

Methods Research Report

Web Interactive Presentation of EPC Reports: A Foray Into Interactive Reports

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 290-2015-00002-I

Prepared by:

Brown University Evidence-based Practice Center
Providence, RI

Duke University Evidence-based Practice Center
Durham, NC

University of Minnesota Evidence-based Practice Center
Minneapolis, MN

Investigators:

Thomas A. Trikalinos, M.D.
Gaelen P. Adam, M.L.I.S.
Jens Jap, M.S.
Birol Senturk, B.S.
Stacey Springs, Ph.D.
Gillian Sanders-Schmidler, Ph.D.
Kathryn Lallinger, M.S.L.S.
Mary Butler, Ph.D.
Michelle Brasure, Ph.D.
Lionel Banez, M.D.
Elise Berliner, Ph.D.
Aysegul Gozu, M.D., M.P.H.
Ethan Balk, M.D.

**AHRQ Publication No. 19-EHC003-EF
September 2019**

Key Messages

- We developed an online interactive tool that allows users to obtain descriptive and analytic results for a systematic review comparing at least three treatments in various populations.
- In a limited evaluation, stakeholders were satisfied with the functionality of the tool and found it easy to use.
- We organized stakeholder feedback in three themes, namely expository, analytic, and information-sharing augmentations. We propose that a roadmap to realizing these augmentations involves the following: (i) integrating analytic capabilities, (ii) incorporating ways to assess the tradeoffs between several distinct benefits and harms of interventions with minimal assumptions, (iii) defining an information technology standard for systematic reviews, and (iv) developing a rigorous representation of evidence synthesis objects as mathematical graphs.

This report is based on research conducted by the Brown, Duke and Minnesota Evidence-based Practice Centers (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2015-00002-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EPC@ahrq.hhs.gov.

Suggested citation: Trikalinos TA, Adam GP, Jap J, Senturk B, Springs S, Sanders-Schmidler G, Lallinger K, Butler M, Brassure M, Banez L, Berliner E, Gozu A, Balk E. Web Interactive Presentation of EPC Reports: A Foray Into Interactive Reports. Methods Research Report. (Prepared by the Brown, Duke, and Minnesota Evidence-based Practice Centers under Contract No. 290-2015-00002-I.) AHRQ Publication No. 19-EHC003-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2019. Posted final reports are located on the Effective Health Care Program [search page](#).
<https://doi.org/10.23970/AHRQEPCMETHENGAGEINTERACTIVE>.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

If you have comments on this Methods Research Project they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Gopal Khana, M.B.A.
Director
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.
Director
Center for Evidence and Practice
Improvement
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Aysegul Gozu, M.D., M.P.H.
Elise Berliner, Ph.D.
Lionel Banez, M.D.
Task Order Officers
Center for Evidence and Practice
Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Jonathan Bae, M.D., C.P.P.S.

Associate Chief Medical Officer for Patient Safety and Clinical Quality
Assistant Professor of Medicine
Duke University Health System

George Cheely, M.D., M.B.A.

Program Director for Care Redesign
Assistant Professor of Medicine
Duke University Health System

Alison Weidner M.D.

Professor of Obstetrics and Gynecology, Urogynecology
Duke School of Medicine

Web Interactive Presentation of EPC Reports: A Foray Into Interactive Reports

Structured Abstract

Background. Systematic reviews are consumed by different types of users to address an ever-expanding set of questions and needs. It is unlikely that a single static report will efficiently satisfy the diverse needs of diverse users. It might be practical and satisfactory to give users the ability to generate their own report from the information produced by the systematic review.

Methods. We developed an open-source web-based interactive report presentation (IRP) of a systematic review on 14 nonsurgical treatments for Urinary Incontinence (UI) in women. We used data from a systematic review we conducted through the Agency for Healthcare Research and Quality Evidence-based Practice Center Program for the Patient Centered Outcomes Research Institute. We presented wireframes (mockups) of the IRP to a panel of three Duke Health System stakeholders in teleconferences, and incorporated their feedback in the final IR, which includes a visual representation of the evidence base, and allows users to access generic or detailed descriptive and analytic information through a point-and-click interface. Subsequently, the stakeholders piloted the IRP and evaluated its ease of use by answering open ended questions. We proposed a roadmap for scalable IRPs of systematic reviews starting from a thematic analysis of these suggestions.

Findings. The final tool (accessible at <https://to1.infalliblekitty.com/#index>) allows users to obtain descriptive and analytic results for a network of 14 treatment categories, various outcomes (cure, improvement, satisfaction, quality of life) and several subgroups (all women, older women, or those with stress or urge UI). Users can access descriptions overall or for specific sets of studies. The stakeholders were satisfied with the functionality of the tool, and commented on its ease of use. They proposed numerous augmentations, which we organized in themes of presentation (e.g., present information on numbers of trials on figures), analytic (e.g., allow on-the-fly subgroup analyses, explore tradeoffs between several outcomes) and information sharing (e.g., ability to import/export data from/to other software). A roadmap to satisfying these augmentations involves the following: (i) integrating analytic capabilities in the tool, (ii) incorporating ways to assess the tradeoffs between several distinct benefits and harms of interventions with minimal assumptions, (iii) defining an information technology standard for evidence synthesis objects (i.e., what information, in in what format, completely describes an evidence-base and its synthesis), and (iv) developing a rigorous representation of evidence synthesis objects as mathematical graphs that can be analyzed.

