American Society of Plastic Surgeons
Don't routinely use extended fractionation schemes (>10 fractions) for palliation of bone metastases.	American Society for Radiation Oncology
Don't initiate non-curative radiation therapy without defining the goals of treatment with the patient and	American Society for Radiation Oncology

considering palliative care referral.	
Don't recommend radiation following hysterectomy for endometrial cancer patients with low-risk disease.	American Society for Radiation Oncology
Don't use aloe vera on skin to prevent or treat radiodermatitis.	American Academy of Nursing
Don't use mixed medication mouthwash, commonly termed "magic mouthwash," to prevent or manage cancer treatment-induced oral mucositis.	American Academy of Nursing
Don't use L-carnitine/acetyl-L-carnitine supplements to prevent or treat symptoms of peripheral neuropathy in patients receiving chemotherapy for treatment of cancer.	American Academy of Nursing
Don't treat gastroesophageal reflux in infants routinely with acid suppression therapy.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't routinely use bronchodilators in children with bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't order chest radiographs in children with uncomplicated asthma or bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use continuous pulse oximetry routinely in children with acute respiratory illness unless they are on supplemental oxygen.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use systemic corticosteroids in children under 2 years of age with an uncomplicated lower respiratory tract infection.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't initiate routine evaluation of carotid artery disease prior to cardiac surgery in the absence of symptoms or other high-risk criteria.	The Society of Thoracic Surgeons
Don't perform a routine pre-discharge echocardiogram after cardiac valve replacement surgery.	The Society of Thoracic Surgeons

CWI: Choosing Wisely Initiative; GoR: Grad of recommendation

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	n.a.
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	n.a.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	n.a.
Bias	9	Describe any efforts to address potential sources of bias	10 to 12
Study size	10	Explain how the study size was arrived at	n.a.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	n.a.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n.a.
		(*b*) Describe any methods used to examine subgroups and interactions	n.a.
		(*c*) Explain how missing data were addressed	n.a.
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n.a.
		(*e*) Describe any sensitivity analyses	n.a.
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	n.a.
		(c) Consider use of a flow diagram	n.a.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n.a.
		(b) Indicate number of participants with missing data for each variable of interest	n.a.
Outcome data	15*	Report numbers of outcome events or summary measures	8 to 10

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n.a.
		(*b*) Report category boundaries when continuous variables were categorized	n.a.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n.a.
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11, 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10, 12, 13
Generalisability	21	Discuss the generalisability (external validity) of the study results	n.a.
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Choosing Wisely: assessment of current US top five list recommendations`trustworthiness using a pragmatic approach

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012366.R2
Article Type:	Research
Date Submitted by the Author:	26-Jul-2016
Complete List of Authors:	Horvath, Karl; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (2) Department of Internal Medicine, Division of Endocrinology and Diabetology Semlitsch, Thomas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Jeitler, Klaus; Medical University Graz, (1) Institute of General Practice and Evidence-based Health Services Research; Medizinische Universitat Graz, (3) Institute for Medical Informatics, Statistics and Documentation Abuzahra, Muna; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Posch, Nicole; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Domke, Andreas; Medizinische Universitat Graz, (1) Institute of General Practice and Evidence-based Health Services Research Siebenhofer, Andrea; Institute of General Practice and Evidence-based Health Services Research, Medical University Graz; Institute for General Practice, Goethe University
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Public health, Health services research, Patient-centred medicine
Keywords:	Choosing Wisely, top five lists, trustworthiness, guidelines

Choosing Wisely: assessment of current US top five list recommendations` trustworthiness using a pragmatic approach

Authors: Karl Horvath (1,2) karl.horvath@medunigraz.at, Thomas Semlitsch (1) Thomas.semlitsch@medunigraz.at, Klaus Jeitler (1,3) klaus.jeitler@medunigraz.at, Muna E. Abuzahra (1) muna.abuzahara@medunigraz.at, Nicole Posch (1) nicole.posch@medunigraz.at, Andreas Domke (1) andreas.domke@medunigraz.at, Andrea Siebenhofer (1,4) andrea.siebenhofer@medunigraz.at

(1) Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz, Austria
(2) Department of Internal Medicine, Division of Endocrinology and Diabetology, Medical University of Graz, Austria
(3) Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria
(4) Institute of General Practice, Goethe University Frankfurt am Main, Germany

Karl Horvath, MD

Auenbruggerplatz 2/9

8036 Graz, Austria

Tel: +43-316-385-73559

Email: karl.horvath@medunigraz.at

Key words: Choosing Wisely, top five lists, methodological quality, guidelines

Word count: 3622

ABSTRACT

Objectives: Identification of sufficiently trustworthy top five list recommendations from the US choosing wisely campaign.

Setting: Not applicable

Participants: All top five list recommendations available from the American Board of Internal Medicine Foundation website.

Main outcome measures/interventions: Compilation of US top five lists and search for current German highly trustworthy (S3) guidelines. Extraction of guideline recommendations, including grade of recommendation (GoR), for suggestions comparable to top five list recommendations. For recommendations without guideline equivalents, the methodological quality of the top five list development process was assessed using criteria similar to that used to judge guidelines, and relevant meta-literature was identified in cited references. Judgement of sufficient trustworthiness of top five list recommendations was based either on an "A" GoR of guideline equivalents or on high methodological quality and citation of relevant meta-literature.

Results: 412 top five list recommendations were identified. For 75 (18%), equivalents were found in current German S3 guidelines. 44 of these recommendations were associated with an "A" GoR, or a strong recommendation based on strong evidence, 26 had a "B" or a "C" GoR. No GoR was provided for 5 recommendations. 337 recommendations had no equivalent in the German S3 guidelines. The methodological quality of the development process was high and relevant meta-literature was cited for 87 top five list recommendations. For a further 36, either the methodological quality was high without any meta-literature citations, or meta-literature citations existed but the methodological quality was lacking. For the remaining 214 recommendations, either the methodological quality was lacking and no literature was cited, or the methodological quality was generally unsatisfactory.

Conclusions: 131 of current US top five list recommendations were found to be sufficiently trustworthy. For a substantial number of current US top five list recommendations, their trustworthiness remains unclear. Methodological requirements for developing top five lists are recommended.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This is a systematic assessment of the trustworthiness of all current top five list recommendations from the US Choosing Wisely Initiative.
- Matching top five list recommendations with equivalents from trustworthy German S3 guidelines or assessing the methodological quality of the lists' development process together with quoted supporting meta-literature allowed for a safe identification of sufficiently trustworthy top five list recommendations.
- Only recommendations from the US campaign were considered.
- Underestimation of the trustworthiness of some recommendations might have occurred because recommendations were actually based on the best current evidence, but either no meta-literature was available or it was not quoted or no meta-literature but sufficient evidence from primary studies was available. Another source of possible misjudgement is that recommendations were actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient.

