ONLINE SUPPLEMENTARY DATA 1

MATERIALS AND METHODS

Phase 1. The Canadian Rheumatology Association (CRA) established a steering committee to oversee the development of its Choosing Wisely Top 5 recommendations. Similar to the American College of Rheumatology (ACR) methodology, a multistage process combining consensus methodology and literature reviews was used. Our objective was to derive the list from practising rheumatologists based on their clinical experience and observations in their communities, and to reflect services directly relevant to the specialty. We attempted to engage the entire Canadian rheumatology workforce to begin a national conversation about this topic and to reach a national consensus. A literature review evaluated the quality of the scientific evidence supporting the final items appearing on the list.

A steering committee (SC, CT) solicited a group of 16 practising rheumatologists, a pediatric rheumatologist, an allied health professional, and a patient from across the country and from diverse clinical settings to form the Choosing Wisely Committee. The adult rheumatologists worked in either academic or community practises, were of different sex and ages, had different years in practise, and were geographically disperse. The patient volunteer had rheumatoid arthritis and regularly volunteers with an arthritis advocacy group. The committee generated candidate recommendations using the Delphi method. The Delphi process accommodated opinions from participants who were geographically dispersed and was anonymous so opinions could be expressed without bias[8]. This encouraged the overarching goal to engage rheumatologists and patients in a

meaningful discussion about high-value care. The instructions asked rheumatologists to reflect on the relevance of this list to current practise.

In round 1 of the Delphi, participants were asked to generate items and were provided examples of the ACR's Top 5 Choosing Wisely items. They were encouraged to suggest items that were commonly ordered by rheumatologists, among the most expensive or frequent services ordered or provided, and shown by current evidence not to provide meaningful benefit to at least some group for whom the test or treatment was provided. Supporting evidence to justify the item was solicited. Responses were organized by themes and statements were constructed using a uniform structure ("Do not perform…"). These statements were then used to create the round 2, web-based survey.

In round 2 of the Delphi, the items were ranked by the committee on a 5-point Likert scale based on their agreement of the suggestion (anchored 1 = strongly disagree to 5 = strongly agree), prevalence of the item in their community (anchored "seen very commonly" to "not seen commonly at all"), and highest effect on patients based on how it may change their practise and patient cost (anchored as "Yes" or "No" for should this be part of the Top 5 items). Comments, additional items, and suggestions for revision on each item were sought.

Data was aggregated and items with high content agreement (≥ 70% of those surveyed agreed), and either high prevalence (≥ 50% of those surveyed said they see this commonly) or high affect ranked in the top 20 items were used for the next Delphi round. A few items were revised and additional items were added.

In round 3 of the Delphi, participants were allowed to view the group results and change their ratings and rankings in light of their colleagues' responses. As in the

previous round, committee members rated each item based on agreement with content, prevalence, and effect on practice. Additional comments, suggestions, or revisions were also solicited. Statements with high subject agreement (> 80% agreement), and either high prevalence (> 50% of those surveyed said they see this commonly) or high affect (among top 20) were retained. Similar statements were combined and additional revisions for clarity were made. Items that were not in the realm of the rheumatologist were discarded.

Phase 2. This included inviting the entire membership of the CRA through e-mail to participate in a web-based survey. Although all the responses were anonymous, we asked demographic questions including sex, age range, practise location, and years in practise to assess whether the respondents were representative of the membership. Two e-mail reminders were sent and as an incentive to increase responses, anonymous respondents were entered into a draw for a prize (waived registration to the CRA Annual Scientific Meeting). Participants were asked to rate their agreement with the top items on a 5-point Likert scale (1 = anchored strongly disagree to 5 = strongly agree), rate whether they believed the item was high affect (yes/no) based on its prevalence, patient cost, and potential to reduce patient harm, and rank the items they believed rheumatologists should consider to be low-value care. In addition, they were able to comment on individual items and on the campaign in general.

Based on the results, we were able to calculate the highest combined rank in each of the 3 categories of subject agreement, affect, and rank. Qualitative analysis of the comments submitted by the CRA members were reviewed by members of the committee and these substantive comments helped revise the statements and identify major themes.

A methodology subcommittee discussed the items in light of their relevance to the health system, potential effect on patients, and the member survey results. The top 5 candidate items were selected based on item rating in the membership survey, adequate representation of diverse aspects of rheumatology practise, measurability, and potential affect. These items advanced for literature review.

A literature search for evidence supporting these items was performed in Cochrane Library, PubMed, and Embase from 2000–2013 (see additional Supplementary Material, available online at jrheum.org, for search strategy and specific search dates) with the help of medical reference librarians. In addition, an Internet grey-literature search was performed on Uptodate.com, Clinicalevidence.com, Guidelines.gov, and Tripdatabase.com since 2004. The search was limited to clinical practice guidelines, consensus statements, position papers, systematic reviews, and randomized controlled trials, and was completed between October to December 2013. Relevant papers from the reference list on included studies were also reviewed. Records were first screened by title and abstract, and then by full-text by 2 independent reviewers. Agreement was reached by consensus. Evidence reported for each of the 5 candidate items were conducted by 5 rheumatology fellows and 2 academic rheumatologists (MB, SC), and reviewed by the CRA Choosing Wisely Methodology subcommittee. Each report included summary tables of the key references and a summary paragraph. The subcommittee reviewed the evidence for each recommendation.

The final recommendations and evidence were then presented to the Choosing Wisely committee and key opinion leaders in the field. Suggestions were incorporated into the final wording of items and recommendations. The list was presented and

approved by the CRA Board of Directors. It was also presented to 3 patients for review (the original patient from the Choosing Wisely committee and 2 other patients, 1 with rheumatoid arthritis and the other with ankylosing spondylitis) who were members of the Canadian Arthritis Patient Alliance. Along with Consumer Reports, the largest, most trusted independent product testing organization in the world, these patient collaborators worked with the CRA to ensure the statements were translated into lay-language and made accessible to other patients and the public.

Grading evidence. We used a modified system developed by the Scottish Intercollegiate Guideline Network to grade evidence[9]. See Table below for custom system for assigning level of evidence and strength of recommendations.

Levels of Evidence		Strength of Recommendation	
I	Metaanalyses, systematic reviews of RCT, or individual RCT	A	Strong recommendation: • Direct level I evidence
II	Metaanalysis, systematic reviews of observational studies (cohort/case control studies), or individual observational studies OR RCT subgroup/post-hoc analyses	B	Moderate recommendation: • Direct level II evidence or extrapolated level I evidence
III	Nonanalytic studies, e.g., case reports, case series	C	Weak recommendation • Direct level III evidence or extrapolated level II evidence
IV	Expert opinion	D	Consensus recommendation:
NR	Recommendations are not linked to evidence		• Expert opinion based on very limited evidence

RCT: randomized controlled trial; NR: not reported.