Conclusions: In sum, we described a foray into an alternative view of a systematic review that complements a static systematic review report. Health system representatives found it useful and practical, and requested analytical, expository, and information sharing augmentations to the tool.

Contents

Background	1
Interactive Report Presentations (IRPs).....	1
Methods	2
Interactive Presentation	2
Outcomes With Network Meta-analyses	2
Other Outcomes	2
Feedback From Health Care Systems	2
Stage 1: Tool Development	3
Stage 2: Pilot Implementation.....	3
Results	4
Interactive Report	4
Evaluation	9
Tool Development	9
Thematic Analysis of Feedback on Pilot Implementation.....	10
Discussion	12
References	15

Tables

Table 1. Selected feedback on how to improve interactive reports	11
--	----

Figures

Figure 1. Landing page of the interactive tool (screenshot from the tool)	4
Figure 2. Example: summary of the evidence base (screenshot from the tool)	5
Figure 3. Example: evidence map for satisfaction in women with urge UI (screenshot from the tool)	6
Figure 4. Example: odds ratios for satisfaction between all intervention categories in women with urge UI (screenshot from the tool)	7
Figure 5. Example: estimated rates for satisfaction between all intervention categories in women with urge UI (screenshot from the tool)	8
Figure 6. Example: study details for the studies that give results for satisfaction between all intervention categories in women with urge UI (not all columns are shown – screenshot from the tool)	9

Appendixes

Appendix A. Foray Into Computable Reports Presentation

Appendix B. Example Guiding Questions

Background

Interactive Report Presentations (IRPs)

Systematic reviews have emerged as a key tool for critically summarizing an evidence-base. Increasingly, systematic review reports are consumed by different types of users to address an ever-expanding set of questions and needs. Consider, for example, a recently-completed systematic review report on the comparative effectiveness of non-surgical treatments for Urinary Incontinence in women.¹ It could be used by specialist or generalist frontline providers to inform their treatment decisionmaking, health systems to decide which interventions to offer, professional societies to develop practice guidelines, clinical teachers to train health professionals, funders to shape their future research portfolio, patients who wish to participate in decisions about themselves, methodologists to test new evidence synthesis methods, and so on. Intuitively, it is unlikely that a single static report will efficiently satisfy the diverse needs of these diverse users.

While it makes sense for the Evidence-based Practice (EPC) Program to produce somewhat tailored reports (e.g., for primarily patient or for primarily clinical audiences), it would be impractical for the Program to generate products that are too-finely-tailored to the specific needs of specific types of users. By contrast, it might be practical and satisfactory to *give users the ability to generate their own report* from the information produced by the systematic review. This conjecture is motivated by observations from an EPC White Paper that summarized interviews with leaders of health systems regarding how they and their organizations use “evidence” (more precisely, evidence synthesis products).² Briefly, these related to the fact that EPC reports are voluminous and not easy to navigate. Users prefer evidence from systematic reviews to be summarized into short digestible summaries of the specific question with the option to click on a link for the more detailed information. In addition, users would like to see the results reported separately for specific subgroups and interventions or combinations of interventions.

This project addresses the above needs through an interactive evidence report that presents a user with a visual representation (“map”) of the evidence base and allows the user to access specific results based on intervention comparisons or by population subgroup to inform individual clinician’s decisions about the management of patients with urinary incontinence, as well as health system decisions related to improving patient safety and clinical quality and efficiency.

Methods

Interactive Presentation

This project took a current EPC report on nonsurgical treatments of Urinary Incontinence in adult women (hereafter, UI report)¹ and moved it to an online format, thereby making the information contained in the report interactive. The interactive format allows the user to select populations, interventions, and outcomes of interest and see specific study and summary information for included comparisons. We added no new information in the tool compared to the contents of the EPC report.

Outcomes With Network Meta-analyses

For outcomes that have been meta-analyzed, the tool has been developed to allow a user to obtain information about the direct and indirect evidence informing specific comparisons (e.g., number of trials and patients, characteristics of populations and compared interventions, risk of bias assessments), and the results of the analyses for different outcomes (direct comparison, combined direct and indirect comparisons, predicted mean outcomes with each compared treatment, subgroup analyses).

Other Outcomes

Outcomes that were not meta-analyzed, such as quality of life outcomes, were extracted from the AHRQ report into a data shell linked to the interactive report, analogously to what we did for the meta-analyzed outcomes.

Feedback From Health Care Systems

We evaluated the usefulness of the tool at the Duke Health System. This evaluation took place in two stages, which we describe below. We engaged the Associate Chief Medical Officer (CMO) for Patient Safety and Clinical Quality and the Program Director for Care Redesign at the Duke University Health System. Because we used the UI report as an example, we also engaged an experienced physician with expertise in Urogynecology who is also faculty at the Duke School of Medicine.

The Associate Chief CMO for Patient Safety and Clinical Quality is tasked with setting Duke's quality agenda and coordinating patient safety and clinical quality activities across the Duke Health System. He also leads the Quality Oversight Committee at Duke and the Health System's efforts to related to quality and safety. The Program Director for Care Redesign works within the CMO's office and partners with multidisciplinary teams of clinicians and administrators to identify, design, and execute improvements in patient care delivery targeting outcomes and efficiency improvements. They were selected because of their decision making roles within the Duke learning health system, the importance of evidence in their daily administrative and clinical roles, their ability to reflect on the evidence needs of the broader Duke health system. The physician stakeholder was selected because, as a senior faculty member, has been responsible for developing guidance for the management of patients with UI.