INTRODUCTION

The Choosing Wisely Initiative (CWI), a campaign led by the American Board of Internal Medicine (ABIM) Foundation, promotes doctor-patient communication and reducing waste in health care.[1] Within the initiative different medical societies develop and publish so called top five lists, naming (at least) five tests, interventions or services which are commonly overused in their respective specialties and should be questioned by doctors and patients. In light of the fact that for years rigorous guidelines have been published and yet they were not widely adopted or implemented in practice, a deliberately pragmatic approach was chosen to engage as many physicians and patients as possible. Because of this, only some loose methodological requirements for the development of top 5 lists were formulated, but among them was the prerequisite that all recommendations had to be evidence based. [1 2]

However, the campaign is currently experiencing some setbacks.[3] There is criticism and questions about the trustworthiness of the top five list recommendations because of the lack of comprehensive methodological requirements for the development of top five lists.[4] It was also noted that some lists might be influenced by financial self-interests.[5] To date only a few and limited attempts have been made to determine how evidence-based the available CWI recommendations are.[6-8] Uncertainty about the trustworthiness of the top five lists can impede the implementation of top five lists in daily practice.[9 10] Also, recommendations lacking a basis in evidence might not only not reduce waste but lead to possible harm. Trustworthy recommendations are necessary to minimize the chance for error in decisions made by patients, doctors and policymakers. Differentiating between sufficiently trustworthy recommendations and recommendations for which trustworthiness is unclear is also a key issue since top five lists will have increasing influence, as the Choosing Wisely campaign is being adopted in more countries.[11-13]

The aim of this study was to identify top five list recommendations from the US choosing wisely campaign which can be regarded as sufficiently trustworthy based on a pragmatic assessment approach

METHODS

We carried out a search for top five lists on the ABIM website on April 24th, 2015. All identified top five lists were included. From the available lists we extracted all stated recommendations, information on which medical society was responsible for developing the top five list, the methods used for their development, the rationale, and the cited supporting literature. Multiple items from different lists with nearly identical recommendations were combined and considered as one single item.

To assess the trustworthiness of top five list recommendations, we aimed to identify equivalent items in German S3 guidelines. We used German S3 guidelines with the following rationale: To be considered trustworthy, guidelines must meet certain quality criteria specified in the AGREE II instrument [14] or in the paper by Quaseem et al [15]. The Association of the Scientific Medical Societies

in Germany (AWMF) classifies guidelines into three categories: S1 expert recommendations developed by informal consensus, S2 guidelines requiring a formal consensus finding and/or a search for evidence and S3 denoting guidelines of the highest methodological quality. S3 guidelines must contain all elements of the AGREE II instrument, including a multidisciplinary development group, a systematic search for and a systematic appraisal of relevant literature, and a structured process for finding consensus. Thus all German S3 guidelines can a priori be considered trustworthy without further assessment. Also, in these guidelines, a sufficiently solid evidence base is a prerequisite for the highest "A" grade of recommendation (GoR). In the web portal of the AWMF all available German S 3 guidelines from many different medical specialist societies can be found. It thus allows for an efficient way of identifying highly trustworthy guidelines on a wide variety of medical topics. Also, a justified GoR and the level of evidence (LoE) must be stated for every guideline item.[16,17] A high level of evidence is a prerequisite for the highest GoR. Thus items from German S3 guidelines with such a high GoR can safely be regarded as evidence based. Top five list recommendations for which such guideline equivalents exist would then be classified as trustworthy themselves. Guidelines will most likely differ regionally in regard to prioritization and importance of guideline topics and items, because of differences in the health care system, ethnicities, local practice and so on. But as long as they have been developed in a way that assured a comprehensive structured consideration of the available evidence, all guidelines should agree on the evidence for or against a test or intervention. Thus while it might not be adequate to judge a US top five list recommendation`s importance, with respect to its overuse, based on German guidelines, its evidence base can very well be judged using highly trustworthy German guidelines.

We conducted a search for all available German S3 guidelines in the web portal of the AWMF without restrictions concerning medical specialities or topics. We then matched the top five list recommendations with the identified current (as of the year 2015) guidelines based on the guidelines' title and the issuing medical societies. We only considered guideline items as equivalent to top 5 list recommendations if they referred directly to omitting tests or interventions, that is if they recommended against them. If a guideline item with a low GoR or insufficient evidence did not specifically state that a service should be avoided, we did not consider it to be equivalent to a top five list recommendation. Relevant guideline items and their associated GoR were extracted. Because different guidelines used different terms for their GoR, a standardised GoR scheme was developed (table 1) and assigned to the respective items. Matching and extraction was done by two authors independently and any differences were resolved by discussion.

Table 1: Standardised Grade of Recommendation

Standardised GoR	Strength of Recommendation in Guideline	Level of Evidence
A	**Strong recommendation** against a test, medical intervention or health care service based on strong solid evidence.	**Strong evidence** (e.g. systematic reviews of RCTs or level 1 diagnostic studies, individual RCTs)
B	**Recommendation** against a test, medical intervention or health care service based on moderate evidence.	**Moderate evidence** (e.g. systematic reviews of cohort studies or level >2 diagnostic studies, individual cohort studies, ecological studies)
C	**Recommendation** against a test, medical intervention or health care service based on expert consensus.	**No evidence possible** or sought
D	**No recommendation** for or against a test, medical intervention or health care service because of unclear or conflicting evidence.	**Weak evidence** (e.g. systematic reviews of case control studies or level 3b diagnostic studies, individual case control studies, case series, poor or non-independent reference standard, expert opinion)

A standardized GoR was then assigned to all top five list recommendations with guideline equivalents resulting in five categories (table 2). Top five list recommendations for which the equivalent in German S3 guidelines was a standardized "A" GoR were considered as trustworthy (category 1A in table 2, figure 1), because within the S3 guidelines a high GoR always reflects a high level of evidence (table 1). Top five list recommendations with guideline equivalents associated with a lesser GoR were classified as being of unclear trustworthiness (figure 1).

