

RESEARCH ARTICLE

Open Access

Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada

Michèle Tony¹, Monika Wagner¹, Hanane Khoury¹, Donna Rindress¹, Tina Papastavros², Paul Oh³ and Mireille M Goetghebeur^{1,4*}

Abstract

Background: Consistent healthcare decisionmaking requires systematic consideration of decision criteria and evidence available to inform them. This can be tackled by combining multicriteria decision analysis (MCDA) and Health Technology Assessment (HTA). The objective of this study was to field-test a decision support framework (EVIDEM), explore its utility to a drug advisory committee and test its reliability over time.

Methods: Tramadol for chronic non-cancer pain was selected by the health plan as a case study relevant to their context. Based on extensive literature review, a by-criterion HTA report was developed to provide synthesized evidence for each criterion of the framework (14 criteria for the MCDA Core Model and 6 qualitative criteria for the Contextual Tool). During workshop sessions, committee members tested the framework in three steps by assigning: 1) weights to each criterion of the MCDA Core Model representing individual perspective; 2) scores for tramadol for each criterion of the MCDA Core Model using synthesized data; and 3) qualitative impacts of criteria of the Contextual Tool on the appraisal. Utility and reliability of the approach were explored through discussion, survey and test-retest. Agreement between test and retest data was analyzed by calculating intra-rater correlation coefficients (ICCs) for weights, scores and MCDA value estimates.

Results: The framework was found useful by the drug advisory committee in supporting systematic consideration of a broad range of criteria to promote a consistent approach to appraising healthcare interventions. Directly integrated in the framework as a "by-criterion" HTA report, synthesized evidence for each criterion facilitated its consideration, although this was sometimes limited by lack of relevant data. Test-retest analysis showed fair to good consistency of weights, scores and MCDA value estimates at the individual level (ICC ranging from 0.676 to 0.698), thus lending some support for the reliability of the approach. Overall, committee members endorsed the inclusion of most framework criteria and revealed important areas of discussion, clarification and adaptation of the framework to the needs of the committee.

Conclusions: By promoting systematic consideration of all decision criteria and the underlying evidence, the framework allows a consistent approach to appraising healthcare interventions. Further testing and validation are needed to advance MCDA approaches in healthcare decisionmaking.

* Correspondence: mireille_goetghebeur@biomedcom.org

¹BioMedCom Consultants inc, Montréal, Québec Canada

Full list of author information is available at the end of the article

Background

Making decisions about the appropriate allocation of scarce healthcare resources is a necessary but difficult task. It involves consideration of a number of decision criteria, processing disparate streams of information and balancing individual and group/jurisdictional perspectives, not to mention ethical principles [1]. This complex process demands transparency, consistency, and accountability to be perceived as legitimate by the public and healthcare providers and to increase the likelihood of making good decisions [2,3].

Cost-effectiveness analysis (CEA), an economic method that aims to maximize efficiency, is the paradigm that currently dominates many healthcare policy decisionmaking processes. However, while potentially useful as a measure for productivity in healthcare, [4] sole reliance on CEA fails to address broader societal and political issues, such as disease severity, availability of alternatives, equity, and budget impact [4,5]. Even agencies that espouse the cost-effectiveness paradigm, such as NICE, acknowledge that other factors are being considered in their decisions; [5,6] however, these factors are not consistently integrated into the decisionmaking process and not revealed in a transparent fashion [5]. There are additional concerns surrounding the CEA paradigm, particularly the approach centered on the cost per quality-adjusted life-year (QALY), including methodological difficulties in valuing health states and fundamental ethical questions regarding the underlying objectives and outcomes of the approach [4,5]. Further, the complexity of some economic models can hamper understanding by the public and even by some decisionmakers of the key issues involved in the decision [6]. There is a need for a process that systematically and explicitly addresses all key factors impacting decisions, while promoting transparency and consistency in decisionmaking.

Multicriteria decision analysis (MCDA), an emerging tool in healthcare decisionmaking, goes beyond CEA by allowing systematic and explicit consideration of multiple factors that may impact the decision [1,7-10]. MCDA structures complex problems into a comprehensive set of criteria. Each criterion is weighted –a step that allows decisionmakers to clarify their fundamental objectives and perspectives [11]– and the performance of each healthcare intervention with respect to each criterion is scored allowing identification of weaknesses and strengths [1,7-9]. Although MCDA may be perceived as not intuitive and potentially usurping decisionmaking authority, if kept simple, it facilitates an important dialog and forces decisionmakers to think hard about and clearly express what they value, why they value it, and in what context they value it.

In addition to the emergence of MCDA, recent decades have seen growth in the field of Health Technology Assessment (HTA) [12]. With a mission “to assist and advise healthcare decisionmakers in defining health policies at all levels”, [13] HTA acts as a bridge between evidence and decisionmaking to ensure better synthesis, communication and dissemination of information [14]. It is now widely recognized that to fulfill its mission, in addition to clinical and economic factors, HTA needs to address social, organizational, ethical and legal dimensions of health technology [15-18]. Current efforts to develop international standards for HTA reports [19,20] point to the need for a structured format that can provide full access to the underlying evidence, thereby enhancing transparency and usability of the report to decisionmakers and stakeholders.

MCDA, HTA, and knowledge translation have a common objective: enlightened and evidence-based healthcare decisionmaking. Reflection on the drivers behind healthcare decisions is essential to ensure that these fields of research are aligned with the practical needs of decisionmakers. A pragmatic framework (EVIDEM), proposes a standard set of criteria with detailed methodology i) to provide validated synthesized evidence for each criterion (by-criterion HTA report), and ii) to systematically consider each criterion using a MCDA Core Model and a Contextual Tool [21]. Criteria were defined based on an extensive analysis of the literature and decision processes around the world, as well as discussion with stakeholders; tools were developed to stimulate reflection on priorities, to support systematic deliberation and to facilitate pragmatic knowledge transfer [21,22].

The EVIDEM framework underwent proof-of-concept evaluation by a panel composed of a broad range of Canadian stakeholders appraising 10 medications [23]. As a support for policy decisionmaking, the framework was field-tested with the reimbursement advisory committee of a private health plan in South Africa using a cervical cancer screening tool as a case study [24]. For clinical decisionmaking, the framework was tested by a panel of pediatric endocrinologists and other Canadian stakeholders using growth hormone for Turner syndrome as a complex case study with far reaching ethical issues; this study led to further development of the framework and explicit integration of ethical and system-related criteria into the decision process [22]. The objective of this study was to field-test a MCDA-based framework (EVIDEM), explore its utility to a drug advisory committee and test the stability of estimates over time using tramadol for chronic non-cancer pain (CNCP) as a case study relevant to their context.

