BMJ Open

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

Karl Horvath,[1,2] Thomas Semlitsch,[1] Klaus Jeitler,[1,3] Muna E Abuzahra,[1] Nicole Posch,[1] Andreas Domke,[1] Andrea Siebenhofer[1,4]

CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Karl Horvath; karl.horvath@medunigraz.at

ABSTRACT

Objectives: Identification of sufficiently trustworthy top 5 list recommendations from the US Choosing Wisely campaign.

Setting: Not applicable.

Participants: All top 5 list recommendations available from the American Board of Internal Medicine Foundation website.

Main outcome measures/interventions: Compilation of US top 5 lists and search for current German highly trustworthy (S3) guidelines. Extraction of guideline recommendations, including grade of recommendation (GoR), for suggestions comparable to top 5 list recommendations. For recommendations without guideline equivalents, the methodological quality of the top 5 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 5 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 5 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, and 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 5 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 5 list recommendations, their trustworthiness

---

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.

remains unclear. Methodological requirements for developing top 5 lists are recommended.

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 healthcare.[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 specialities 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 five 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 reduce waste and also lead to possible harm. Trustworthy recommendations are necessary to minimise 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 24 April 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 Qaseem 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 S3 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 must be stated for every guideline item.[14] [16] 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 prioritisation and importance of guideline topics and items, because of differences in the healthcare 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 recommendations' 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 five 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. Since 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 performed by two authors independently and any differences were resolved by discussion.

A standardised 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 standardised 'A' GoR were considered as trustworthy (category 1A in table 2, figure 1), because

Table 1 Standardised GoR

Standardised GoR	Strength of recommendation in guideline	Level of evidence
A	Strong recommendation against a test, medical intervention or healthcare service based on strong solid evidence	Strong evidence (eg, systematic reviews of RCTs or level 1 diagnostic studies, individual RCTs)
B	Recommendation against a test, medical intervention or healthcare service based on moderate evidence	Moderate evidence (eg, systematic reviews of cohort studies or level >2 diagnostic studies, individual cohort studies, ecological studies)
C	Recommendation against a test, medical intervention or healthcare service based on expert consensus	No evidence possible or sought
D	No recommendation for or against a test, medical intervention or healthcare service because of unclear or conflicting evidence	Weak evidence (eg, 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)

GoR, grade of recommendation; RCT, randomised controlled trial.

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; SG, systematic guideline; SR, systematic review.

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).

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 17 18] 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 17] 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 using 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 were performed 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 in table 2, figure 1). Top five list recommendations for which the top five list development process was judged to be of lesser quality and/or for which no supporting
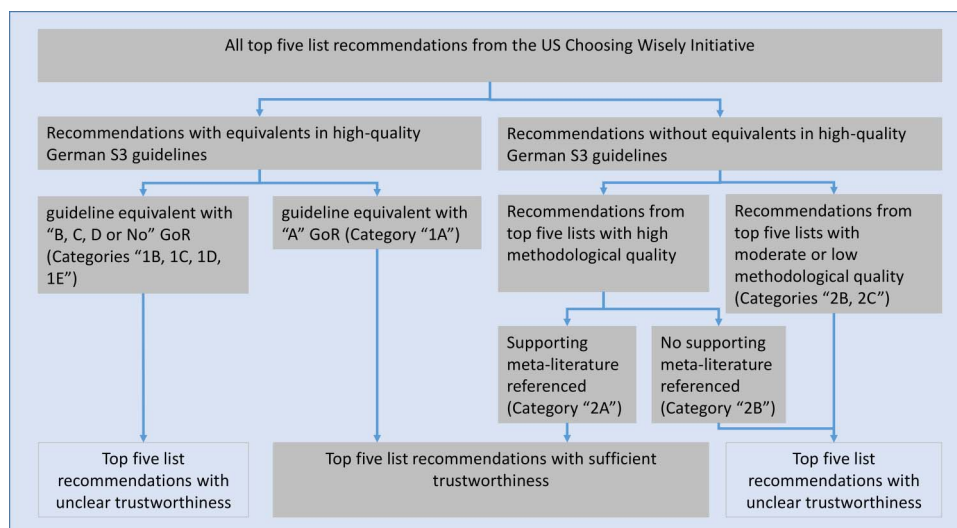
Figure 1 Is this top five list recommendation sufficiently trustworthy? (GoR, grade of recommendation).

meta-literature was available from the reference lists were categorised to be of unclear trustworthiness. The classification process is summarised 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 24 April 2015,[19] 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 2 June 2015) yielded 139 methodologically high-quality German S3 guidelines.[20] We excluded 23 guidelines because they were outdated (expiration dates before 1 January 2015).