Table 2: Categories of top five list recommendations

Categories	Criteria
1. CWI recommendations with corresponding equivalents from S3 guidelines	
1A	standardised GoR A
1B	standardised GoR B
1C	standardised GoR C
1D	standardised GoR D
1E	no GoR available
2. CWI recommendations without corresponding equivalents from S3 guidelines	
2A	high methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2B	high methodological quality but no supporting systematic meta-literature (SG, SR, MA, HTA) cited or moderate methodological quality and supporting systematic meta-literature (SG, SR, MA, HTA) cited
2C	moderate methodological quality and no supporting systematic meta-literature (SG, SR, MA, HTA) cited or low methodological quality

CWI: Choosing Wisely initiative; GoR: grade of recommendation; HTA: health technology assessment; MA: meta-analysis; SR: systematic review; SG: systematic guideline

In the case of top five list recommendations for which no guideline equivalent could be identified, we assessed the trustworthiness of the respective top five lists. For this, in a first step, we appraised the methodological quality of the development process of these lists using a validated rapid-assessment tool [4 18 19] based on criteria otherwise applied for the evaluation of guideline trustworthiness: systematic literature searches, involvement of a multidisciplinary group of experts, patient participation, management of conflicts of interests, method of consensus finding and planned updates.[4 18] We only considered information reported in the "How the list was developed" sections of the top five lists without additional searches for further information. Based on these criteria we judged the methodological quality of the development process as high (requirements fully or largely met), moderate (requirements partially met) or low (requirements not or mostly not met). In a second step, we searched the references quoted in the top five lists for supporting systematic meta-literature (meta-analyses, systematic reviews, health technology reports and evidence based guidelines utilizing systematic searches), because we hypothesised that the citation of such relevant meta-literature would increase the chance of a full consideration of the available evidence with appraisals of the effect sizes, the chance for bias and the consistency of results by the top five list authors. We evaluated the relevance of the identified meta-literature based on their full text publications. For top five list recommendations with a low quality development process, we omitted the meta-literature assessment. Quality assessment and assessment of the meta-literature was done by two authors independently and discrepancies were resolved by discussion. The resulting categories of top five list recommendations are shown in table 2.

Top five list recommendations were considered as sufficiently trustworthy if they came from a top five list with a high quality development process and supporting meta-literature was included in the lists' references (category 2A table 2, figure 1). Top five lists recommendations for which the top five list development process was judged to be of lesser quality and/or for which no supporting meta-literature was available from the reference lists were categorized to be of unclear trustworthiness. The classification process is summarized in figure 1.

Patient involvement

Patients were not involved in formulating the research question, the design or conduct of this study. Since patients were not involved in this investigation and no data linked to persons were used, this project was not reviewed by the ethics committee.

RESULTS

From the ABIM website, searched on April 24th 2015,[20] we identified 412 top five list recommendations developed by 66 different medical societies. Of these, 96 (23%) items represented nearly identical recommendations.

Top five list recommendations with S3 guideline equivalents

The search in the web portal of the AWMF (search date June 2nd 2015) yielded 139 methodologically high quality German S3 guidelines.[21] We excluded 23 guidelines because they were outdated (expiration dates before January 1st 2015).

For 75 (18%) top five list recommendations we identified guideline equivalents. For 9 recommendations we found equivalents in more than one (up to five) guideline. In these instances, we based our assessments on the guideline with the closest fit of content. 44 (11%) top five list recommendations were equivalent to a standardised "A" GoR, or a strong recommendation based on strong evidence. For 16 (4%) and 10 (2%) recommendations, the corresponding standardised GoR was "B" or "C" respectively. There were no recommendations classified as "D" GoR but 5 (1%) could not be classified because no GoR was available for their guideline equivalents (for all see figure 2).

We did not find any guideline items contradicting its associated top five list recommendation.

Top five list recommendations without S3 guideline equivalents

The majority of the top five list recommendations, 337 or 82%, had no equivalent in current German S3 guidelines. For 103 (25%) recommendations we judged the methodological quality of the respective top five list`s development process as high. Relevant systematic meta-literature was included in the references lists of 87 (21%) of these recommendations. For further 36 (9%) recommendations, either the methodological quality of the top five list development process was high without citation of relevant meta-literature, or literature citations existed but the quality of the development process was only moderate. For the remaining 214 (52%) top five list recommendations, either the methodological quality of the respective top five lists was judged as moderate and no relevant meta-literature was cited, or the methodological quality was generally unsatisfactory (for all see figure 2).

Concerning the quality criteria (table 3), a systematic search was reported for 91 (22%) top five list recommendations. We found indications for patient participation in the development process for 17 (4%) and for the involvement of a multidisciplinary group of experts for 208 (50%) recommendations. An expiration date or information on planned updates was not given for any of the recommendations. Also, information concerning the management of potential conflicts of interests of top five list authors was not available for 16 (4%) recommendations. All remaining recommendations contained references

only to the respective very general policies as stated on the websites of the different medical societies but no specific information on potential conflicts of interests of the development group members. While for 328 (80%) recommendations some information on the process for formulating the recommendations was available, a structured, validated process was described only for 98 (24%) recommendations.

Table 3: Top five list recommendations without S3 guideline equivalents, methodological quality

	Systematic search (n)	**Multidisciplinary expert team** (n)	**Patient participation** (n)	**Structured consensus finding** (n)	**Management of CoI** (n)	**Expiration date** (n)
yes	91	208	17	98	0	0
no	184	129	320	239	16	337
unclear	62	0	0	0	321	0

CoI: conflict of interest

Trustworthiness of top five recommendations

Of all 412 available top five list recommendations, we judged 131 (32%) to be sufficiently trustworthy, 44 (11%) because their S3 guideline equivalents were associated with an "A" GoR indicating a strong recommendation with strong supporting evidence, and 87 (21%) because their methodological quality of the respective top five lists was high and relevant systematic meta-literature was cited in their support of the recommendation (figure 2 and supplementary material table A).

The trustworthiness of 281 top five list recommendations remained unclear.

DISCUSSION

Principal findings

Our study provides evidence that about a third of current US top five list recommendations up to April 2015 provide sufficiently trustworthy information on tests, interventions or services which are commonly overused. Methodological quality of the top five lists` development process varied considerably, especially with regard to conducting systematic searches for evidence, the methods for achieving a structured consensus, and the involvement of experts from multiple disciplines. Patient participation in the development of top five lists, and information on the management of potential conflicts of interest were scarce.

While it is likely that the results reflect mainly the lack of adequate methodological requirements on how to develop top five lists,[4] other possible causes such as discrepancy of actual methods and their reporting, or financial self-interest[5], cannot be ruled out completely.

Strengths and limitations

All current top five list recommendations were included in our investigation. We systematically assessed the trustworthiness of the recommendations. Searching guidelines for equivalents identified recommendations with sufficient importance for daily practice. German S3 guidelines are required to incorporate all aspects of the AGREE II instrument and the given GoR in those guidelines always also reflects the quality and level of the underlying evidence. Thus we were able to judge top five list recommendations for which we identified guideline equivalents associated with the highest GoR (category "1A") as sufficiently trustworthy with a high level of certainty. A guideline GoR below "A" is an indication of uncertain or insufficient evidence and we thus judged the trustworthiness of top five list recommendations with guideline equivalents which were associated with a GoR below "A" as unclear (categories "1B, 1C, 1D, 1E"). Using only high quality S3 guidelines might also have resulted in an underestimation of the trustworthiness of recommendations for which good evidence but no S3 guidelines exist. Also, employing only German guidelines might have led us to underrate recommendations for which there are no equivalents in Germany, but would be available from highly trustworthy international guidelines. But since we did not a priori judge the trustworthiness of recommendations without guideline equivalents as unclear, but assessed them using a different method, this should not have resulted in misjudgement of many recommendations.