Stage 1: Tool Development

We created wireframes (mockups, non-interactive “sketches”) of the tool using hyperlinked PowerPoint slides, which were presented to a panel of three Duke Health System stakeholders in a teleconference (Appendix A). The goal of this stage was to understand the evidence needs of potential users and integrate their preferences into the tool. Semi-structured interviews of the same stakeholders informed key attributes of the tool and were used to optimize implementation (See Appendix B for a list of guiding questions). To the extent that it was feasible, we incorporated their suggestions into the design of the draft interactive tool.

Stage 2: Implementation of the Interactive Tool

The three stakeholders participated in piloting the tool. Using open-ended questions, we qualitatively assessed user perspectives on the usefulness of the tool toward informing decisions and achieving the objectives elucidated in earlier interactions. Specifically, we asked whether there are features the stakeholders found particularly useful, whether there are features that they would have liked to see, whether such a tool would meet their evidence needs, and whether they had any additional feedback.

Overall, the stakeholder engagement process provided feedback and guidance on tool development and informed on how EPC report analyses could be operationalized within health systems.

Results

Interactive Report

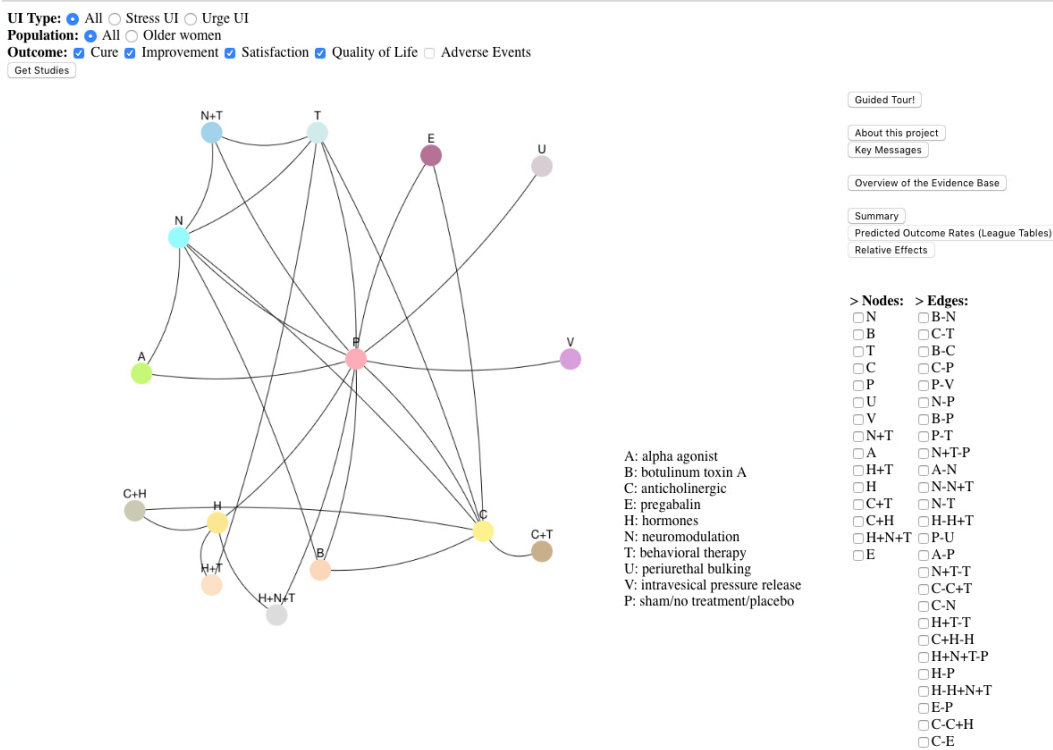
As of the time of this report, the prototype tool had been completed for the outcomes with network meta-analysis (<https://to1.infalliblekitty.com/#index>). The prototype tool is 508 compliant.

Landing Page

In the current version of the tool, the user is first presented with an interactive version of a figure with a representation of the evidence on the comparisons between all 15 intervention categories plus placebo (Figure 1). The evidence graph comprises nodes, which represent interventions, and edges (lines linking nodes). Edges connect a pair of nodes only if the corresponding interventions have been compared in at least one head-to-head study.

Figure 1. Landing page of the interactive tool (screenshot from the tool)

UI = urinary incontinence



Overview of the Evidence Base

Clicking on the “Overview of the Evidence Base” button on the right side of the screen provides the user with a summary description of the evidence graph (Figure 2). This information includes how many comparisons have been observed between interventions and intervention categories; whether there are groups of interventions that have been compared between them but not with other groups of interventions; how many and which studies have compared various interventions, a summary of key characteristics of the underlying studies, including number of

arms, sample size and risk of bias scores; and for interventions that have not been compared head to head, which are the comparisons that most contribute to the indirect estimate of the effect. This information can also be generated for specific population subgroups and outcomes, as well as for specific nodes or comparisons.