Methods

Study design

In a move to build on existing decisionmaking processes, the EVIDEM framework was field-tested by the Drug Advisory Committee of the Ontario Workplace Safety Insurance Board (WSIB), which provides benefits, including healthcare benefits, to workers suffering injury or illness directly related to work. Tramadol for CNCP was selected by the committee as a relevant case study to the context of the population covered by the WSIB. The study design is presented in Figure 1. Based on an extensive literature review, a structured HTA report on tramadol for CNCP was produced, tailored to provide detailed information, as available, for each criterion of the EVIDEM framework, i.e., 14 criteria for the MCDA Core Model and 6 qualitative criteria for the Contextual

Tool [21,22]. Synthesized data integrated in the MCDA Core Model and the Contextual Tool is referred to as the “by-criterion HTA report”. During workshop sessions, committee members tested the framework in three steps by assigning: 1) weights to each criterion of the MCDA Core Model representing individual perspective; 2) scores for tramadol for each criterion of the MCDA Core Model using synthesized data (by-criterion HTA report); and 3) qualitative impacts of the Contextual Tool criteria on the appraisal. Utility and reliability of the approach were explored through discussion, survey and test-retest.

By-criterion Health Technology Assessment report

An extensive analysis of the published and grey literature was performed to identify relevant data for each

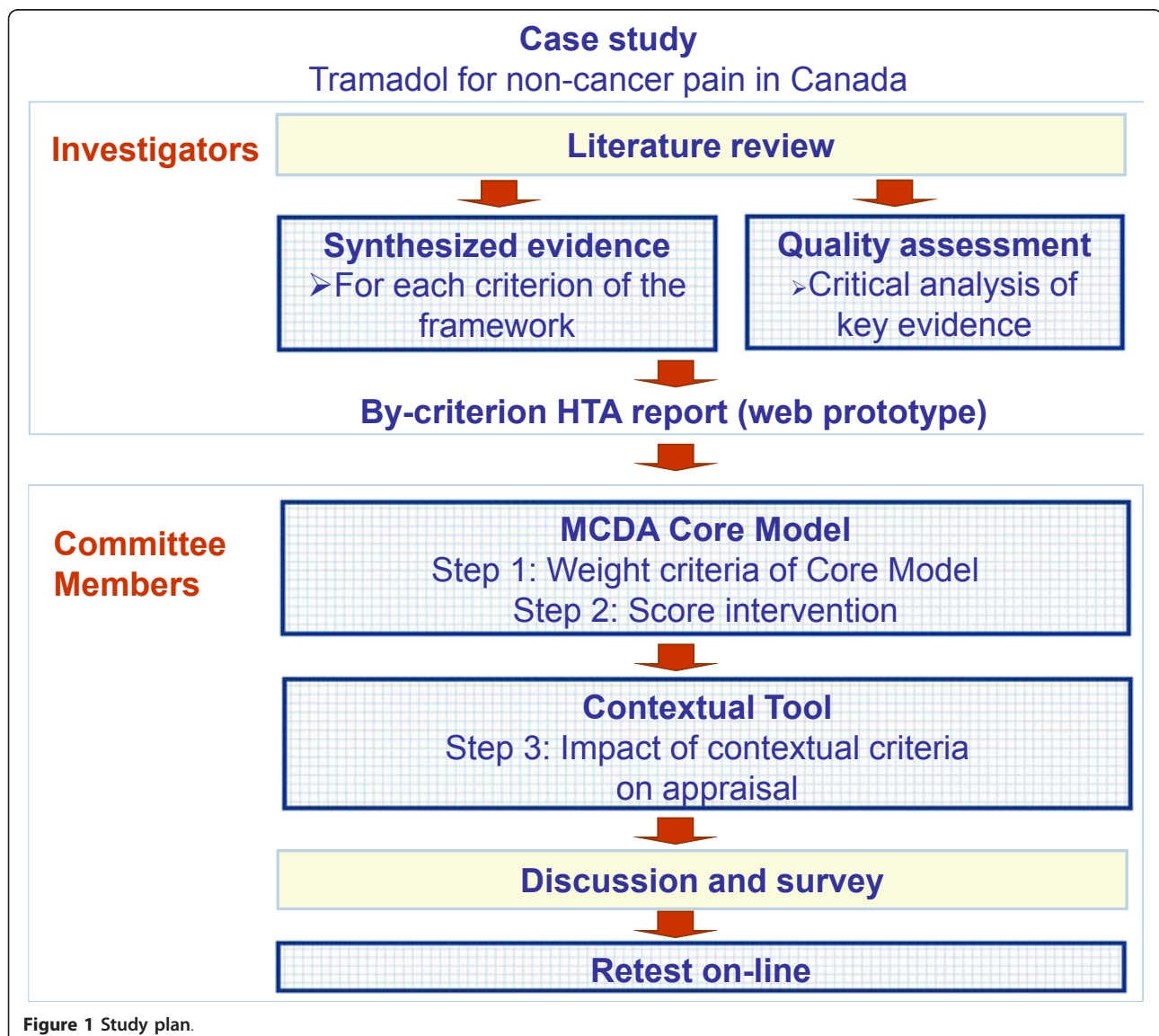


Figure 1 Study plan.

criterion of the framework. Databases and sources searches, including PubMed, EMBASE, Cochrane, Disease Association web sites (Canadian Pain society; Chronic Pain Association of Canada), and websites of the Agency for Healthcare Research and Quality (AHRQ), the National Institute for Clinical Excellence (NICE), the Canadian Agency for Drugs and Technologies in Health (CADTH), and the World Health Organization (WHO), were completed by hand searching of bibliographies. Search terms included: tramadol, opioid/NSAID/COX-2, chronic pain, chronic non-cancer pain, osteoarthritis/low back pain/neuropathic pain/fibromyalgia, randomized controlled trial/non-randomized trial, WOMAC/Pain and sleep Questionnaire/Chronic Pain Sleep Inventory, abuse/dependence, quality of life/QoL/HRQoL, epidemiol*/prevalence/incidence, mortality, guideline/recommendation/clinical practice, pain management, patient-reported outcomes*/PRO, burden, depression/anxiety, cost*/econom*, productivity, ethic*.