While at first sight it seems odd that equivalents in German guidelines were only identified for 18% of top five list recommendations, this finding becomes more plausible when one realises that in the AWMF-web portal alone over 700 guidelines can be found, but only 139 of them (around 18%) are S3 guidelines. Because of the methodological requirements for developing a S3 guideline, many guideline development groups settle for less methodologically robust S2 or S1 guidelines. Also there are further German guidelines not included in the AWMF portal. But since they could not a priori be considered methodologically sound, we did not consider them.

Top five list recommendations without S3 guideline equivalents were only judged as sufficiently trustworthy if a methodological quality of the top five lists` development process was found to be high. This was determined by applying indicators such as a transparent and structured development process including multidisciplinary experts and patients, and the quotation of supporting meta-literature (category "2A"). However, since we did not check whether additional meta-literature potentially contradicting the quoted references was available, the trustworthiness might have been overestimated in some cases. On the other hand, using this approach, it seems likely that we underestimated some of the recommendations for which the trustworthiness remained unclear because

the respective top five lists were either of a lesser methodological quality (category "2C") or no meta-literature was quoted (category "2B"). This might be the case when recommendations which were actually based on the best current evidence, but either no meta-literature was available or it was not quoted. Also the trustworthiness of recommendations for which no meta-literature but sufficient evidence from primary studies was available might have been underestimated. Another source of possible misjudgement is that top five lists were was actually developed in a structured way and based on evidence but the reporting on the methods used was insufficient. Also we considered only top five list recommendations from the US while many more countries have now started to produce their own[13].

To assess the trustworthiness of top five list recommendations without guideline equivalents with the highest level of certainty, it would be necessary to conduct systematic reviews, based on primary or secondary literature, for each of these recommendations. This is the only method to assure that all available evidence will be considered, and the effect sizes and the likelihood of bias are sufficiently assessed.[22] But conducting such systematic reviews is highly time consuming. We thus used a pragmatic approach, based on the hypothesis that developing recommendations according to stringent methodological criteria[18] which are used in developing high-quality guidelines would suffice to assume a low likelihood of error.

In conclusion we think that our proposed method identifies trustworthy recommendations (categories "1A" and "2A") with a high specificity but a lesser sensitivity. Because of this, it was not possible to use the category "not trustworthy". Thus in the end we distinguished only between two categories, that is top five list recommendations with sufficient or unclear trustworthiness.

Comparison with other studies

To our knowledge, this is the first study to comprehensively assess the trustworthiness of all currently available US top five list recommendations. In a somewhat similar attempt Hipkins et al investigated the top five lists in regard to a thorough literature search and an evidence based process used in the development of the lists.[6] They considered the information given by the authors in the "How the list was developed" sections and any additional information from searches in MEDLINE, Google Scholar, relevant websites and publications. They found a description of some review of literature in more than a brief, non-specific way for only 20% to 35% of the lists they examined, and an evidence based process for about 38% of the lists. These results are in good accordance with our own findings. Gliwa and Pearson in their 2014 study did not assess the quality of the development process or reliability, but categorized the reported evidence according to the evidentiary rationales given by the top five list authors.[8] Institute for Clinical and Economical Review (ICER) reports[7] are only available for a small number of lists and the evaluation of the supporting evidence is based on the work by Gliwa and Pearson.

Potential implications for clinicians or policymakers

The lack of stringent standards for developing top five lists should not so much be viewed as a flaw, but rather as a necessary pragmatic approach for the campaign to gain momentum. But from the results of our study, it is clear that methodological requirements for the development of top five lists need to be formulated. An explicit, comprehensive consideration of the current best evidence and a transparent development should be mandatory. Attention should also be given to an adequate management of possible conflicts of interests and to patient participation. While an evidence based development process is imperative, additional criteria such as the extent of potential harm, disease severity and urgency, health resources consumption and others have to be considered when prioritizing recommendations to allow for a substantial impact on the health system. Better reporting is necessary. To keep top five lists concise, a comprehensive description might be given on the medical societies' websites with a link provided in the published lists.

New ways of developing top five lists, for example using big data or utilizing high quality guidelines[23 24], need to be explored. Different groups have already developed new top five lists emphasising a solid evidence base, consideration of the potential impact and a structured transparent development process as important criteria.[25-27] While such an approach strengthens the trustworthiness of recommendations, the higher effort needed in their development will perhaps raise the barrier for creating and implementing top five lists. In the context of overuse, study results showing no differences between interventions are helpful findings in providing a solid evidence base for respective recommendations. Thus it is important that such negative studies are published.

Unanswered questions and future research

The proposed method for assessing the trustworthiness of top five list recommendations still needs to be validated, which we have planned as a follow-up project. The assessment also needs to be expanded to include international top five list recommendations and guidelines.

Contributors: KH, TS, KJ and AS designed the study. KH, TS, KJ, AS, MEA, NP and AD were involved in the conduct of the study, data analysis and interpretation. KH drafted the manuscript and TS, KJ, AS, MEA, NP and AD critically revised it for important intellectual content. KH is the guarantor.

Funding: This study was supported by the Techniker Krankenkasse, a German health insurance provider. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Competing interests: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that (1) KH, TS, MA, NP, AD, KJ, AS have support from the Techniker Krankenkasse, a German health insurance provider, for the submitted work; (2) KH, TS, MA, NP, AD, KJ, AS have no relationships with companies that might have an interest in the submitted work in the previous 3 years; (3) their spouses, partners, or children have no financial relationships that may be relevant to the submitted work; and (4) KH, TS, MA, NP, AD, KJ, AS have no non-financial interests that may be relevant to the submitted work.

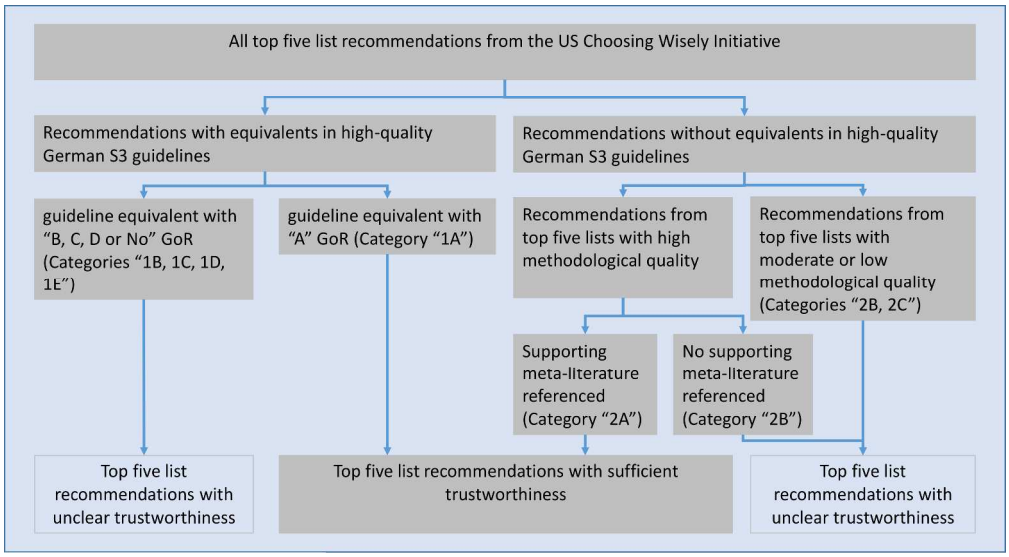
Ethical approval: No ethical approval was sought.