Figure 2. Example: summary of the evidence base (screenshot from the tool)

Outcomes: Cure, Improvement, Satisfaction, Quality of Life

UI Type: All

Population: All

Total Evidence Base

Characteristic	Description	Value
Trials, number (people)	Total number of trials (people) in the evidence-base	213 (24951)
Publication year, median (range)	A proxy for the years patients were enrolled in the study	2004.5 (1990, 2017)
Sample size, median (range)	Sample size per study	90 (18, 2758)

Comparisons

Characteristic	Description	Value
All possible comparisons, number	Number of possible comparisons across specific interventions	105
Comparisons with data, number (% of all)	Pairs of intervention categories compared in at least one trial	26 (26/105 = 24.8%)
Trials, median (range)	Median of number of trials informing each comparison with data	3 (1, 39)
Cumulative sample size, median (range)	Median sample size of trials per comparison with data	458 (66, 6848)

Intervention Categories

Characteristic	Description	Value
Intervention Categories, number	Number of intervention categories	15
Trials, median (range)	Median of number of trials per intervention category	5 (1, 39)
Sample size, median (range)	Median sample size per intervention category	885 (66, 7252)

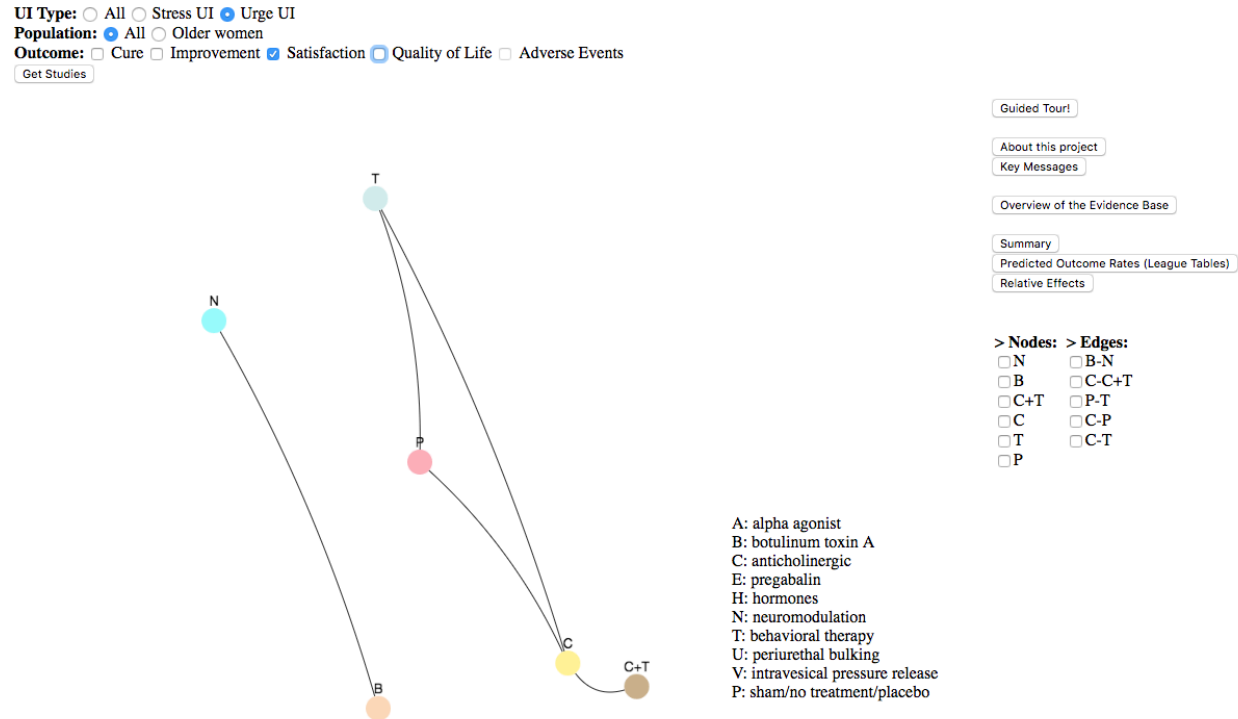
UI = urinary incontinence

Information on Specific Outcomes

For users interested in a specific outcome (e.g., satisfaction) or a specific population (e.g., urge UI), an analogous evidence graph can be presented, depicting only the interventions with data for this outcome (Figure 3).

Figure 3. Example: evidence map for satisfaction in women with urge UI (screenshot from the tool)

UI = urinary incontinence



The user can then request information about (1) the relative differences between any two interventions, or (2) the predicted mean frequency of the outcome for an intervention (a metric that can be looked at across interventions to rank their effectiveness). The user may opt to receive information about some comparisons, or about all possible comparisons. In its current form, the tool does not provide or allow for additional analyses not included in the original report.

As an example of looking at the relative differences between any interventions, Figure 4 describes odds ratio estimates and 95% CIs for all pairwise comparisons in the graph in Figure 3. Each cell has a (row, column) address, and reports the estimated odds ratio between the intervention in the row versus the intervention in the column. The unshaded cells correspond to comparisons for which there is head-to-head information (i.e., there is an edge between these corresponding nodes in the evidence graph). The estimated treatment effects in these cells are informed by direct and indirect evidence. The shaded cells correspond to comparisons that have not been empirically observed (there is no edge between these corresponding nodes in the evidence graph), and the odds ratios are based only on indirect comparisons.

Figure 4. Example: odds ratios for satisfaction between all intervention categories in women with urge UI (screenshot from the tool)

Table H-8A. Odds ratios for SATISFACTION between all intervention categories: URGENCY UI (subgraph 1)

C	C	0.8 (0.26, 2.5)	0.32 (0.07, 1.41)	2.6 (0.57, 11.87)
C+T	1.25 (0.4, 3.88)	C+T	0.39 (0.06, 2.48)	3.24 (0.5, 20.75)
T	3.16 (0.71, 14.09)	2.53 (0.4, 15.93)	T	8.2 (1.7, 39.43) ^a
P	0.39 (0.08, 1.76)	0.31 (0.05, 1.98)	0.12 (0.03, 0.59) ^a	P

See Figures above for code. Results are given as odds ratios (95% confidence intervals). Values above the diagonal >1 favor the row intervention (to the left) over the column intervention (below).