To provide the most relevant data to the committee, evidence in the Canadian context was researched, supplemented with information from other countries. Disease information was obtained from prominent reviews and epidemiological studies in Canada and elsewhere. Analyses of the limitations of current options and of current clinical guidelines for pain management in Canada and elsewhere were performed. Clinical, safety and patient-reported outcomes (PRO) evidence for tramadol and comparators for the most important primary outcomes [25,26] were obtained from active-controlled randomized clinical trials (RCTs) and product monograph. Although inclusion of observational studies can supplement evidence from RCTs, [27,28] no such studies were identified. Economic data was obtained from the literature supplemented by analyses provided by the health plan. Critical analysis of key clinical and economic studies was performed using the EVIDEM tools described previously [21] to explore the quality of evidence. For the contextual criteria, scientific and grey literature was searched for information on ethical, historical, and contextual aspects of tramadol and opioid treatment.

Data collected was synthesized for each criterion of the framework following the EVIDEM methodology [21] based on HTA best practices recommendations (Busse et al., 2002 [29]). To streamline access to evidence and limit data overload, an interactive web prototype (Tikiwiki v3.0) was developed to provide highly synthesized data for each criterion ('Lite' version, for a quick grasp of issues), hyperlinked to a version with more details ('Full' version) with further hyperlinks to the full text sources from which data was extracted. The web prototype was also designed to allow committee members to enter weights, scores and impacts for each criterion for online appraisal of the selected medicine.

Field-testing with committee

To explore individual perspectives, during the workshop session (test), each member of the committee (n = 9) were instructed to assign weights individually (on a scale of 1-low to 5-high) to each criterion of the MCDA Core Model, from their perspective in the context of the health plan. For consistency across interventions, committee members were instructed to attribute these weights independently of the intervention; these weights are expected to be defined once and then used throughout appraisals.

Time was allotted on an as-needed basis. This was followed by a period of discussion on each criterion, and committee members were allowed to modify their weights, on an individual basis.

To appraise the intervention, committee members were instructed to score individually (on a scale of 0 to 3) each criterion of the MCDA Core Model, using evidence synthesized for each of them (by-criterion HTA report). This was followed by a period of discussion on each criterion, and committee members were allowed to modify their scores, on an individual basis.

Committee members then explored the six contextual criteria and assigned the type of impact (negative, none or positive) each criterion would have on the appraisal of tramadol, using the colloquial and scientific evidence integrated into the Contextual Tool.

Feedback on the framework, included criteria and process was collected during discussion periods at the first workshop and at a follow-up workshop, and from a questionnaire administered during the follow-up workshop. To explore reliability, a retest was performed at least two weeks after the last session either using the web-based prototype on-line or a hardcopy document.

Data collection and analyses

For the test, weights, scores and impacts were obtained using the hardcopy documents distributed to committee members and entered into Microsoft Excel software. Data entered on-line by panelists (retest) was recorded in a MySQL database and transferred to the Excel software, which was then used to perform statistical analyses.

Descriptive statistics were used and mean \pm standard deviations (SD) were reported for weights and scores.

The MCDA value estimate of the perceived value of tramadol for the treatment of CNCP was obtained by applying an MCDA linear additive model to combine weights and scores [30]. The MCDA value estimate is anchored on a scale of no value (0, e.g., minor symptom relief for a rare, mild condition with numerous alternative treatment options that provides worse efficacy, safety and quality of life and produces no public health benefits but results in major additional spending) to

maximum value (1, e.g., cure for an endemic severe disease with limited treatment alternatives that provides significant improvement in efficacy, safety and quality of life, and produces major public health benefits and healthcare savings) [21].

$$V = \sum_{x=1}^n V_x = \sum_{x=1}^n \left(\frac{W_x}{\sum_{x=1}^n W_x} \times S_x \right)$$

The estimate was thus calculated as the sum of value contribution (V_x) of combined normalized weights (W_x) and scores (S_x) for applicable criteria ($n = 14$) of the MCDA Core Model. Calculations were performed for each committee member (e.i., combining weights and scores for each individual) and then averaging the MCDA value estimates for the 9 committee members.

Agreement between test and retest data was analyzed by calculating intra-rater correlation coefficients (ICCs) for weights, scores and MCDA value estimates. One type of ICCs was calculated following Shrout and Fleiss (1979) [31] methods and classification: the ICC (3, 1) which is based on a two-way mixed analysis of variance (ANOVA) model (general effects of the test and the retest were assumed to be fixed). In addition, the proportion of data pairs that did not differ between test and retest, that differed by 1 point, and that differed by 2 points, was calculated for weights and scores.

Inter-rater correlation coefficients were not calculated since the tool is designed to capture individual perspectives, which are expected to vary among individuals.

Results

By-criterion Health Technology Assessment report

Synthesized data was based on 69 references covering all the criteria of the framework. The highly synthesized version of the by-criterion HTA report with assessment scales is reported in Additional file 1. The web HTA report with all levels of synthesis is available online [32].

The report provides an overview of tramadol, a weak opioid indicated for the treatment of CNCP, and the context of this treatment in Canada. In summary, CNCP is a disabling condition affecting 25% of the Canadian population [33]. Efficacy and safety of current treatments are limited due to ceiling effects and the potential for organ damage [34-46]. In randomized trials, tramadol significantly reduced pain intensity from baseline, but this reduction was not significantly different from NSAIDs, COX-2 inhibitors and opioids [47-51]. The tolerability of tramadol is comparable to that of other analgesics, [47,49,51-56] but drug abuse may be lower with tramadol than with other opioids [57,58]. Significant changes from baseline were observed for patient-

reported outcomes, which were not significantly different than those reported for comparators, including placebo [47,48]. Relevance and validity of trial data was limited by: the trial population (mostly older patients with osteoarthritis (OA) not relevant to WSIB); trial durations (a few weeks), high attrition rates in 3 trials and unclear reporting in 2 trials.