Data sharing: No additional data available.

REFERENCES

1. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely. About the Campaign. 2015 [updated 25.11.2015. Available from: http://www.choosingwisely.org/wp-content/uploads/2015/04/About-Choosing-Wisely.pdf.
2. Wolfson D, Santa J, Slass L. Engaging Physicians and Consumers in Conversations About Treatment Overuse and Waste: A Short History of the Choosing Wisely Campaign. Acad Med 2014;**89**(7):990-95.
3. Lenzer J. Choosing Wisely: setbacks and progress. BMJ 2015;**351**:h6760.
4. Strech D, Follmann M, Klemperer D, et al. When Choosing Wisely meets clinical practice guidelines [Wenn "Choosing Wisely" auf Leitlinien trifft]. Z Evid Fortbild Qual Gesundh wesen (ZEFQ) 2014;**108**(10):601-03.
5. Morden NE, Colla CH, Sequist TD, et al. Choosing wisely--the politics and economics of labeling low-value services. N Engl J Med 2014;**370**(7):589-92.
6. Hipkins B, Cuervo C, Barclay C, et al. Choosing Wisely: Progress in use of evidence to develop top 5 lists? [Poster]. Preventing Overdiagnosis, Winding back the harms of too much medicine; September 1-3, 2015; Bethesda, Maryland USA 2015.
7. American Board of Internal Medicine (ABIM) Foundation. ICER Baseline Reports [webpage] 2016 [Available from: http://www.choosingwisely.org/resources/icer-baseline-reports/.
8. Gliwa C, Pearson SD. Evidentiary rationales for the Choosing Wisely Top 5 lists. JAMA 2014;**311**(14):1443-4.
9. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. JAMA 1999;**282**(15):1458-65.
10. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess 2004;**8**(6):iii-iv, 1-72.
11. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Gemeinsam Klug Entscheiden [webpage] 2016 [Available from: http://www.awmf.org/medizin-versorgung/gemeinsam-klug-entscheiden.html.
12. Malhotra A, Maughan D, Ansell J, et al. Choosing Wisely in the UK: the Academy of Medical Royal Colleges' initiative to reduce the harms of too much medicine. BMJ 2015;**350**:h2308.
13. Levinson W, Kallewaard M, Bhatia RS, et al. 'Choosing Wisely': a growing international campaign. Bmj Qual Saf 2015;**24**(2):167-74.

14. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)-Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien“ - 1. Auflage 2012 [updated 06.11.2012. Available from: http://www.awmf.org/leitlinien/awmf-regelwerk.html
15. Qaseem A, Forland F, Macbeth F, et al. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med 2012;**156**(7):525-31.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. AWMF-Regelwerk „Leitlinien“. 1. Auflage 2012 [updated 2012-11-06. Available from: http://www.awmf.org/leitlinien/awmf-regelwerk.html.
17. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. Klassifikation S3-Leitlinien 2012 [Available from: http://www.awmf.org/leitlinien/awmf-regelwerk/ll-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/klassifikation-s3.html.
18. Semlitsch T, Jeitler K, Kopp IB, et al. [Development of a workable mini checklist to assess guideline quality]. Z Evid Fortbild Qual Gesundhwes 2014;**108**(5-6):299-312.
19. Siebenhofer A, Semlitsch T, Herborn T, et al. Validation and reliability of a guideline appraisal mini-checklist for daily practice use. BMC Med Res Methodol 2016;**16**:39.
20. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely - Clinician Lists 2015 [Available from: http://www.choosingwisely.org/wp-content/uploads/2015/01/Choosing-Wisely-Recommendations.pdf.
21. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). AWMF online; Das Portal der wissenschaftlichen Medizin - Aktuelle Leitlinien 2015 [Available from: http://www.awmf.org/leitlinien/aktuelle-leitlinien.html.
22. Higgins JPT, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [update March 2011]: The Cochrane Collaboration; 2011 [Available from: www.cochrane-handbook.org.
23. Grad R, Pluye P, Tang D, et al. Patient-oriented evidence that matters (POEMs) suggest potential clinical topics for the Choosing Wisely campaign. J Am Board Fam Med 2015;**28**(2):184-9.
24. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Manual Entwicklung von Empfehlungen im Rahmen der Initiative Gemeinsam Klug Entscheiden (GKE) [webpage] 2016 [Available from: "http://www.awmf.org/fileadmin/user_upload/Medizinische_Versorgung/GKE/Manual_GKE_AWMF_V1.0-Konsultationsfassung.pdf"]
25. Hicks LK, Bering H, Carson KR, et al. Five hematologic tests and treatments to question. Hematology Am Soc Hematol Educ Program 2014;**2014**(1):599-603.
26. McMahon LF, Jr., Beyth RJ, Burger A, et al. Enhancing patient-centered care: SGIM and choosing wisely. J Gen Intern Med 2014;**29**(3):432-3.
27. Schuur JD, Carney DP, Lyn ET, et al. 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.



Is this top five list recommendation sufficiently trustworthy?

1237x675mm (72 x 72 DPI)