Table H-8B. Odds ratios for SATISFACTION between all intervention categories: URGENCY UI (subgraph 2)

B	B	1.4 (0.93, 2.12)
N	0.71 (0.47, 1.08)	N

See Figures above for code. Results are given as odds ratios (95% confidence intervals). Values above the diagonal >1 favor the row intervention (to the left) over the column intervention (below).

UI = urinary incontinence, C = anticholinergic, T = behavioral therapy, P = sham/no treatment/placebo, B = botulinum toxin A, N = neuromodulation; cells shaded gray are based only on indirect comparisons.

As an example of evaluating the predicted mean frequency of the outcome for an intervention, the user may receive the expected (mean) satisfaction frequency in each intervention category (Figure 5), which we call the league table. The table in this figure describes the mean proportion of patients with the outcome of interest (satisfaction) for each intervention category over the populations included in the meta-analysis and the corresponding forecasted proportion in a new setting that is analogous to the settings of the analyzed studies. The results in the Figures 4 and 5 are from the same analysis. The league table explains what the relative effects imply about the probability of the outcome under each treatment. The expected frequency of the event in a setting that is analogous to the settings in which the meta-analyzed studies were conducted is shown in the forecast column. Note that the confidence intervals for the forecast are always larger than the confidence intervals for the mean.

Imagine that you are hiking along a trail from east to west, through six camp sites. The camp sites serve as the analogue for the interventions. A table showing the signed distances (which encode the direction of movement and the distance traveled between pairs of campsites) would be the analogue of the relative effects table. A table showing how far each campsite is from the easternmost end of the trail would be the analogue of the league table.

Figure 5. Example: estimated rates for satisfaction between all intervention categories in women with urge UI (screenshot from the tool)

BTX= onabotulinum toxin a; CI = confidence interval

Table H-14B. Estimated and forecast rates of SATISFACTION by intervention category: URGENCY UI

Urgency UI		
Intervention category	Mean Percent* (95% CI)	Forecast Percent (95% CI)
<i>Pharmacological</i>		
BTX (B)	59.5 (52.3, 66.2)	59.5 (52.3, 66.2)
Anticholinergic (C)	48.9 (16.3, 82.5)	48.9 (4.1, 95.6)
<i>Nonpharmacological</i>		
Neuromodulation (N)	51.1 (43.8, 58.5)	51.1 (43.8, 58.5)
Behavioral Therapy (T)	75.1 (29.1, 95.7)	75.1 (9.7, 98.8)
<i>Combination</i>		
Anticholinergic + Behavioral Therapy (C+T)	54.4 (17.9, 86.7)	54.4 (4.7, 96.6)
<i>No Treatment</i>		
Placebo/Sham/No Treatment (P)	26.9 (4.7, 73.4)	26.9 (1.3, 91.3)

*The summary mean percentage (with confidence interval) of women in the trials receiving the intervention with the outcome.

UI = urinary incontinence, BTX = botulinum toxin A, CI = confidence interval

Information on Specific Studies

Clicking on the “Get Studies” button allows the user to see descriptions of the studies included in this analysis, their results and key characteristics, including number of participants, specific interventions, between arm odds ratios, and nine risk of bias indicators (Figure 6). This helps the user understand where (most) of the information that is encoded in the estimate comes from.

Figure 6. Example: study details for the studies that give results for satisfaction between all intervention categories in women with urge UI (not all columns are shown – screenshot from the tool)

Column visibility Search:

PMID ▲	Study Author	Year	Age	Older_women	UI type	Followup	Counts	Coarse trt	Granular trt	trt description1
9863850	Burgio	1998	NA	1	uui	NA	45/65 vs 14/65	H vs K	H4+H6 vs K1	pfmt & biofeedback
9863850	Burgio	1998	NA	1	uui	NA	27/67 vs 14/65	A vs K	A1 vs K1	oxybutynin
9863850	Burgio	1998	NA	1	uui	NA	27/67 vs 45/65	A vs H	A1 vs H4+H6	oxybutynin
18678843	Burgio	2008	56.9	0	uui	NA	31/153 vs 51/154	A vs A+H	A3 vs A3+H1+H4	tolterodine
20639023	Burgio	2010	58.4	0	uui	NA	28/32 vs 25/32	A vs A+H	A1 vs A1+H4	oxybutynin
27701661	Amundsen	2016	63	0	uui	6	113/190 vs 89/174	B vs G	B1 vs G2	botox

Showing 1 to 6 of 6 entries Previous Next

NA = not applicable, pfmt = pelvic floor muscle training, Botox = Onabotulinum toxin A. uui: urge urinary incontinence. Counts: number of people with the event out of total per arm. Coarse trt: coding that coarsely categorizes interventions in treatment categories. Followup: followup time. granular trt: Coding for specific interventions. The key for coarse and granular codes is provided in the UI report and in the graphs in the interactive report tool.

Other Outcomes

For all outcomes, clicking on the “Summary” button brings the user to a list of key findings for that outcome. These are narrative summaries of the evidence base with strength of evidence assessments. For outcomes that are not meta-analyzed, such as quality of life and adverse events, the information available are the narrative summaries and detailed lists of the studies; but no summary tables.