Economic data indicated that, based on a daily cost of tramadol of \$2.24-2.27 (comparators costs range \$0.11 to \$6.61), reimbursement of tramadol would result, depending on models assumptions, in savings of 0.32% of additional expenditures and 0.27% of the current budget for pain analgesics for the health plan budget [unpublished data]. A Dutch cost-minimization analysis indicated that, compared to NSAIDs, tramadol would result in savings of \$534 to \$574 per patient over six months when considering adverse events associated with NSAIDs [59]. Relevance and validity of this study were limited by the setting and OA population, a short time horizon and costs considered (e.g., no costs from productivity losses). To stimulate reflection and discussion in the context of the health plan, information was provided in the Contextual Tool on: the utility of the treatment, [60,61] its efficiency and potential opportunity costs, fairness and access to care for opioid analgesics, [61,62] risk of abuse, [57,63,64] pressures from the Canadian Pain Society to keep tramadol out of the controlled drug schedule, [65,66] historical reviews of the WHO on tramadol [67] and recommendations on tramadol from Canadian agencies. This report was used for field-testing of the framework with the committee.

Field-testing with committee

Decision criteria and committee perspective - weights

Independently of the specific case of tramadol, each member of the committee assigned weights on a scale of 1 to 5 to the criteria of the MCDA Core Model to express individual perspectives. Mean weights and standard deviations are reported in Figure 2. At the committee level, the greatest importance was given to the criteria "Improvement of efficacy/effectiveness" (4.6 ± 0.5) and "Relevance and validity of evidence" (4.3 ± 0.7). The least important criterion was "Comparative interventions limitations" (3.1 ± 1.1). A perfect consensus among committee members was observed for importance of the criterion "Disease severity" (SD: 0). Weights for "Improvement of efficacy/effectiveness" and "Type of medical service" also varied little among committee members (SD: 0.5). The largest divergence of opinions was recorded for the criterion "Comparative intervention limitations" (SD: 1.1).

Survey data and discussion revealed that committee members felt that most of the criteria considered in the MCDA Core Model and the six contextual criteria were

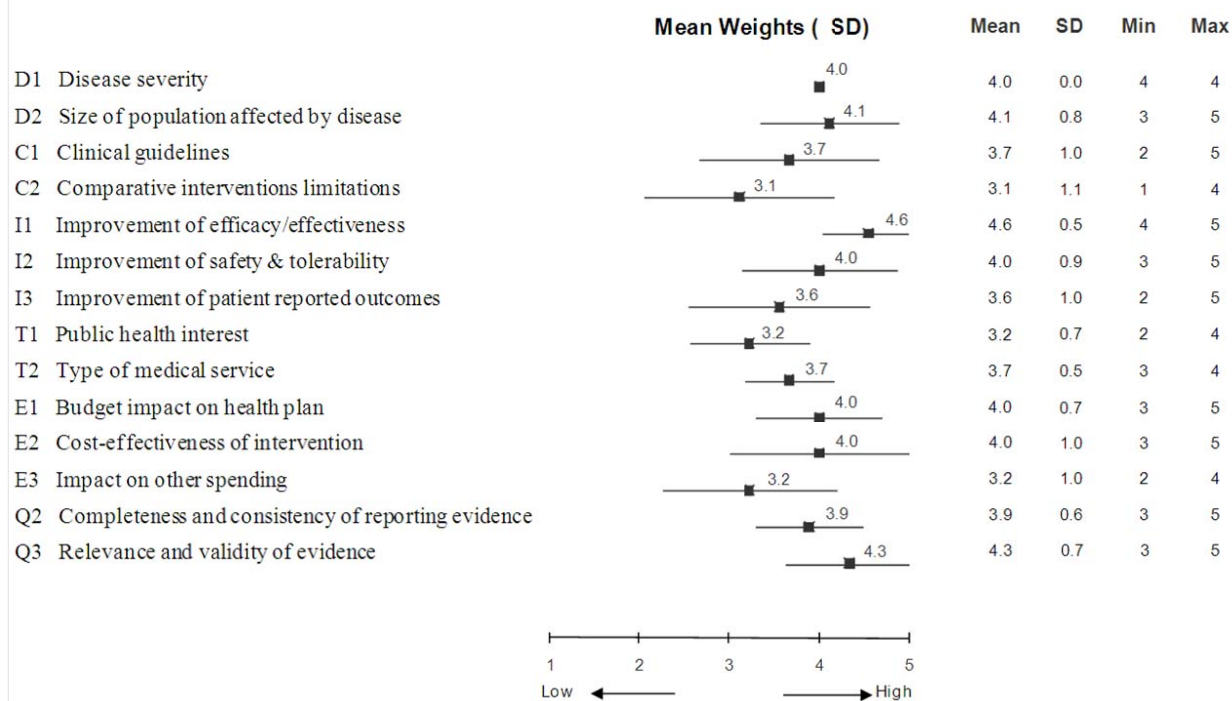


Figure 2 Mean weights for decision criteria of the MCDA Core Model from the drug advisory committee. A five-point weighting scale was used with 1 for lowest weight and 5 for highest weight.

relevant in their context and should be systematically considered. Although the criterion “Disease severity” was considered important by the committee (weight: 4.0 ± 0), the necessity of this criterion was discussed; some committee members noted that because of the type of conditions covered by the health plan (i.e., work-related illness or injury), scores for criterion “Disease severity” may always be high in the committee context. However, it was also noted that some of the population covered by the health plan suffer from really severe diseases and that the framework would capture this aspect. “Public health interest” was considered by some committee members as irrelevant in their context, and there was some controversy on considering the criterion “Political/historical context”.

MCDA Core Model - scores for tramadol

Using synthesized data integrated in the MCDA Core Model and the Contextual Tool (by-criterion HTA report) (Additional file 1), committee members assigned scores to appraise tramadol for each criterion (Figure 3). The highest scores were assigned to the criteria “Size of population affected by disease” (2.6 ± 0.5 , on a scale of 0 to 3), “Disease severity” (1.9 ± 0.3) and “Impact on other spending” (1.8 ± 0.4). In contrast, low scores were given to “Improvement of patient-reported outcomes” (0.9 ± 0.3), “Improvement of efficacy/effectiveness” (1.0 ± 0.0) and “Improvement of safety & tolerability” (1.0 ± 0.5).

The committee demonstrated a unanimous agreement concerning the performance of tramadol with respect to “Improvement of efficacy/effectiveness” (SD: 0.0). In contrast, differences of 2 and 3 points (on a scale of 0 to 3) were observed for “Type of medical service” (SD: 0.8) and “Comparative intervention limitations” (SD: 0.7), respectively. Committee members indicated that average scores for the MCDA criteria presented in Figure 3 had good face validity and represented their opinion of tramadol.

