



Published in final edited form as:

J Health Serv Res Policy. 2019 July ; 24(3): 201–206. doi:10.1177/1355819618815521.

THE CHALLENGE OF DETERMINING APPROPRIATE CARE IN THE ERA OF PATIENT-CENTERED CARE AND RISING HEALTH CARE COSTS

Coulter ID^{1,2}, Herman PM¹, Ryan GW¹, Hilton LJ¹, Hays RD², CERC team*

¹ RAND Corporation, Los Angeles

² University of California, Los Angeles

Abstract

Objective—To expand the appropriateness of care methodology to include patient preferences and resource utilization, and the impact of care appropriateness on patient outcomes.

Data Sources/Study Setting—Primary data from expert panels, focus groups, chiropractors, chiropractic patients with chronic low back pain (CLBP) and chronic cervical pain (CCP), and from internet “workers” via crowdsourcing. Study setting is a cluster sample of 125 chiropractic clinics from six US regions.

Study Design—This multicomponent methods study includes analysis of longitudinal data on patient outcomes, preferences, CLBP and CCP symptoms and healthcare utilization.

Data Collection/Extraction Methods—Data were collected bi-weekly for 3 months via online surveys that included both new and legacy measures, including PROMIS and CAHPS.

Principle Findings—Appropriateness panels generated ratings for 1800 CLBP and 744 CCP indications which will be applied to patient charts. Data from 2025 patients are being analyzed.

Conclusions—Patient-centered care is a significant policy initiative but translating it into policy that has been clinician and research-expert based, poses significant methodological issues. Nonetheless, we make the case that patient preferences, self-reported outcomes, and financial burden should be considered in the evaluation of the appropriateness of healthcare.

Introduction

The ultimate goal of all medical research is to ensure that patients receive appropriate care (i.e., “suitable or fitting for a particular purpose, person, occasion”).¹ In general, appropriate

Corresponding author: Ian D. Coulter, coulter@rand.org, (310) 393-0411 x. 7455, (310) 260-8175 (fax).

*CERC is a multi-disciplinary group of investigators from RAND, UCLA and the Samueli Institute and support staff. They conduct systematic reviews, evidence syntheses, derive indications for care, hold focus groups, expert panels, design data collection instruments, do economic modelling, recruit subjects nationally, analyze both qualitative and quantitative data, prepare reports and articles for submission. *Members include (alphabetically): Gursel Aliyev, Eric Apaydin, Judy Bearer, Marika Booth, Ian Coulter, Andrew Dick, Prodyumna Goutam, Carlos Gutierrez, Ron Hays, Patricia Herman, Lara Hilton, Eric Hurwitz, Praise Iyiewuare, Mallika Kommareddi, Lisa Kraus, Nishtha Mishra, Carol Roth, Carolyn Rutter, Gery Ryan, Paul Shekelle, Cathy Sherbourne, Melony Sorbero, Howard Vernon, and Margaret Whitley.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest.

care involves getting the right care for the right patient for the right problem at the right time from the right provider. It has also been shown that inappropriate care is costly. It is estimated that 30% of healthcare costs are wasted on inappropriate or useless care.² All patients should get appropriate care but the challenge comes in defining what that means (i.e., appropriateness), establishing the rate of inappropriate care and increasing delivery of appropriate care.

The introduction of patient-centered care in the Patient Protection and Affordable Care Act (commonly referred to as the Affordable Care Act, Public Law 111–148) and the establishment of the Patient-Centered Outcomes Research Institute have posed at least two new challenges to the traditional approach to measuring appropriateness. The first is the challenge of how to include patient preferences into what has largely been clinician-based decision making. A second challenge is a move away from explanatory randomized controlled trials (RCTs) to Comparative Effectiveness Research (CER) and a focus on what is important to patients, providers, and other stakeholders (effectiveness in everyday practice).³ The patient wants to know if it will work for me and the provider wants to know if it will work for my patients. Because of the strict protocols and inclusion/exclusion criteria of RCTs, neither patients nor providers can tell whether an intervention generalizes. Finally, with healthcare costs rising and more and more of the costs of care left as the responsibility of the patient, does the relative cost of care affect its appropriateness?