Evaluation

Tool Development

The three stakeholders from the Duke University Health System were scheduled for one-hour conference calls with the EPC and AHRQ members of our working group. Approximately 30 minutes of each call was devoted to the web interactive tool. During the call, they were told about the rationale for the interactive systematic review tool and were shown a mock-up of the tool for the sample urinary incontinence report. After the demonstration, the health system these stakeholders were asked to comment on the novelty of the interactive tool, the relevance of the tool to their evidence needs, and for any suggestions on making the interactive evidence tool more effective or more efficient.

All of the stakeholders liked the interactive tool and agreed with the goal of allowing users to explore deeper when they need specific details on treatments or outcomes. They suggested providing a preface screen that would give a general summary of the overarching comparisons and conclusions. The stakeholders mentioned that the addition of cost effectiveness would be helpful in their decisionmaking in terms of the outcomes of interest. It was explained that not all EPC reports evaluate the impact of interventions on costs and/or cost effectiveness but that they would be included in the map if relevant to individual case reports. One concern was that

information for different outcomes is not provided side by side, but only for one outcome at a time. Finding ways to allow users to quickly get to the supporting evidence for comparisons and outcomes of interest should be explored.

Thematic Analysis of Feedback on the Implementation of the Interactive Tool

A link to the interactive tool, along with a tutorial giving an overview of how to access various types of information through the tool, was sent to the same stakeholders. In general, the stakeholders found that the tool was helpful and intuitive, allowing them to explore deeper when they needed specific details on treatments or outcomes.

Their suggestions for improvement included *analytical, expository, and information sharing* augmentations. We did not implement these augmentations in the prototype tool. We propose that future iterations of interactive tools should include as many of these augmentations as practical. Key suggestions, edited for uniformity of language and brevity, are listed in Table 1.

Analytical Augmentations

We identified two themes pertaining to analytical augmentations. The first theme was the desire to have *user-specified subsets* of populations, interventions, outcomes, or other ways to define subgroups of studies, and obtain analogous descriptions of the evidence base, the individual studies, and their quantitative summaries. This need can be served by an interactive system that has the ability to perform analyses on demand.

The second theme pertained to the *desire to consider information on diverse outcomes at the same time*. For decisions that are influenced by two or more outcomes, it is desirable to have the ability to describe their syntheses in parallel. Further, users asked for a way to examine tradeoffs between competing outcomes, or more generally, to explore implications of multiple outcomes on the preferability of each treatment considering their own preferences. This need can be served by providing a framework for exploring decisions affected by multiple attributes, taking into account the user's preferences for each outcome while minimizing assumptions about the valuations and aggregations of the valuations of various outcomes.

Expository Augmentations

Most of the requested expository augmentations pertained to the tool itself. These included the need for (i) more informative exhibits, including but not limited to, figures that show the number of studies, their design (e.g., randomization unit, if applicable, number of arms, parallel or crossover), their sample size, the intervention intensity (e.g., drug dose), the followup at which outcomes were measured, and the risk of bias, (ii) dynamic exhibits that show the impact of assumptions about evidence structures on the evidence base (e.g., how the evidence graph changes with alternative definitions of interventions), and (iii) inserting hyperlinks to cross reference information within the tool, and from the tool to external resources (e.g., the corresponding evidence report, PubMed, ClinicalTrials.gov). All of these tasks can be served by appropriate modeling of the evidence network with mathematical graphs.

Information Sharing Augmentations

Finally, the tool should be able to provide salient information about the evidence synthesis programmatically to other software clients. The salient information about the evidence synthesis

can be served by specifying (i) an evidence synthesis object that is documented, and (ii) a documented programmatic communication interface.

Table 1. Selected feedback on how to improve interactive reports

Suggestion	Thematic classification
Provide ability to conduct customized subgroup analyses	Analytic augmentation (dynamic analyses)
Show information on two or more outcomes in parallel	Analytic augmentation (show diverse outcomes)
Include key messages on main page.	Expository augmentation (more-informative exhibits)
When selecting two interventions, give the user the ability to visually represent effect size.	Expository augmentation (more-informative exhibits)
Provide additional information on individual studies: study design, author conflicts, study quality, timeframe for follow-up, inclusion/exclusion criteria.	Expository augmentation (more-informative exhibits)
Color coding/sorting of evidence grade of study/outcome.	Expository augmentation (dynamic exhibits)
Give a way to scan head-to-head comparisons for high-quality evidence, to help the user know where to dive deeper.	Expository augmentation (dynamic exhibits)
Give the ability to export/print tailored evidence tables.	Expository augmentation (dynamic exhibits) / Information sharing augmentation (human-readable)
Insert a hyperlink to PubMed for all studies.	Expository augmentation (cross-linking out of the report)
Import information from other software systems (e.g. Systematic Review Data Repository)	Information sharing augmentation (machine readable)
Export information for use with other software systems including local systems	Information sharing augmentation (machine readable)

Discussion

Stakeholder-driven systematic reviews, such as those conducted by the EPC program, summarize the practically available empirical information on important questions and, in this sense, have the capacity to address essential key evidentiary needs of their consumers. First, and most importantly, they describe the status of the evidence and the key inferences that can be drawn about the effectiveness and safety of interventions. Secondly, and of particular importance to many users, EPC systematic review reports discuss the implications of the evidence they describe and identify any obvious and universally-applicable solutions to decisional problems. However, systematic review reports, including those by the EPC Program do not provide specific recommendations for practice, especially for value-laden decisions or when there is pervasive uncertainty. For such problems, explicit recommendations are outside the scope of systematic reviews and are not satisfactorily addressed by their methods. Those developing them, however, benefit from a meaningful synthesis of the pertinent evidence. For this reason, in the current project, we focused on improvements to the primary function of systematic reviews, namely, the description of an evidence base and of the key inferences that can be drawn from empirical data.