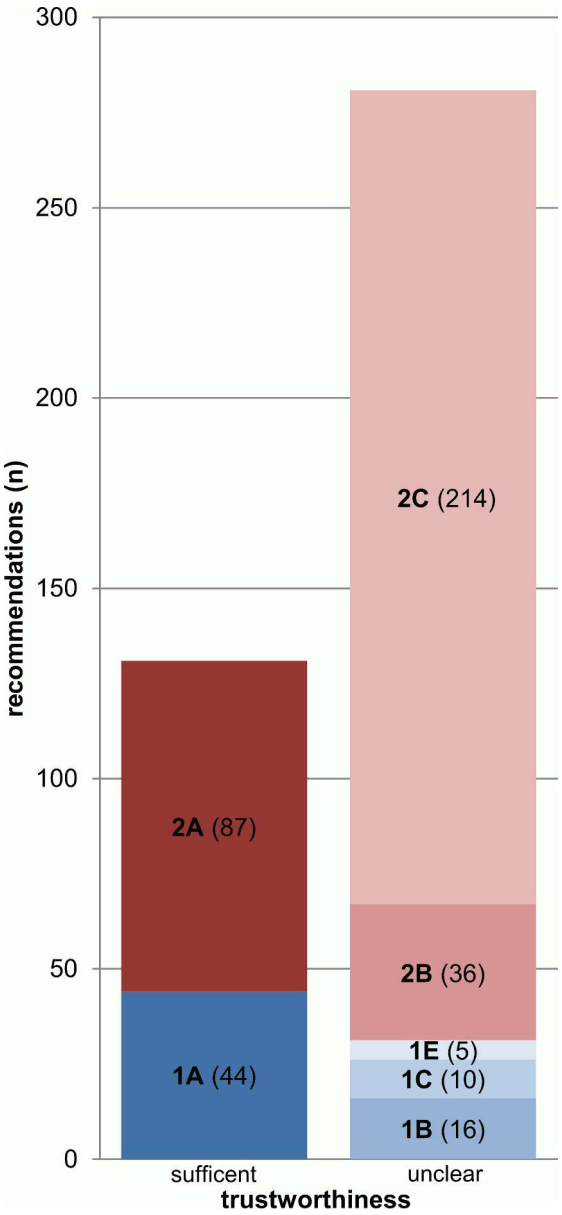


Figure 2: Trustworthiness of top five list recommendations. Blue columns represent top five list recommendations with guideline equivalents, red columns top five list recommendations without guideline equivalents. Numbers and letters in brackets denote different categories of top five recommendations (see table 2).

Table A: Top five list recommendations with sufficient reliability

Recommendation	Publishing Medical Society
CWI recommendations with S3 guideline equivalents associated with an "A" GoR (category "1A")	
Don't prescribe bed rest for acute localized back pain without completing an evaluation.	American Academy of Physical Medicine and Rehabilitation
Don't order an imaging study for back pain without performing a thorough physical examination.	American Academy of Physical Medicine and Rehabilitation
Avoid lumbar spine imaging in the emergency department for adults with non-traumatic back pain unless the patient has severe or progressive neurologic deficits or is suspected of having a serious underlying condition (such as vertebral infection, cauda equine syndrome, or cancer with bony metastasis).	American College of Emergency Physicians
Don't do imaging for low back pain within the first six weeks, unless red flags are present.	American Academy of Family Physicians
Don't obtain imaging (plain radiographs, magnetic resonance imaging, computed tomography [CT], or other advanced imaging) of the spine in patients with non-specific acute low back pain and without red flags.	American Association of Neurological Surgeons and Congress of Neurological Surgeons
Don't obtain imaging studies in patients with non-specific low back pain.	American College of Physicians
Avoid imaging studies (MRI, CT or X-rays) for acute low back pain without specific indications.	American Society of Anesthesiologists – Pain Medicine
Don't recommend advanced imaging (e.g., MRI) of the spine within the first six weeks in patients with non-specific acute low back pain in the absence of red flags.	North American Spine Society
Avoid prescribing antibiotics in the emergency department for uncomplicated sinusitis.	American College of Emergency Physicians
Don't order sinus computed tomography (CT) or indiscriminately prescribe antibiotics for uncomplicated acute rhinosinusitis.	American Academy of Allergy, Asthma & Immunology
Don't routinely prescribe antibiotics for acute mild-to-moderate sinusitis unless symptoms last for seven or more days, or symptoms worsen after initial clinical improvement.	American Academy of Family Physicians
Antibiotics should not be used for apparent viral respiratory illnesses (sinusitis, pharyngitis, bronchitis).	American Academy of Pediatrics
Avoid prescribing antibiotics for upper respiratory infections.	Infectious Diseases Society of America
Don't perform sentinel lymph node biopsy or other diagnostic tests for the evaluation of early, thin melanoma because they do not improve survival.	American Academy of Dermatology
Don't screen for carotid artery stenosis (CAS) in asymptomatic adult patients.	American Academy of Family Physicians
Don't routinely screen for prostate cancer using a prostate-specific antigen (PSA) test or digital rectal exam.	American Academy of Family Physicians
Don't routinely perform PSA-based screening for prostate cancer.	American College of Preventive Medicine
Don't perform PSA testing for prostate cancer screening in men with no symptoms of the disease when they are expected to live less than 10 years.	American Society of Clinical Oncology
Don't use post-operative splinting of the wrist after carpal tunnel release for long-term relief.	American Academy of Orthopaedic Surgeons
Don't perform annual stress cardiac imaging or advanced non-invasive imaging as part of routine	American College of Cardiology

follow-up in asymptomatic patients.	
Avoid performing routine stress testing after percutaneous coronary intervention (PCI) without specific clinical indications.	Society for Cardiovascular Angiography and Interventions
Don't perform routine annual stress testing after coronary artery revascularization.	Society of Nuclear Medicine and Molecular Imaging
Don't perform stress cardiac imaging or advanced non-invasive imaging in the initial evaluation of patients without cardiac symptoms unless high-risk markers are present.	American College of Cardiology
Don't perform cardiac imaging for patients who are at low risk.	American Society of Nuclear Cardiology
Don't perform stress cardiac imaging or coronary angiography in patients without cardiac symptoms unless high-risk markers are present.	American Society of Nuclear Cardiology
Avoid using stress echocardiograms on asymptomatic patients who meet "low risk" scoring criteria for coronary disease.	American Society of Echocardiography
Don't perform coronary CMR in the initial evaluation of asymptomatic patients.	Society for Cardiovascular Magnetic Resonance
Don't perform stress cardiovascular magnetic resonance (CMR) in the initial evaluation of chest pain patients with low pretest probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Don't screen for ovarian cancer in asymptomatic women at average risk.	American College of Obstetricians and Gynecologists
Don't screen low risk women with CA-125 or ultrasound for ovarian cancer.	Society of Gynecologic Oncology
Don't take a multi-vitamin, vitamin E or beta carotene to prevent cardiovascular disease or cancer.	American College of Preventive Medicine
Don't prescribe biologics for rheumatoid arthritis before a trial of methotrexate (or other conventional non-biologic DMARDs).	American College of Rheumatology
For a patient with functional abdominal pain syndrome (as per ROME III criteria) computed tomography (CT) scans should not be repeated unless there is a major change in clinical findings or symptoms.	American Gastroenterological Association
Don't use antimicrobials to treat bacteriuria in older adults unless specific urinary tract symptoms are present.	American Geriatrics Society
Don't treat asymptomatic bacteriuria with antibiotics.	Infectious Diseases Society of America
Avoid using PET or PET-CT scanning as part of routine follow-up care to monitor for a cancer recurrence in asymptomatic patients who have finished initial treatment to eliminate the cancer unless there is high-level evidence that such imaging will change the outcome.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early breast cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't perform PET, CT, and radionuclide bone scans in the staging of early prostate cancer at low risk for metastasis.	American Society of Clinical Oncology
Don't initiate management of low-risk prostate cancer without discussing active surveillance.	American Society for Radiation Oncology
Don't recommend bed rest for more than 48 hours when treating low back pain.	North American Spine Society
Avoid coronary angiography in post-coronary artery bypass graft (CABG) and post-PCI patients who are asymptomatic, or who have normal or mildly abnormal	Society for Cardiovascular Angiography and Interventions