Determining What Is Appropriate Care

In the last decade, much attention has been given to evidenced-based practice (EBP) with the implicit assumption that this approach ensures appropriate care.^{4–6} This is care that has been scientifically shown to be efficacious and safe.⁷ Where there is clear evidence from a body of research, the appropriateness of care is relatively easily answered. However, across biomedicine there is considerable debate about what percentage of treatments can claim to be evidenced-based. As little as 15%–20% of all medical practice may be evidenced-based.^{8–11}

Hicks notes “it is generally accepted that between 20–60% of patients either receive inappropriate care or are not offered appropriate care.”¹² In addition, in large areas of health care, including complementary and alternative medicine (CAM), the evidence may be equivocal or contradictory.¹³

One of the driving forces historically that drew attention towards the examination of the appropriateness of care was the demonstration of variations (both small and large scale) in the amount of medical care delivered that could not be attributed to variations in those receiving it.^{14–17}

Surgical procedure rates were shown to have wide variation across different geographic areas. This led to investigations into the causes of those variations and while the incidence and prevalence of disease, socioeconomic factors, and underlying differences in the health care delivery system have all been shown to contribute, they do not adequately explain these

observed variations.¹⁸ One major factor found to contribute to variations in the health care delivered was the appropriateness of care.¹⁹

A second factor driving the interest in appropriateness was the failure of NIH Consensus Conferences to reach consensus on many indications for treatments. In the NIH approach consensus is required and the panels are secluded until consensus is reached. This means that where consensus does occur, it may be for the lowest common denominator. Finally, it was discovered that the recommendations and guidelines from NIH Consensus Conferences were not changing physician behavior.²⁰

These factors lead us to develop a method that would have more credibility with clinicians, utilize their clinical experience to expand upon the research evidence, address more indications that were relevant to the actual patients they were treating, and not force consensus. It resulted in the development of a formal and replicable method that drew on both traditional evidence but also on clinical experience and acumen.²¹

The Appropriateness Method

In recognition of the real and widespread limitations in evidence, of the need to make good clinical decisions even in the absence of a clear evidence base, and of the need to have more influence on clinician behavior, we pioneered a method to study the appropriateness of care which takes advantage of the available evidence base, but also draws upon the clinical acumen/experience of practitioners.^{22,14, 23}

This approach uses a mixed clinical expert and researcher panel to consider the available evidence and to then judge for a particular treatment, whether “for an average group of patients presenting [with this set of clinical indications] to an average U.S. physician ... the expected health benefit exceeds the expected negative consequences by a sufficiently wide margin that the procedure is worth doing ... excluding considerations of monetary cost”.¹⁴

To date this has been the most widely used and studied method for defining and identifying appropriate care in the U.S., and it has also been used internationally.^{6, 24, 25} Studies have investigated the relationship between the **literature and the ratings**; the **reliability of the ratings**, and **content validity**.^{26–29} We have also conducted **reliability studies** of the panel process (replicating the same panels but with different panel members).^{23, 30}

As shown in Figure 1, the Appropriateness Method begins with a systematic review of the literature (step 1), which includes studies of efficacy and effectiveness.^{21, 31} Based on the review, the research staff and subject matter experts create a set of indications (combinations of patient symptoms, past medical history, and test results) for performing the given clinical procedure or using a drug or supplement (step 2). The goal for these indications is to be comprehensive—i.e., cover all likely patient presentations. However, the list of indications must also be short enough that each indication can be separately and realistically evaluated for appropriateness by the panelists.

In step 3, the appropriateness of the procedure for each indication is rated by a nine-member panel consisting of content experts and clinicians. Ratings range from 1 to 9, with 1

representing that the procedure is *extremely inappropriate* for that patient type, and 9 that the procedure is *extremely appropriate*. Appropriateness is determined by assessing whether “the expected health benefit to the patient (relief of symptoms, improved functional capacity, reduction of anxiety etc.) exceed expected health risks (pain, discomfort etc.) by a sufficiently wide margin that the procedure is worth doing”—i.e., the efficacy/effectiveness and safety of the procedure for each indication. Clinical experience and acumen are encouraged in the process by instructing panelists to evaluate the risks and benefits based on commonly accepted best clinical practice for the year in which the panel is conducted. After panelists individually rate appropriateness for each of the indication, they are brought together in a face-to-face meeting and given the opportunity to rerate the indications after group discussion (step 4).