For 75 (18%) top five list recommendations, we identified guideline equivalents. For nine 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. Forty-four (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 five (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 reference 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
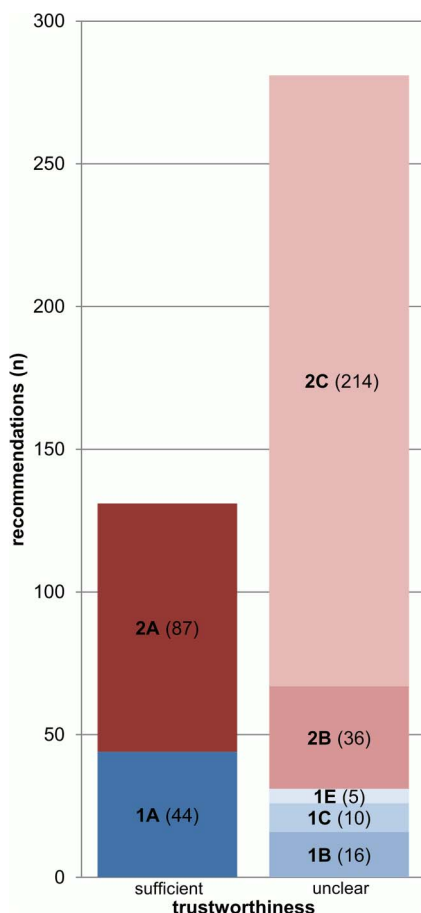
Figure 2 Trustworthiness of top five list recommendations. Blue columns represent top five list recommendations with guideline equivalents, and red columns represent top five list recommendations without guideline equivalents. Numbers in parentheses and letters denote different categories of top five recommendations (see table 2).

process for formulating the recommendations was available, a structured, validated process was described only for 98 (24%) recommendations.

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 see online supplementary table S1).

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 and 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

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.

guidelines exist. Also, employing only German guidelines might have led us to under-rate 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 that 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 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 USA 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.[21] 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[17] 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 the best of 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[6] investigated the top five lists in regard to a thorough literature search and an evidence-based process used in the development of the lists. 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 20–35% 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[8] in their 2014 study did not assess the quality of the development process or reliability, but categorised the reported evidence according to the evidentiary rationales given by the top five list authors. 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 prioritising 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 using high-quality guidelines,[22 23] 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.[24–26] 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.

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

REFERENCES

1. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely. About the Campaign, 2015 (updated 25 Nov 2015). 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;89:990–5.
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 Gesundhwes* 2014;108:601–3.
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;370:589–92.
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*; 1–3 September 2015; Bethesda, MD, USA, 2015.
7. American Board of Internal Medicine (ABIM) Foundation. ICER Baseline Reports [webpage], 2016. http://www.choosingwisely.org/resources/icer-baseline-reports/
8. Gliwa C, Pearson SD. Evidentiary rationales for the Choosing Wisely Top 5 lists. *JAMA* 2014;311:1443–4.
9. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999;282:1458–65.
10. Grimshaw JM, Thomas RE, MacLennan G, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:iii–iv, 1–72.
11. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Gemeinsam Klug Entscheiden [webpage], 2016. 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;24:167–74.
14. *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF)-Ständige Kommission Leitlinien. AWMF-Regelwerk "Leitlinien"—1.* Auflage, 2012 (updated 6 Nov 2012). 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;156:525–31.
16. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), Ständige Kommission Leitlinien. Klassifikation S3-Leitlinien, 2012. http://www.awmf.org/leitlinien/awmf-regelwerk/ll-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/klassifikation-s3.html
17. Semlitsch T, Jeitler K, Kopp IB, et al. [Development of a workable mini checklist to assess guideline quality]. *Z Evid Fortbild Qual Gesundhwes* 2014;108:299–312.
18. 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.
19. American Board of Internal Medicine (ABIM) Foundation. Choosing Wisely—clinician lists, 2015. http://www.choosingwisely.org/wp-content/uploads/2015/01/Choosing-Wisely-Recommendations.pdf
20. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). AWMF online; Das Portal der wissenschaftlichen Medizin—Aktuelle Leitlinien, 2015. http://www.awmf.org/leitlinien/aktuelle-leitlinien.html
21. Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions Version 5.1.0 (update Mar 2011).* The Cochrane Collaboration, 2011. http://www.cochrane-handbook.org
22. 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;28:184–9.
23. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF). Manual Entwicklung von Empfehlungen im Rahmen der Initiative Gemeinsam Klug Entscheiden (GKE) [webpage], 2016. http://www.awmf.org/fileadmin/user_upload/Medizinische_Versorgung/GKE/Manual_GKE_AWMF_V1.0-Konsultationsfassung.pdf
24. Hicks LK, Bering H, Carson KR, et al. Five hematologic tests and treatments to question. *Hematol Am Soc Hematol Educ Program* 2014;2014:599–603.
25. McMahon LF Jr, Beyth RJ, Burger A, et al. Enhancing patient-centered care: SGIM and choosing wisely. *J Gen Intern Med* 2014;29:432–3.
26. 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;174:509–15.