The scoring direction for the criterion “Size of the population”, which stipulates a higher score for interventions for conditions with high prevalence/incidence, was questioned as committee members felt that interventions for rare conditions should not be valued lower than those for more common conditions. Committee members also raised the importance of patient preferences and their perspective on the disease, which could be more explicitly considered under the criterion “Patient-reported outcomes”. When evaluating the criterion “Improvement of efficacy/effectiveness”, committee members noted that while they consider RCTs are the primary sources of data on efficacy, non-randomized studies can sometimes be useful for providing supplementary information on safety and real-life effectiveness.

MCDA Core Model - value estimate for tramadol

The MCDA value estimate for tramadol, which integrates perspectives of committee members (weights) and

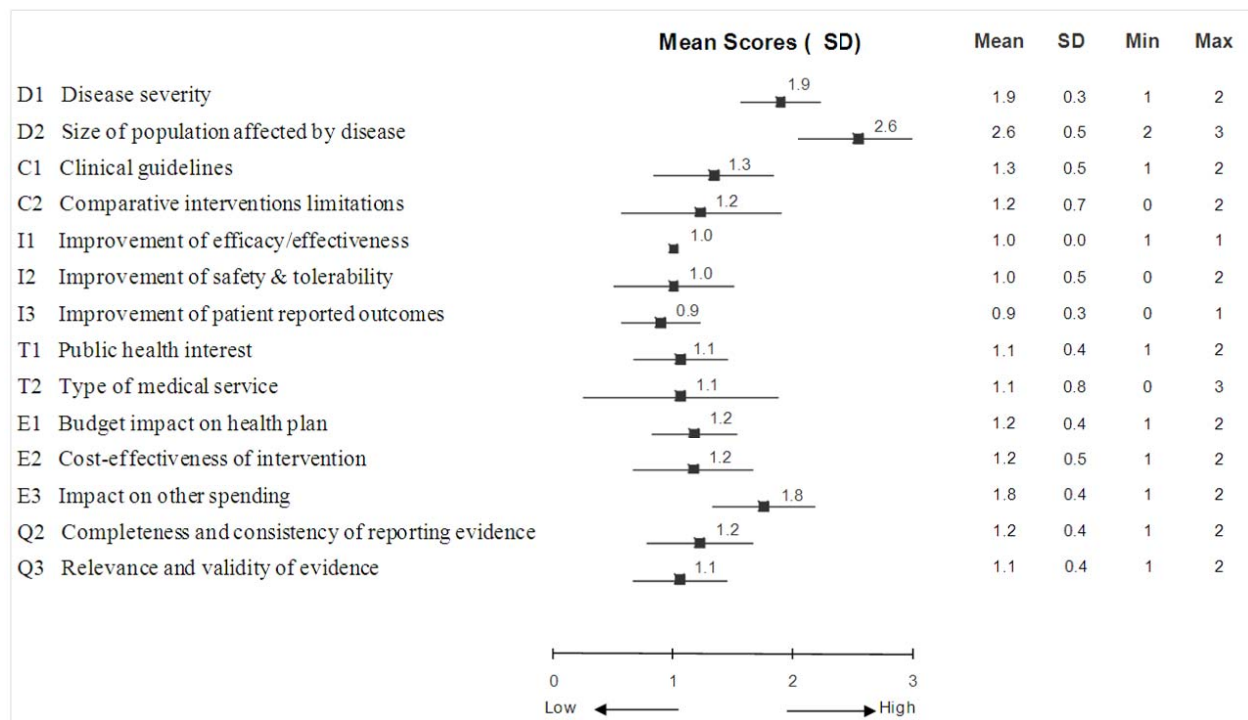


Figure 3 Mean scores for decision criteria of the MCDA Core Model for the appraisal of tramadol by the drug advisory committee. A four-point scoring scale was used with 0 for lowest score and 3 for highest score.

performance measurements (scores), was calculated for each member of the committee by combining normalized weights and scores using a linear additive model. At the individual level, the MCDA value estimates ranged between 0.36 and 0.61 on a scale of 0 to 1. For the group, the overall MCDA value estimate was 0.44 ± 0.07 . More than one quarter (26%) of the estimated value of tramadol was derived from the two criteria of the “Disease impact” cluster (“Disease severity” and “Size of population”) (Figure 4). Relative contributions of the other criteria to the MCDA value estimate for tramadol ranged between 5% and 8%.