Once the work of the expert panel is completed, a set of indications now exists for doing a procedure for which there is consensus (step 5), and indications where there was agreement that the grounds for doing a procedure are indeterminate and indications for which there was no agreement are identified (step 6). The indications are based on the evidence and clinical experience, and can then be used to guide practice but also as a standard against which to compare current practice. This process permits the calculation of rates of appropriate and inappropriate care (step 7). To measure this, files are randomly chosen from a practice and compared to the indications to determine care appropriateness.

Going beyond the Appropriateness Method

The approach to appropriateness makes it feasible to take the best of what is known from research, combine it with the expertise of experienced clinicians, and apply it over the wide range of patients and presentations seen in real-world clinical practice. Clinicians are, after all, the final translators of evidence into practice, and this approach formalizes the process. A major limitation of the approach, however, is that it utilizes a limited definition of appropriateness; one that relies heavily on safety, efficacy and effectiveness. In contrast, the proceedings of an international World Health Organization workshop justifiably describe appropriateness as a “complex issue.” They state that: “Most definitions of appropriateness address ... that care is effective (based on valid evidence); efficient (cost-effective); and consistent with the ethical principles and preferences of the relevant individual, community or society”.³² Thus, there is room to improve the methods for defining and identifying appropriate care.

The Study of the Appropriateness of Care in CAM

We are currently funded to study the expansion of the appropriateness method to include patient preferences and resource utilization (costs). It also examines the impact of care appropriateness on outcomes, which has rarely been directly measured.

The focus is on methods for including patient preferences and resource use in the determination of the appropriateness of chiropractic manipulation and mobilization (M/M) for chronic low back pain (CLBP) and chronic cervical pain (CCP). However, the methods

developed may be also applicable for other types of CAM and other conditions, and for health care in general.

The choice of chiropractic was driven by our previous studies demonstrating the feasibility of working with and recruiting from chiropractic practices, the fact that chiropractors are an oft-used type of CAM for CLBP and CCP, and that the one previous appropriateness study of CAM was performed for chiropractic for low back pain.

An estimated 76.5 million adult Americans reported experiencing pain at some point in their life that lasted more than 24 hours; of those reporting pain, 42 percent said it lasted more than a year³³. Spine disorders are the leading cause of disability in the United States and world-wide. The 2010 Global Burden of Disease study reported comprehensive information on the worldwide impact of 291 diseases, injuries and risk factors. Low back pain is the number one cause of years lived with disability and neck pain is number four.³⁴ Low back pain is ubiquitous in Western society.³⁵⁻³⁷ Its lifetime prevalence is generally accepted to be around 80% and is estimated to be one of the most costly of all medical conditions.^{35, 38, 39}

The Bone and Joint Decade 2000–2010 Task Force noted that in the U.S. the estimated incremental costs of health-care expenditures for adults with spine problems in the U.S. in 2005 were \$85.9 billion, representing 9% of the total national expenditure LBP ranks among the 10 most expensive medical conditions.⁴⁰

The Research Design

The Center has four large scale research projects (R01s) which utilize the services of two cores:

- Project 1: Clinician-Based Appropriateness
- Project 2: Outcomes-Based Appropriateness
- Project 3: Patient Preferences-Based Appropriateness
- Project 4: Resource Utilization-Based Appropriateness
- Systematic Review Core
- Data Collection and Management Core

Our focus project (Project 1: Clinician-Based Appropriateness) applies the Appropriateness Method to determine appropriate care for M/M for chronic low back pain (CLBP) and chronic cervical pain (CCP). This approach uses a mixed (CAM and non-CAM) panel of expert clinicians and researchers to interpret and synthesize all available data on safety, efficacy and effectiveness²¹ and apply those data and their clinical experience to the question of the appropriateness of M/M for CLBP and CCP. This approach has previously been applied to M/M for low back pain⁴¹; and a literature review and expert panel were done for M/M for cervical pain.^{23, 29, 30}⁴² Since these initial studies, performed in the mid-1990s, the evidence base for M/M has expanded considerably.

Appropriateness Panels for CLBP and for CCP have been held and panelists successfully reviewed the literature and rated a comprehensive set of patient indications (1800 different indications for CLBP and 744 indications for CCP) as to whether M/M is appropriate or inappropriate for CLBP and for CCP. The main objectives of Project 1 are to develop these ratings and then to apply them to a national sample of patients' charts (collected through the Data Collection and Management Core) to determine the prevalence of appropriate and inappropriate care in practice. The systematic reviews of the evidence regarding M/M for CLBP and CCP were performed by the Systematic Review Core. The panels' ratings provide feedback to the research community regarding the patient presentations where substantial uncertainty remains as to the appropriateness of M/M, and feedback as to whether guidelines are needed to increase appropriate and decrease inappropriate care.