Starting from our own conception of an interactive presentation of an EPC report and the feedback we received, we envision a rich set of use cases for interactive reports. For the UI case, health systems that want to optimize their patient management approach may be interested in the expected outcomes of different treatments for patient subgroups they serve. Through the interactive tool, they have immediate access to the outcomes that have been examined in various subgroup strata. They can disaggregate information for outcomes that are reported at a coarse level in the systematic review, such as different definitions of cure or specific adverse events (e.g., by focusing on the corresponding part of the evidence base through a simple interface). Further, they can immediately identify outcomes that are not reviewed (e.g., legal implications, cost-effectiveness of interventions), or outcomes that were reviewed but for which data are sparse or nonexistent (e.g., costs, mortality). While for nonreviewed or sparse outcomes, they would have to obtain additional information, they can very efficiently identify the evidence gaps that are relevant to them.

Based on the feedback we received, we propose a roadmap for research that leads to scalable interactive systematic review reports. First, there is a need to perform analyses, including sensitivity analyses, dynamically. This can be accomplished by embedding statistical analyses in the interactive report. Robust open-source meta-analysis-specific and general statistical modeling solutions exist, and they can be readily used for such a scheme. For example, the analyses for the UI project use the code base of the AHRQ-funded Open-Meta-analyst software and various other R libraries, all of which are open-source.³⁻⁶ Enabling on-demand analyses is obviously useful, but not without risk. The most important risk is that users might undertake analyses that are clinically not meaningful and methodologically ungrounded, perhaps because they have not studied the evidence-base closely and are not familiar with evidence synthesis methods. Approaches to managing this risk (e.g., by characterizing any on-demand analysis as “descriptive”) should be considered. However, any publicly available curated dataset can be analyzed badly, and any programming language and statistical package can be abused. We submit that, perhaps, the advantages of offering dynamic analyses outweigh their risks.

Second, most decisionmaking problems require information for several outcomes at the same time. An interactive report may present this information and also facilitate explorations of what the evidence means for various choices, while minimizing assumptions about how users value

and balance the outcomes they want to consider. Several approaches to multicriteria decisionmaking are generally relevant.⁷ For interactive reports, we favor approaches that do not demand that users explicitly value and aggregate the various outcomes they consider. For example, stochastic multi objective acceptability analysis (SMAA) does not require that users express all outcomes into a common scale (e.g., as quality-adjusted life years) or even that they explicitly state how much they value one outcome versus another (e.g., for UI, that one cares x -times more about improving their quality of life outcomes compared to being cured). SMAA allows users to specify their preferences loosely, in the form of directional preferences (for UI, it would encode that someone cares, perhaps, more about bettering their quality of life than about being fully cured, but not exactly how much more; and so on), and can identify treatments that are most compatible with these preferences.^{8,9} SMAA analyses are practical in the context of a systematic review,^{10, 11} and we have developed numerical algorithms to make them computationally efficient.¹² We envision that an instantiation of the tool would, optionally, provide some functionality such as SMAA, to help users gauge the implications of the reviewed evidence for decisions and policy.

Third, it is important that any interactive system efficiently communicates with other existing systems developed by other participants of a broader evidence ecosystem.¹³ The, currently under development, new version of the Systematic Review Data Repository (SDRD 2.0)¹⁴ aims to develop a standard for systematic review objects that would be comprehensive, scalable, and well documented, so that they can be seamlessly used and integrated into practice such as through guidelines and clinical decision support systems. This effort involves the engagement of a group of users, methodologists, information technologists, software developers, and other experts.

The final need is technical, and pertains to developing a unified way to represent, manipulate, describe, meta-analyze (when applicable), present, query (search), and share evidence synthesis objects. This is necessary for the routine production of interactive reports of systematic reviews. An explicit description of this framework is beyond the scope of this document. At a high level, we propose to start with a rigorous representation of studies, their attributes, and the evidence base they comprise as mathematical graphs (i.e., mathematical objects analogous to those that encode the images of the evidence-base in Figures 1 and 3). Mathematical graphs are a very versatile modeling tool,¹⁵ and they are used to describe, analyze, and optimize many problems that can be represented as networks, from telecommunications to social networks,¹⁶ and from integrated circuits to ecology,¹⁷ epidemiology,¹⁸ and evidence bases.¹⁹ For example, any standard for evidence synthesis data can be encoded in a mathematical graph. A proper representation would, at minimum, capture (i) the hierarchical organization of data into arms and studies, or of outcomes reported at various time points, (ii) alternative structures for the evidence base (e.g., allow grouping of treatments into various categories), and (iii) attributes at the arm (e.g., demographics, baseline descriptives), study (e.g., design characteristics), or other level (e.g., risk of bias per outcome, and in multi-arm trials, per outcome and comparison). Then the various tasks of an evidence synthesis, including descriptive, qualitative, and quantitative (if necessary) analyses can be represented as graph operations, and decision analyses are represented as optimization problems over graphs. Graph operations and graph optimization have been extensively studied in mathematics, engineering, and the social sciences. We believe that gains can be had by porting powerful methods to evidence synthesis problems.