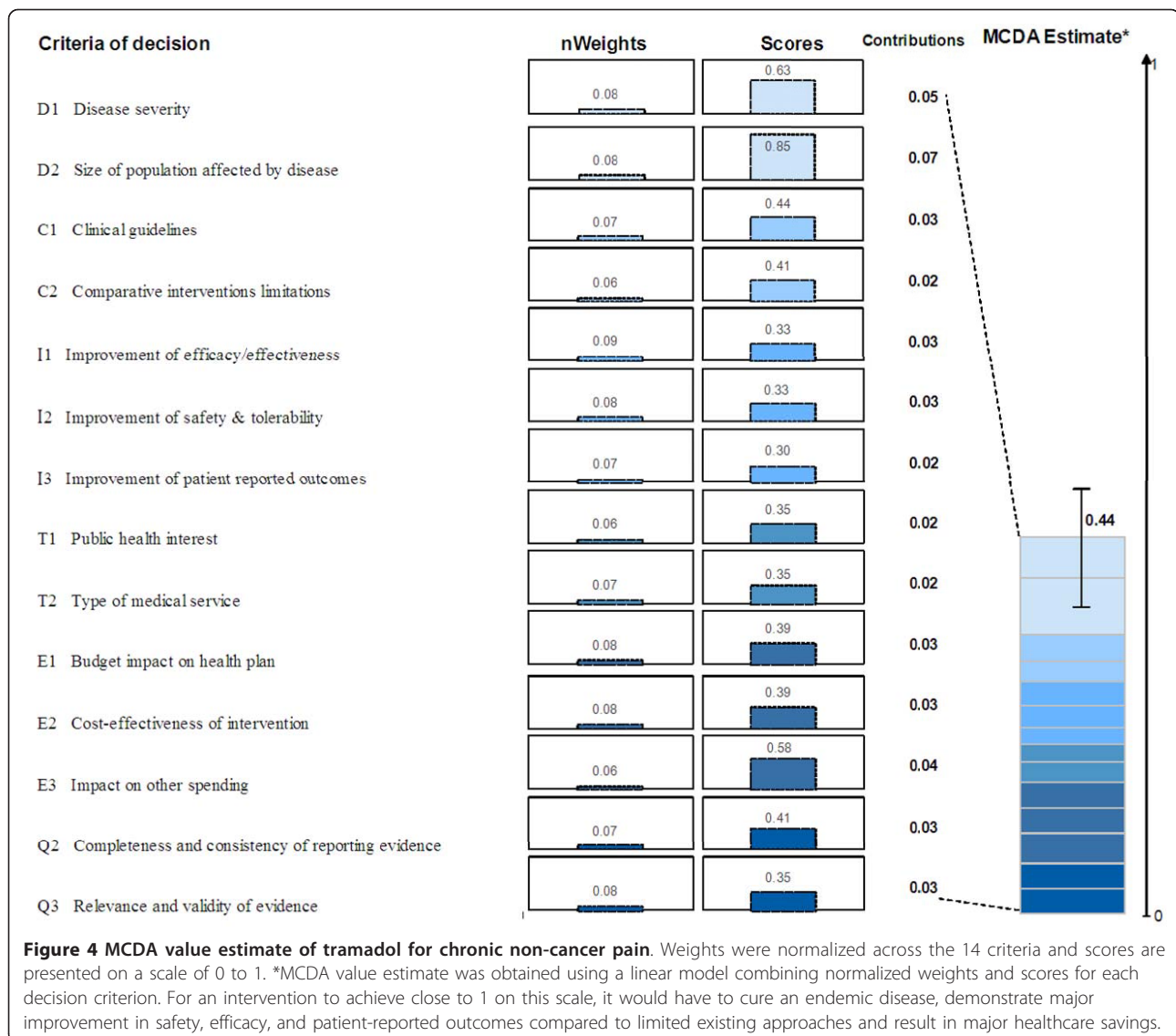
Committee members indicated that quantification of evidence and interpretation of the MCDA value estimate of 0.44 was challenging and required detailed explanations on mathematical calculations used in the framework. It was deemed difficult to interpret the end result of the MCDA Core Model. Scaling with other drugs appraised by the committee (preferably with widely differing MCDA value estimates) would be needed to clarify the meaning of MCDA value estimates and get a better grasp of how the methodology can be used for ranking interventions. Some committee members expressed reluctance to base decisions on cut-off values, which is not the objective of the framework, as it allows for qualitative contextual considerations to modulate the estimate (see section below - Contextual

Tool). In addition, the framework is meant to support the evaluation and deliberation leading to the decision, which was deemed very important by the committee.

Contextual Tool - impacts of contextual criteria on tramadol appraisal

Based on the synthesized information collected for the six contextual criteria (Additional file 1) and their own understanding of the context, committee members considered what type of impact (positive, negative or neutral) each contextual criterion would have on the appraisal of tramadol. The distribution of these impacts within the committee is shown in Figure 5.

For most committee members (8 out of 9), “Goal of healthcare - utility” had a positive impact on the appraisal of tramadol, since relieving pain is aligned with the goals of healthcare in general and the WSIB in particular. Regarding the criterion “System capacity and appropriate use of intervention” opinions were divided (5 indicated a positive, 4 a neutral impact), as committee members indicated: that “abuse with tramadol exists”; that there is a “need for abuse data for the true comparator (e.g., codeine) rather than with more potent hydrocodone which is a misleading comparison”; but that there is a “growing concern/burden about opiate abuse and any drug with potential to decrease that burden deserves special consideration”.



Consideration of “Opportunity cost - efficiency” had a negative impact on the appraisal for most (7 out of 9) committee members as it was noted that “tramadol is more expensive than comparable medicines”. For more than half of committee members (5 out of 9), “Political/historical context” had a negative impact on the appraisal of tramadol. As stated by one committee member “negative recommendations from other agencies have a negative impact on tramadol”.

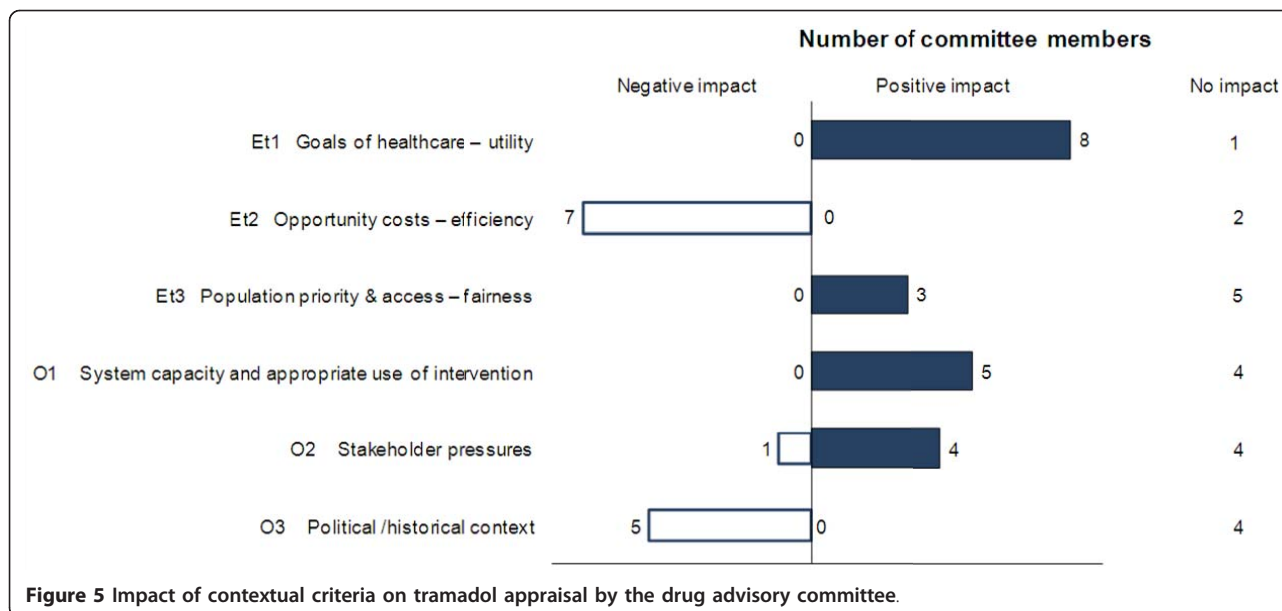
Regarding “Population priority and access and fairness”, 5 out of 8 committee members indicated that this criterion had no (neutral) impact on the appraisal. One member did not provide data for this criterion and indicated that it was not clear how to consider it, highlighting the need for thorough reflection on this criterion which is meant to elicit discussion on priorities of the

health plan. Impacts were mixed for “Stakeholders pressures”.