The three other projects in the Center add to the method in different ways by exploring the broader dimensions of appropriateness from the patient perspective.

Project 2 (Outcomes-Based Appropriateness) examines the relationship of appropriateness of care to patient experiences with care and health-related quality of life. Its objectives include adapting existing state-of-the science measures for patients using M/M for CLBP and CCP and the evaluation of the relationships of appropriate care with these patient-reported measures using data collected by our Data Collection and Management Core from our national sample. The Systematic Review Core performed literature reviews in support of the outcome measures used. We also conducted focus groups and cognitive interviews to help evaluate candidate survey instruments and develop supplemental items meaningful to patients and included domains relevant to them.

Given the prevalence of patient self-referral and the health system-wide focus on patient-centered care, Project 3 (Patient Preference-Based Appropriateness) examines how patient preferences affect what is considered to be appropriate care. Objectives for this study include understanding how CLBP and CCP patients decide to use M/M and determining what they believe is appropriate care. Again, the Systematic Review Core performed literature reviews in support of the measures of patient preference used, and data collected from our national sample by the Data Collection and Management Core will be analyzed to determine patient preferences for M/M care.

Finally, given the high proportion of out-of-pocket costs in CAM and the dramatic rise of healthcare costs in general, Project 4 (Resource Utilization-Based Appropriateness) examines how cost affects both which type of care is most appropriate and the appropriate course of care once a type of care has been chosen. Objectives here include determining the relative cost-effectiveness of care options for CLBP and CCP (these results will be based on two economic simulation models; the reviews of the literature supporting model building were performed by the Systematic Review Core), and analyzing longitudinal data from our national sample, collected by the Data Collection and Management Core, to begin to understand the characteristics of an appropriate course of care once M/M is chosen.

The last act in the project will be to reassemble the two expert panels (chronic low back pain and chronic neck pain) and provide them with the patient-centered information collected

(outcomes, preferences and resources) from our longitudinal patient sample and ask the panels to rerate the appropriateness of M/M for each clinical indication. We are adapting the grade methodology used to get from evidence to decision making to do this.^{7, 43}

Data Collection and Management Core

As noted above the data collection for Projects 1, 2, 3 and 4 is being conducted within a single data core. We used a national cluster sample to recruit chiropractic practices across 6 states: California, Florida, Minnesota, New York, Oregon, and Texas. The states were chosen to ensure geographical representation of the major areas of the U.S. and to ensure variation in chiropractic practice that might be determined by state laws and regulations. In each state a single metropolitan center was chosen and approximately 20 chiropractic clinics were recruited. In each practice patients were recruited over a four-week period through the use of an onsite iPad, and if eligible, asked to complete online surveys at baseline and three months, and shorter surveys every two weeks in between. We successfully recruited 2025 patients from 125 clinics across the 6 states. Retention across the 3-month data collection period was excellent; 91 percent of those who completed baseline also completed the 3-month survey and respondents completed on average four of the five bi-weekly surveys.

The final stage of data collection, not yet completed, will be to scan patient files of those who participated (with their consent) and to scan the files of a random retrospective sample of patients at each clinic to establish the amount of chronic pain being treated in a given clinic. Once scanned, the data from the files will be abstracted and compared to our appropriateness ratings to determine the overall rates of appropriate and inappropriate care.

Crowdsourcing as a Supplement to Appropriateness Methods

One innovative way to increase representativeness, efficiency, and cost of data collection, analysis, and patient inclusion in panel processes may be to engage patients using web-based, labor portals. Crowdsourcing first emerged in 2006 as an online contract labor market where needed services, ideas, or content were obtained for pay from a large group of people.⁴⁴ More recently, platforms such as Amazon's Mechanical Turk have been tested for feasibility of use for medical, psychological, and behavioral sciences research and shown to provide reliable data faster and more cheaply as well as affording access to a broader cross-section of the population than is typical of standard research experiments, including online panels.⁴⁵⁻⁵⁰

We will leverage these projects and data core to explore the feasibility of using crowdsourcing for low-cost patient engagement in the expert panel process through a connected grant. We will test comparability and reliability of crowdsourced data to patient data collected within the study, develop and test a crowdsourcing approach for patient inclusion in qualitative analysis of chronic pain definitions, and assess efficiency and quality of crowdsourced data compared to our data. We will also look at the possibility of providing data from crowdsourced participants to our expert panel. The crowdsourcing methods have the potential to enhance health-related care and research, by providing inexpensive, time-efficient methods to advance pain research and patient-centered care.