stress tests and stable symptoms not limiting quality of life.	
Don't perform stress CMR in patients with acute chest pain and high probability of coronary artery disease.	Society for Cardiovascular Magnetic Resonance
Avoid routine imaging for cancer surveillance in women with gynecologic cancer, specifically ovarian, endometrial, cervical, vulvar and vaginal cancer.	Society of Gynecologic Oncology
Patients with suspected or biopsy proven Stage I NSCLC do not require brain imaging prior to definitive care in the absence of neurologic symptoms.	The Society of Thoracic Surgeons
CWI recommendations without S3-guideline equivalents associated with good methodological quality and relevant meta-literature (category "2A")	
Avoid CT pulmonary angiography in emergency department patients with a low-pretest probability of pulmonary embolism and either a negative Pulmonary Embolism Rule-Out Criteria (PERC) or a negative D-dimer.	American College of Emergency Physicians
Don't perform chest computed tomography (CT angiography) to evaluate for possible pulmonary embolism in patients with a low clinical probability and negative results of a highly sensitive D-dimer assay.	American College of Chest Physicians and American Thoracic Society
Don't place an indwelling urinary catheter to manage urinary incontinence.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't place or maintain a urinary catheter in a patient unless there is a specific indication to do so.	American Academy of Nursing
Avoid placing indwelling urinary catheters in the emergency department for either urine output monitoring in stable patients who can void, or for patient or staff convenience.	American College of Emergency Physicians
Don't place, or leave in place, urinary catheters for incontinence or convenience or monitoring of output for non-critically ill patients (acceptable indications: critical illness, obstruction, hospice, preoperatively for <2 days for urologic procedures; use weights instead to monitor diuresis).	Society of Hospital Medicine – Adult Hospital Medicine
Don't initiate antihypertensive treatment in individuals ≥60 years of age for systolic blood pressure (SBP) <150 mm Hg or diastolic blood pressure (DBP) <90 mm Hg.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't recommend screening for breast, colorectal or prostate cancer if life expectancy is estimated to be less than 10 years.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid colorectal cancer screening tests on asymptomatic patients with a life expectancy of less than 10 years and no family or personal history of colorectal neoplasia.	American College of Surgeons
Don't recommend screening for breast, colorectal, prostate or lung cancer without considering life expectancy and the risks of testing, overdiagnosis and overtreatment.	American Geriatrics Society
Don't perform routine cancer screening for dialysis patients with limited life expectancies without signs or symptoms.	American Society of Nephrology
Don't recommend cancer screening in adults with life expectancy of less than 10 years.	Society of General Internal Medicine
Don't obtain a C. difficile toxin test to confirm "cure" if symptoms have resolved.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't insert percutaneous feeding tubes in individuals with advanced dementia. Instead, offer oral assisted	AMDA – The Society for Post-Acute and Long-Term Care Medicine

feedings.	
Don't use sliding scale insulin (SSI) for long-term diabetes management for individuals residing in the nursing home.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't obtain a urine culture unless there are clear signs and symptoms that localize to the urinary tract.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Avoid the use of surveillance cultures for the screening and treatment of asymptomatic bacteruria.	American Academy of Pediatrics
Don't order annual electrocardiograms (EKGs) or any other cardiac screening for low-risk patients without symptoms.	American Academy of Family Physicians
Don't prescribe antibiotics for otitis media in children aged 2-12 years with non-severe symptoms where the observation option is reasonable.	American Academy of Family Physicians
Don't screen women older than 65 years of age for cervical cancer who have had adequate prior screening and are not otherwise at high risk for cervical cancer.	American Academy of Family Physicians
Don't perform screening for cervical cancer in low-risk women aged 65 years or older and in women who have had a total hysterectomy for benign disease.	American College of Preventive Medicine
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks, 0 days gestational age.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor or Cesarean deliveries before 39 weeks 0 days gestational age.	American College of Obstetricians and Gynecologists
Don't screen women younger than 30 years of age for cervical cancer with HPV testing, alone or in combination with cytology.	American Academy of Family Physicians
Avoid elective, non-medically indicated inductions of labor between 39 weeks, 0 days and 41 weeks, 0 days unless the cervix is deemed favorable.	American Academy of Family Physicians
Don't schedule elective, non-medically indicated inductions of labor between 39 weeks 0 days and 41 weeks 0 days unless the cervix is deemed favorable.	American College of Obstetricians and Gynecologists
Don't perform Pap smears on women younger than 21 or who have had a hysterectomy for non-cancer disease.	American Academy of Family Physicians
Don't screen adolescents for scoliosis.	American Academy of Family Physicians
Don't perform voiding cystourethrogram (VCUG) routinely in first febrile urinary tract infection (UTI) in children aged 2 -24 months.	American Academy of Family Physicians
Don't perform imaging of the carotid arteries for simple syncope without other neurologic symptoms.	American Academy of Neurology
Don't recommend CEA for asymptomatic carotid stenosis unless the complication rate is low (<3%).	American Academy of Neurology
Don't perform electroencephalography (EEG) for headaches.	American Academy of Neurology
Don't prescribe interferon-beta or glatiramer acetate to patients with disability from progressive, non-relapsing forms of multiple sclerosis.	American Academy of Neurology
Don't automatically initiate continuous electronic fetal heart rate (FHR) monitoring during labor for women without risk factors; consider intermittent auscultation (IA) first.	American Academy of Nursing
Don't routinely use blood products to reverse warfarin.	American Association of Blood Banks
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated	American Society of Hematology