Explicit consideration of the six contextual criteria forced reflection on a broad range of issues, which were considered to have mixed impacts on the overall appraisal of tramadol.

Feedback from committee on overall approach

Participants indicated that the framework is a good point of departure to help lead group discussions and to ensure systematic, transparent and consistent consideration of important elements that may affect decisionmaking. To cite one committee member: “EVIDEM is a good tool in the sense that it forces each member to think and weight aspects that otherwise would not have been considered”. Another committee member felt that



it may help validate decisions and demonstrate transparency to stakeholders. Some members voiced concern about the amount of work involved in developing the by-criterion HTA report. Others questioned the complexity of the tool and terminology used. One committee member stated that the framework adds clarity to decisionmaking but repeated use may be required to fully capture its utility and get acquainted with its application.

Exploration of reliability

Weighting and scoring during the test and the retest led to the same mean MCDA value estimate of 0.44 (Table 1). The ICC (3, 1) for these data was 0.698, indicating fair to good reproducibility. ICC (3, 1) values of 0.683 for weights and of 0.676 for scores were also a sign of good reliability of test-retest weights and scores.

Between test and retest, 65.1% of weights were identical, 31.7% differed by 1 point (on a scale of 1 to 5) and 3.2% differed by 2 points. The greatest inconsistency between test and retest weights was recorded for the cluster “Context of intervention” (38.9% identical) and the least for the clusters “Disease impact” and “Intervention outcomes” (77.8% identical).

Scores exhibited better consistency between test and retest than weights: 78.8% were identical, 19.8% differed by 1 point (on a scale of 0 to 3) and 1.6% differed by 2 points. The greatest difference between test and retest was recorded for clusters “Context of intervention” and “Type of benefits” (61.1% identical) and the least for clusters “Disease impact” and “Quality of evidence” (88.9% identical).

Discussion

The framework was found useful by the drug advisory committee in supporting systematic consideration of a broad range of criteria to promote a consistent approach to appraising healthcare interventions. Directly integrated in the framework as a “by-criterion” HTA report, synthesized evidence for each criterion facilitated its consideration, although this was sometimes limited by lack of relevant data. This is in agreement with previous studies in which committee members and panelists indicated that the framework was a useful approach to systematic consideration of all aspects of decision, facilitating consistency, transparency and clarity of appraisal and decisionmaking [22,24]. Test-retest analysis showed fair to good consistency of weights, scores and MCDA value estimates at the individual level (ICC ranging from 0.676 to 0.698), thus lending some support for the reliability of the approach. Overall, committee members endorsed the inclusion of most framework criteria and revealed important areas of discussion, clarification and adaptation of the framework to the needs of the committee.

The EVIDEM framework is aligned with the four key features of HTA identified by Battista et al.: [14] policy orientation; interdisciplinary content and process; synthesis of information; and facilitation of dissemination and communication of information. The framework proposes a comprehensive set of criteria to promote systematic and explicit consideration of all aspects of decision, including ethical and system-related issues, considered vital for HTA to fulfill its mandate [68]. Testing revealed the need for adjusting criteria and processes to ensure MCDA approaches can support existing

Table 1 Agreement at the individual level between test-retest for weights, scores and MCDA value estimates for tramadol

	Weights	Scores	Value Estimates
Number of test-retest pairs	126*	126*	9
Mean of test data	3.81	1.31	0.44
Mean of retest data	3.81	1.33	0.44
ICC (3, 1)	0.683	0.676	0.698
Proportion of pairs with no test-retest difference (%)	65.1	78.6	NA
Proportion of pairs with test-retest difference of 1 point (%)	31.7	19.8	NA
Proportion of pairs with test-retest difference of 2 points (%)	3.2	1.6	NA

*9 evaluators × 14 MCDA criteria

NA: Not applicable

ICC: intra-class correlation coefficient, defined according to Shrout and Fleiss (1979) [31]

processes and meet the needs of decisionmakers. The framework is meant to be adapted and some tools have been developed since to facilitate adaptation of the framework based on collaborative development with users and researchers [69].

Standardized procedures for gathering, synthesizing, distilling, and presenting information to inform each criterion, an integral part of the framework, can also contribute to enhancing consistency of the decisionmaking process. As advocated by Straus, [70] Robeson [71] and their coworkers, data was made available on the web in two hyperlinked levels: a 'lite' version with data distilled into key messages –to facilitate knowledge transfer and to reduce the time required to integrate the information presented– and a comprehensive, detailed synthesis with full access to underlying sources of evidence. Although presenting the data at these two levels of synthesis was found to be helpful by some committee members, others would have preferred the 'lite' version to provide more details. Thus, while the comprehensive version of the by-criterion HTA report should always serve as a basis, subsequent knowledge distillation may need to be further optimized to strike the right balance between conciseness and detail. Collaborative development is ongoing to further advance the methodology to synthesize data for each criterion and for different levels of details.

The weighting exercise and following discussion revealed the different perspectives of committee members as captured by the large SDs for some criteria. Variations may also be due to different understanding of the criteria, although detailed definitions were provided and were further clarified during discussions. For consistency across interventions, it is recommended that these weights be attributed once and then used throughout appraisals. Although at the individual level, relative weights vary largely between criteria, once weights are averaged at the committee level, extremes disappear and weights tend to be less distinguishable. Exploration of how the weight elicitation methods (e.g., analytical

hierarchy process [AHP], Simple Multi-Attribute Rating Technique [SMART], point allocation, ranking)[72], impact weight attribution and the overall MCDA value estimate is ongoing to further advance the approach and provide additional tools to adapt the framework to the preferences and needs of users.