Conclusion

While the move to patient-centered care is to be applauded and represents a significant move towards shared decision making in health care, it poses significant challenges in terms of implementation. In the words of the traditional expression, “It is easier said than done”. On the one hand, the care patients receive should be clinically appropriate and clinically necessary. This will involve clinical judgments which for the most part, patients cannot make unassisted. So judgments by clinicians based on evidence, experience and clinical acumen will be necessary. But these judgments should be and hopefully will be increasingly tempered by patient input. The challenge is how to make sure the one does not contradict the other. Traditionally providers are supposed to know the difference between patients’ needs and wants, the perspicacity to know the difference, and the courage to tell the patients. Unnecessary and ineffective procedures should not be used simply because the patients want them. There is evidence that providers have pushed back against use of patient judgments about care such as CAHPS and claims that patient input can pressure providers to provide inappropriate care (e.g., demanding antibiotics for conditions not amenable to them).^{51, 52}

But in situations where alternative therapies are available and where the patient choice can play a significant role, patients certainly should be involved in the decision making. Back pain provides a very good exemplar where there are competing approaches that have roughly the same strength of evidence for their use but where patient preferences, costs, associated risks, and outcomes can be significantly different. One could make a case here that the patients’ preferences and financial burden should be considered criteria in the clinical decision making once safety and efficacy/effectiveness have been established.

Previous appropriateness studies determined appropriate care based only on efficacy, effectiveness and safety. This Center proposes to expand the definition of appropriateness to match not only the clinical needs of patients but also to take into account, patient preferences and resource limitations. Finally, it is important to determine whether a focus on the appropriateness of care actually improves patient-centered outcomes.

In this era of rising health care costs, it is increasingly urgent to evaluate the appropriateness of therapies provided to Americans. This is even more the case with CAM therapies, as these remain less well-researched, even as their utilization remains high.

References

1. Dictionary.com. Thesaurus.com, (2017, 2017).
2. Berwick DM and Hackbarth AD. Eliminating waste in US health care. *JAMA* 2012; 307: 1513–1516. [PubMed: 22419800]
3. Coulter ID. Comparative effectiveness research: does the emperor have clothes? *Alternative Therapies in Health & Medicine* 2011; 17: 8–15.
4. Brook RH. Appropriateness: the next frontier. *BMJ* 1994; 308: 218–219. [PubMed: 8111255]
5. Coulter ID and Adams AH. Consensus methods, clinical guidelines, and the RAND study of chiropractic. 1992.
6. Coulter ID. Evidence-based practice and appropriateness of care studies. *Journal of Evidence Based Dental Practice* 2001; 1: 222–226.

7. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *bmj* 2016; 353: i2089. [PubMed: 27365494]
8. Banta H, Behney C and Andrulid D. Assessing the efficacy and safety of medical technologies. Washington: Office of Technology Assessment 1978.
9. Michaud G, McGowan JL, van der Jagt R, et al. Are therapeutic decisions supported by evidence from health care research? *Archives of Internal Medicine* 1998; 158: 1665–1668. [PubMed: 9701101]
10. Imrie R and Ramey D. The evidence for evidence-based medicine. *Complementary therapies in Medicine* 2000; 8: 123–126. [PubMed: 10859606]
11. Congress U. Office of Technology Assessment. The impact of randomized clinical trials on health care policy and medical practice. Washington, DC: US Congress, Office of Technology Assessment 1983 OTA-BP-H-22.
12. Hicks NR. Some observations on attempts to measure appropriateness of care. *BMJ: British Medical Journal* 1994; 309: 730. [PubMed: 7950529]
13. Coulter ID, Lewith G, Khorsan R, et al. Research methodology: Choices, logistics, and challenges. *Evidence-Based Complementary and Alternative Medicine* 2014; 2014.
14. Chassin MR, Brook RH, Park RE, et al. Variations in the use of medical and surgical services by the Medicare population. Rand, 1986.
15. Wennberg JE and Gittelsohn AM. Small area variations in health care delivery. In: 1973, American Association for the Advancement of Science, Washington, DC.
16. Lewis CE. Variations in the incidence of surgery. *New England Journal of Medicine* 1969; 281: 880–884. [PubMed: 5812257]
17. Roos NP. Hysterectomy: variations in rates across small areas and across physicians' practices. *American Journal of Public Health* 1984; 74: 327–335. [PubMed: 6703159]
18. Wennberg JE. The paradox of appropriate care. *Jama* 1987; 258: 2568–2569. [PubMed: 3669227]
19. Chassin MR, Koseoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services?: A study of three procedures. *Jama* 1987; 258: 2533–2537. [PubMed: 3312655]
20. Lomas J Words without action? The production, dissemination, and impact of consensus recommendations. *Annual review of public health* 1991; 12: 41–65.
21. Coulter I, Elfenbaum P, Jain S, et al. SEaRCH™ expert panel process: streamlining the link between evidence and practice. *BMC research notes* 2016; 9: 16. [PubMed: 26744077]
22. Coulter ID, Shekelle PG, Mootz RD, et al. The Use of Expert Panel Results: The RAND Panel for Appropriateness of Manipulation and. *Chiropractic Technologies* 1999: 149.
23. Coulter I, Adams A and Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: a comparison of a multi-and single disciplinary panel. *Health services research* 1995; 30: 577. [PubMed: 7591782]
24. Andreasen B Consensus conferences in different countries. *International journal of technology assessment in health care* 1988; 4: 305–308. [PubMed: 10287625]
25. McClellan M and Brook RH. Appropriateness of care: A comparison of global and outcome methods to set standards. *Medical care* 1992; 30: 565–586. [PubMed: 1614228]
26. Park RE, Fink A, Brook RH, et al. Physician ratings of appropriate indications for six medical and surgical procedures. *American Journal of Public Health* 1986; 76: 766–772. [PubMed: 3521341]
27. Kahn KL, Park RE, Vennes J, et al. Assigning appropriateness ratings for diagnostic upper gastrointestinal endoscopy using two different approaches. *Medical care* 1992; 30: 1016–1028. [PubMed: 1434955]
28. Shekelle P and Adams A. The appropriateness of spinal manipulation for low back pain: Indication and ratings by a multidisciplinary expert panel. RAND, Santa Monica, CA Monograph No 1991.
29. Shekelle P, Adams A, Chassin M, et al. The Appropriateness of Spinal Manipulation for Low-Back Pain: Indications and Ratings by an all Chiropractic Expert Panel. Santa Monica, CA: The RAND Corporation 1992 R-4025/3-CCR/FCER.

30. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *New England Journal of Medicine* 1998; 338: 1888–1895. [PubMed: 9637810]
31. Coulter ID. Expert panels and evidence: The RAND alternative. *Journal of Evidence Based Dental Practice* 2001; 1: 142–148.
32. Organization WH. Appropriateness in health care services: report on a WHO workshop, Koblenz, Germany 23–25 3 2000 2000.
33. Nahin RL, Boineau R, Khalsa PS, et al. Evidence-based evaluation of complementary health approaches for pain management in the United States In: *Mayo Clinic Proceedings* 2016, pp.1292–1306. Elsevier.
34. Mick CA. Spine is #1. *Spine* 1973; 8.
35. Al N, Waddell G and Norland A. Epidemiology of Neck and Low Back Pain, in. *Neck and Back Pain: The scientific evidence of causes, diagnoses, and treatment*. Philadelphia: Lippincott Williams & Wilkins, 2000.
36. Maniadakis N and Gray A. The economic burden of back pain in the UK. *Pain* 2000; 84: 95–103. [PubMed: 10601677]
37. Deyo RA, Mirza SK and Martin BI. Back pain prevalence and visit rates: estimates from US national surveys, 2002. *Spine* 2006; 31: 2724–2727. [PubMed: 17077742]
38. Luo X, Pietrobbon R, Sun S, et al. PNP24: ESTIMATES AND PATTERNS OF HEALTHCARE EXPENDITURES AMONG INDIVIDUALS WITH BACK PAIN IN THE US. *Value in Health* 2003; 6: 280.
39. Ritzwoller DP, Crouse L, Shetterly S, et al. The association of comorbidities, utilization and costs for patients identified with low back pain. *BMC musculoskeletal disorders* 2006; 7: 72. [PubMed: 16982001]
40. Walsh NE, Brooks P, Hazes JM, et al. Standards of care for acute and chronic musculoskeletal pain: the Bone and Joint Decade (2000–2010). *Archives of physical medicine and rehabilitation* 2008; 89: 1830–1845. e1834. [PubMed: 18760171]
41. Shekelle PG, Coulter I, Hurwitz EL, et al. Congruence between decisions to initiate chiropractic spinal manipulation for low back pain and appropriateness criteria in North America. *Annals of internal medicine* 1998; 129: 9–17. [PubMed: 9653012]
42. Manipulation Coulter I. and mobilization of the cervical spine: the results of a literature survey and consensus panel. *Journal of musculoskeletal pain* 1996; 4: 113–124.
43. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *bmj* 2016; 353: i2016. [PubMed: 27353417]
44. Howe J Crowdsourcing: A definition. *Crowdsourcing: Tracking the rise of the amateur* 2006.
45. Hays RD, Liu H and Kapteyn A. Use of Internet panels to conduct surveys. *Behavior research methods* 2015; 47: 685–690. [PubMed: 26170052]
46. Weiner M The Potential of Crowdsourcing to Improve Patient-Centered Care. *The Patient-Patient-Centered Outcomes Research* 2014; 7: 123–127. [PubMed: 24671920]
47. Mason W and Suri S. Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior research methods* 2012; 44: 1–23. [PubMed: 21717266]
48. Azzam T and Jacobson MR. Finding a comparison group: Is online crowdsourcing a viable option? *American Journal of Evaluation* 2013; 34: 372–384.
49. Azzam T and Harman E. Crowdsourcing for quantifying transcripts: An exploratory study. *Evaluation and program planning* 2016; 54: 63–73. [PubMed: 26519690]
50. Goodman JK, Cryder CE and Cheema A. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making* 2013; 26: 213–224.
51. Price RA, Elliott MN, Cleary PD, et al. Should health care providers be accountable for patients’ care experiences? *Journal of general internal medicine* 2015; 30: 253–256. [PubMed: 25416601]

52. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Medical Care Research and Review* 2014; 71: 522–554. [PubMed: 25027409]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

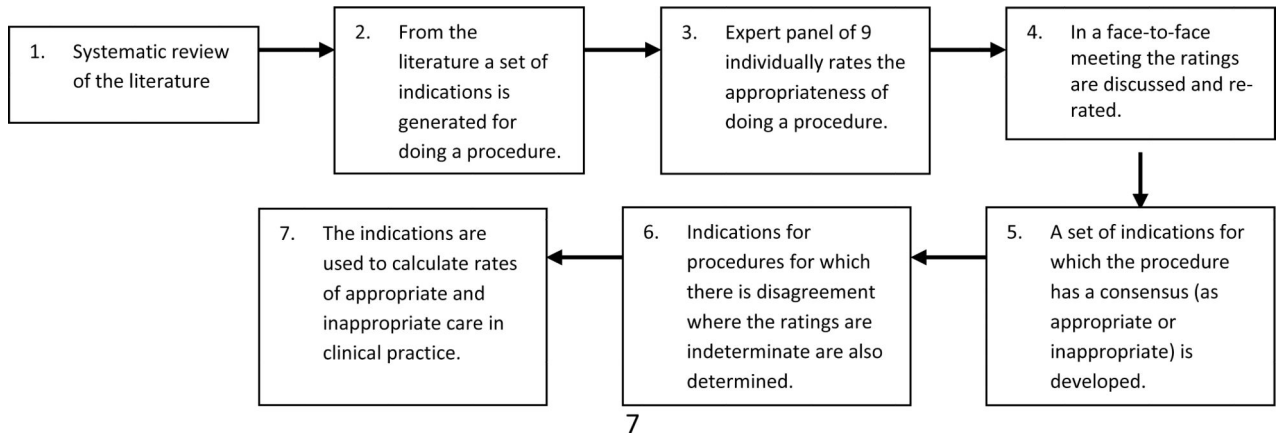


Figure 1:
The Process of the RAND Appropriateness Method