emergent surgery).	
Don't transfuse more units of blood than absolutely necessary.	American Association of Blood Banks
Don't transfuse more than the minimum number of red blood cell (RBC) units necessary to relieve symptoms of anemia or to return a patient to a safe hemoglobin range (7 to 8 g/dL in stable, non-cardiac in-patients).	American Society of Hematology
Don't perform stress cardiac imaging or advanced non-invasive imaging as a pre-operative assessment in patients scheduled to undergo low-risk non cardiac surgery.	American College of Cardiology
Don't obtain baseline diagnostic cardiac testing (trans-thoracic/esophageal echocardiography – TTE/TEE) or cardiac stress testing in asymptomatic stable patients with known cardiac disease (e.g., CAD, valvular disease) undergoing low or moderate risk non-cardiac surgery.	American Society of Anesthesiologists
Don't perform cardiac imaging as a pre-operative assessment in patients scheduled to undergo low- or intermediate- risk non-cardiac surgery.	American Society of Nuclear Cardiology
Don't perform stress CMR as a pre-operative assessment in patients scheduled to undergo low-risk, non-cardiac surgery.	Society for Cardiovascular Magnetic Resonance
Patients who have no cardiac history and good functional status do not require preoperative stress testing prior to non-cardiac thoracic surgery.	The Society of Thoracic Surgeons
Avoid cardiovascular testing for patients undergoing low-risk surgery.	Society for Vascular Medicine
Avoid computed tomography (CT) scans of the head in emergency department patients with minor head injury who are at low risk based on validated decision rules.	American College of Emergency Physicians
Avoid ordering a brain CT or brain MRI to evaluate an acute concussion unless there are progressive neurological symptoms, focal neurological findings on exam or there is concern for a skull fracture.	American Medical Society for Sports Medicine
Avoid instituting intravenous (IV) fluids before doing a trial or oral rehydration therapy in uncomplicated emergency department cases of mild to moderate dehydration in children.	American College of Emergency Physicians
Don't order low back X-rays as part of a routine preplacement medical examination.	American College of Occupational and Environmental Medicine
Don't prescribe opioids for treatment of chronic or acute pain for workers who perform safety-sensitive jobs such as operating motor vehicles, forklifts, cranes or other heavy equipment.	American College of Occupational and Environmental Medicine
Don't routinely order sleep studies (polysomnogram) to screen for/diagnose sleep disorders in workers suffering from chronic fatigue/insomnia.	American College of Occupational and Environmental Medicine
Don't routinely order X-ray for diagnosis of plantar fascitis/heel pain in employees who stand or walk at work.	American College of Occupational and Environmental Medicine
Don't initially obtain X-rays for injured workers with acute non-specific low back pain.	American College of Occupational and Environmental Medicine
Don't test ANA sub-serologies without a positive ANA and clinical suspicion of immune-mediated disease.	American College of Rheumatology
Don't order autoantibody panels unless positive antinuclear antibodies (ANA) and evidence of rheumatic disease.	American College of Rheumatology – Pediatric Rheumatology
Don't perform methotrexate toxicity labs more often	American College of Rheumatology – Pediatric

than every 12 weeks on stable doses.	Rheumatology
Don't perform MRI of the peripheral joints to routinely monitor inflammatory arthritis.	American College of Rheumatology
Don't routinely repeat DXA scans more often than once every two years.	American College of Rheumatology
Don't routinely perform surveillance joint radiographs to monitor juvenile idiopathic arthritis (JIA) disease activity.	American College of Rheumatology – Pediatric Rheumatology
Don't test for Lyme disease as a cause of musculoskeletal symptoms without an exposure history and appropriate exam findings.	American College of Rheumatology
Don't use antipsychotics as the first choice to treat behavioral and psychological symptoms of dementia.	American Geriatrics Society
Don't routinely use antipsychotics as first choice to treat behavioral and psychological symptoms of dementia.	American Psychiatric Association
Don't prescribe antipsychotic medications for behavioral and psychological symptoms of dementia (BPSD) in individuals with dementia without an assessment for an underlying cause of the behavior.	AMDA – The Society for Post-Acute and Long-Term Care Medicine
Don't treat with an anticoagulant for more than three months in a patient with a first venous thromboembolism (VTE) occurring in the setting of a major transient risk factor.	American Society of Hematology
Don't perform baseline or routine surveillance computed tomography (CT) scans in patients with asymptomatic, early-stage chronic lymphocytic leukemia (CLL).	American Society of Hematology
Don't use inferior vena cava (IVC) filters routinely in patients with acute VTE.	American Society of Hematology
Don't administer plasma or prothrombin complex concentrates for non-emergent reversal of vitamin K antagonists (i.e. outside of the setting of major bleeding, intracranial hemorrhage or anticipated emergent surgery).	American Society of Hematology
Don't routinely transfuse patients with sickle cell disease (SCD) for chronic anemia or uncomplicated pain crisis without an appropriate clinical indication.	American Society of Hematology
Don't test for thrombophilia in adult patients with venous thromboembolism (VTE) occurring in the setting of major transient risk factors (surgery, trauma or prolonged immobility).	American Society of Hematology
Don't test or treat for suspected heparin-induced thrombocytopenia (HIT) in patients with a low pre-test probability of HIT.	American Society of Hematology
Don't treat patients with immune thrombocytopenic purpura (ITP) in the absence of bleeding or a very low platelet count.	American Society of Hematology
Avoid using drains in breast reduction mammaplasty.	American Society of Plastic Surgeons
Avoid continuing prophylactic antibiotics for greater than 24 hours after a surgical procedure.	American Society of Plastic Surgeons
Avoid performing routine and follow-up mammograms of reconstructed breasts after mastectomies.	American Society of Plastic Surgeons
Avoid performing routine mammagrams before breast surgery.	American Society of Plastic Surgeons
Don't routinely use extended fractionation schemes (>10 fractions) for palliation of bone metastases.	American Society for Radiation Oncology
Don't initiate non-curative radiation therapy without defining the goals of treatment with the patient and	American Society for Radiation Oncology

considering palliative care referral.	
Don't recommend radiation following hysterectomy for endometrial cancer patients with low-risk disease.	American Society for Radiation Oncology
Don't use aloe vera on skin to prevent or treat radiodermatitis.	American Academy of Nursing
Don't use mixed medication mouthwash, commonly termed "magic mouthwash," to prevent or manage cancer treatment-induced oral mucositis.	American Academy of Nursing
Don't use L-carnitine/acetyl-L-carnitine supplements to prevent or treat symptoms of peripheral neuropathy in patients receiving chemotherapy for treatment of cancer.	American Academy of Nursing
Don't treat gastroesophageal reflux in infants routinely with acid suppression therapy.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't routinely use bronchodilators in children with bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't order chest radiographs in children with uncomplicated asthma or bronchiolitis.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use continuous pulse oximetry routinely in children with acute respiratory illness unless they are on supplemental oxygen.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't use systemic corticosteroids in children under 2 years of age with an uncomplicated lower respiratory tract infection.	Society of Hospital Medicine – Pediatric Hospital Medicine
Don't initiate routine evaluation of carotid artery disease prior to cardiac surgery in the absence of symptoms or other high-risk criteria.	The Society of Thoracic Surgeons
Don't perform a routine pre-discharge echocardiogram after cardiac valve replacement surgery.	The Society of Thoracic Surgeons

CWI: Choosing Wisely Initiative; GoR: Grad of recommendation

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	page
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1, 2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	n.a.
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	n.a.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	n.a.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	n.a.
Bias	9	Describe any efforts to address potential sources of bias	10 to 12
Study size	10	Explain how the study size was arrived at	n.a.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	n.a.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	n.a.
		(*b*) Describe any methods used to examine subgroups and interactions	n.a.
		(*c*) Explain how missing data were addressed	n.a.
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	n.a.
		(*e*) Describe any sensitivity analyses	n.a.
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8
		(b) Give reasons for non-participation at each stage	n.a.
		(c) Consider use of a flow diagram	n.a.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	n.a.
		(b) Indicate number of participants with missing data for each variable of interest	n.a.
Outcome data	15*	Report numbers of outcome events or summary measures	8 to 10

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	n.a.
		(*b*) Report category boundaries when continuous variables were categorized	n.a.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n.a.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	n.a.
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11, 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10, 12, 13
Generalisability	21	Discuss the generalisability (external validity) of the study results	n.a.
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.