Finally, this work is preliminary and does not directly inform on the generalizability of the findings to users other the stakeholders we engaged. This is almost universally the case with

qualitative studies. We are also not able to examine how likely it is that users from different types of health systems, e.g., health systems that are smaller, not affiliated with tertiary care facilities school, or covering primarily in rural areas, would find the tool useful, because we did not succeed in securing feedback from the ones we contacted. We hypothesize that users who seek a description of the evidence (i.e., a description of what is known) will find this tool to be on the right track. However, many users seek clinical practice recommendations, especially for controversial topics or topics with a paucity of evidence (i.e., algorithms for what to do based on what is known). No EPC report nor any tool that summarizes evidence would satisfy this need, because EPC reports critically describe and summarize the evidence base, but generally refrain from making explicit clinical practice recommendations. More generally, evidence informs, but does not fully determine, clinical and policy decisions.^{20, 21} Decisions are based on the evidence but also take into account the decisionmaker's preferences, must weigh numerous inputs, including information on various endpoints, the decisionmakers' preferences and attitudes towards risk, and resource constraints. Thus, at best a tool that describes evidence would provide prospective users with a means of exploring the implications of alternative actions, as we alluded above in describing the need to evaluate the balance of benefits and harms by considering information on several outcomes jointly.

In terms of practical implementation, we believe that the contents of any EPC systematic review, including technology assessments, can be presented in an interactive format. For example, the part of the tool that describes which treatments have been compared between them and for which populations and outcomes (the interactive evidence graph) can be generated in a straightforward manner for any systematic review. The same is true for study-level descriptions. Systematic reviews that are limited to pairwise comparisons and do not do a network meta-analysis would only have direct comparisons to report. If a network meta-analysis is not undertaken, however, the review cannot explicitly rank treatments and some of the exhibits we developed (e.g., analogous to Figures 4 and 5) would be omitted. We estimate that the development of the tool required at least 50 hours of programmers' time. However, this time investment can be substantially sped up if a dedicated system is built, and as long as the underlying data are available electronically in the correct format. We believe that, if a standard for a systematic review object is used, the generation of the interactive report can be fully automated.

In sum, we described a foray into an alternative view of a systematic review that complements a static systematic review report. Health system representatives found it useful and practical, and requested analytical, expository, and information sharing augmentations to the tool. We discussed a roadmap for practical and theoretical developments for scaling the production of informative interactive reports.

References

1. Balk E. Protocol for Nonsurgical Treatments for Urinary Incontinence in Adult Women: A Systematic Review Update. Rockville, MD: Agency for Healthcare Research and Quality; 2017. <https://effectivehealthcare.ahrq.gov/topics/urinary-incontinence-update/research-protocol>.
2. White C, Sanders GS, Butler M, et al. Understanding Health-Systems' Use of and Need for Evidence To Inform Decisionmaking. Rockville, MD: Agency for Healthcare Research and Quality; 2017. <https://effectivehealthcare.ahrq.gov/topics/health-systems/research>.
3. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems*. 2006;1695(5):1-9.
4. van Valkenhoef G, Lu G, de Brock B, et al. Automating network meta-analysis. *Res Synth Methods*. 2012;3(4):285-99.
5. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3).
6. Wallace BC, Schmid CH, Lau J, et al. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Med Res Methodol*. 2009 Dec 4;9:80. doi: 10.1186/1471-2288-9-80. PMID: 19961608.
7. Greco S, Figueira J, Ehr Gott M. *Multiple criteria decision analysis*: Springer; 2016.
8. Lahdelma R, Hokkanen J, Salminen P. SMAA-stochastic multiobjective acceptability analysis. *Eur J Oper Res*. 1998;106(1):137-43.
9. Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. *Eur J Oper Res*. 2007;178(2):500-13.
10. van Valkenhoef G, Tervonen T, Zhao J, et al. Multicriteria benefit-risk assessment using network meta-analysis. *J Clin Epidemiol*. 2012;65(4):394-403.
11. Van Valkenhoef G, Tervonen T, Zwinkels T, et al. ADDIS: a decision support system for evidence-based medicine. *Decis Support Syst*. 2013;55(2):459-75.
12. Trikalinos T, van Valkenhoef G. Efficient sampling from uniform density n-polytopes. 2014. <https://www.brown.edu/academics/public-health/research/evidence-synthesis-in-health/sites/brown.edu/academics/public-health/research/evidence-based-medicine/files/uploads/EfficientsamplingSMAA.pdf>
13. Agoritsas T. Harnessing the Potential of a Digital Evidence Ecosystem for Shared Decision Making. *BMJ*. 2015;350:g7624.
14. Ip S, Hadar N, Keefe S, et al. A Web-based archive of systematic review data. *Syst Rev*. 2012;1(1):15.
15. Clark J, Holton DA. *A first look at graph theory*: World Scientific Publishing Company; 1991.
16. Carrington PJ, Scott J, Wasserman S. *Models and methods in social network analysis*: Cambridge university press; 2005.
17. Fortin M-J, James PM, MacKenzie A, et al. Spatial statistics, spatial regression, and graph theory in ecology. *Spat Stat*. 2012;1:100-9.
18. Richardson TS, Robins JM. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper. 2013;128(30):2013.
19. Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods*. 2012;3(4):312-24.
20. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 1: introduction. *Med Decis Making*. 2013 Jul;33(5):597-606. doi: 10.1177/0272989X13487604. PMID: 23804506.

21. Trikalinos TA, Sanders-Schmidler G, Russell L. Evidence Synthesis. In: Neumann P, Sanders G, Russell L, Siegel J, Ganiats T, eds. Cost-effectiveness in health and medicine. Oxford University Press; 2016.