Interpretation and utility of the MCDA value estimate (i.e., the figure 0.44) was found challenging by committee members. Indeed, the MCDA value estimates are meant to be used in a comparative manner for ranking healthcare interventions, which was beyond the scope of this case study. An MCDA model adapted from the EVIDEM framework by a district health board uses such an approach to systematically evaluate and prioritize a wide range of healthcare interventions (Sharon Kletchko, MD, personal communication 2011). Although basic principles are explicit, interpretation of the MCDA scale requires acquaintance with the broad range of criteria that are incorporated in a single number. For example, an intervention with a score close to 1, which would constitute a high-end reference point, would have to be something like: a cure for an endemic severe disease with limited or no alternatives that provides significant improvement in efficacy, safety and quality of life, and produces major public health benefits and healthcare savings [21]. A pilot study testing the framework with 13 Canadian healthcare stakeholders showed that the MCDA value scale had discriminating properties, with mean estimates ranging from 0.42 to 0.64 across 10 medicines [23]. However, it should be kept in mind that MCDA value estimates are not meant to be used in a prescriptive fashion, but rather as “a framework conducive for focused discussion.”[73] MCDA value estimates can serve as a basis for establishing a ranking scheme, [9] which can be modulated by ethical and context related considerations. This is often done implicitly in healthcare decisionmaking and is meant to be facilitated by the Contextual Tool. It should also be noted that MCDA value estimates obtained by applying this framework are committee-specific since they reflect the

individual perspectives of committee members as captured by weights.

Although there are numerous approaches to improve healthcare coverage decisionmaking –including the prevailing cost-effectiveness paradigm, program-budgeting and marginal analysis, and the HTA Core Model [74-76], for example – there is no accepted and validated way to identify successful evaluation and decisionmaking and still less consensus concerning the best framework to support decisionmaking [77] or even the most reliable process for weight elicitation [78]. In that sense, our study suffered limitations common to other studies on the validity of decisionmaking approaches. Exploration of reliability revealed good consistency at the individual level which constitutes a first positive step. Higher consistency with scores, which are based on evidence, than with weights, which are based on individual perspective, were observed in this study and may reveal the difficulty to explicit its one's perspective. This might stem from the fact that perspectives are often implicit and that decisionmakers need to reflect on the implications of the criteria. As for every evaluation of healthcare interventions, the information presented on the case study was limited by the information available at the time of the study. Nevertheless, this field study has shown that, through bridging HTA and MCDA, the EVIDEM framework can support appraisal in practice, particularly by promoting systematic consideration of a comprehensive set of carefully selected criteria. A strength of the framework also lies in the acknowledgment and incorporation into its application that decisionmaking is a “fundamentally value-laden enterprise” [68]. These features are combined with firm grounding in scientific evidence, which includes rigorous synthesis and quality assessment, to make the committee's deliberative process as well-informed, comprehensive and explicit as possible. The framework promotes a consistent approach to decisionmaking that can help legitimize decisions and be aligned with the A4R framework set forth by Daniels [79,80].

Conclusions

In a field test with a public health plan drug advisory committee, the EVIDEM framework supported appraisal and deliberations in practice by bridging HTA and MCDA. Feedback from some committee members confirmed that the framework promoted the explicit consideration of a wide range of criteria relevant to decisionmakers. Further field testing is required to establish a frame of reference for appraisal outcomes, optimize and adjust its use in practice, and establish consistency. Further research is needed to collaboratively advance pragmatic MCDA-based frameworks for

appraisal of healthcare interventions, decisionmaking, priority setting and pragmatic knowledge translation.

Additional material

Additional file 1: By-criterion health technology assessment report on tramadol ('Lite' highly synthesized version). References for each statement and links to sources are available on the web version of this report [32].

List of abbreviations used

AHRQ: US Agency for Healthcare Research and Quality; ANOVA: Analysis of variance; CEA: Cost-effectiveness analysis; CNCP: Chronic non-cancer pain; EVIDEM: Evidence and Value: Impact on DEcisionMaking; H2RA: Histamine receptor antagonist; HRQOL: Health-related quality of life; ICC: Intra-rater correlation coefficients; MCDA: MultiCriteria Decision Analysis; NSAID: Nonsteroidal anti-inflammatory drug; OA: Osteoarthritis; PRO: Patient-reported outcomes; QALY: Quality-adjusted life-year; QoL: Quality of life; RCT: Randomized clinical trials; WOMAC: The Western Ontario and McMaster Universities Arthritis Index; WSIB: Ontario Workplace Safety Insurance Board.

Acknowledgements

The authors wish to acknowledge all the members of the drug advisory committee of the WSIB of Ontario for their participation in this study as well Patricia Campbell and Peter Melnyk, BioMedCom Consultants, Montreal for the development of the web-based prototypes. No sources of funding were used to conduct this study and internal sources of support for the study were provided by the WSIB and BioMedCom Consultants.

Author details

¹BioMedCom Consultants inc, Montréal, Québec Canada. ²Workplace Safety Insurance Board of Ontario, Toronto, Ontario, Canada. ³Toronto Rehabilitation Institute, Toronto, Ontario, Canada. ⁴Centre Hospitalier Universitaire Ste Justine, Montréal Québec, Canada.

Authors' contributions

MMG, DR & MW designed the study and reviewed the HTA report and MCDA analyses. MT performed data collection and analyses. HK & MW participated in data analyses. TP and PO participated in data collection and reviewing tools and synthesized evidence. MT, MMG, MW and HK drafted the manuscript. All authors reviewed the manuscript and approved the final version.

Competing interests

The authors declare that they have no competing interests.

Received: 1 October 2010 Accepted: 30 November 2011

Published: 30 November 2011

References

1. Baltussen R, Niessen L: Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc* 2006, **4**:14.
2. Dhalla I, Laupacis A: Moving from opacity to transparency in pharmaceutical policy. *CMAJ* 2008, **178**:428-431.
3. Daniels N, Sabin J: Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philos Public Aff* 1997, **26**:303-350.
4. Nord E, Daniels N, Kamlet M: QALYs: some challenges. *Value Health* 2009, **12**(Suppl 1):S10-S15.
5. Schlender M: The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): no(t yet an) exemplar of a deliberative process. *J Med Ethics* 2008, **34**:534-539.
6. Williams I, McIver S, Moore D, Bryan S: The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technol Assess* 2008, **12**:iii, ix-iii, 175.

7. Baltussen R, Stolk E, Chisholm D, Aikins M: **Towards a multi-criteria approach for priority setting: an application to Ghana.** *Health Econ* 2006, **15**:689-696.
8. Baltussen R, ten Asbroek AH, Koolman X, Shrestha N, Bhattarai P, Niessen LW: **Priority setting using multiple criteria: should a lung health programme be implemented in Nepal?** *Health Policy Plan* 2007, **22**:178-185.
9. Baltussen R, Youngkong S, Paolucci F, Niessen L: **Multi-criteria decision analysis to prioritize health interventions: Capitalizing on first experiences.** *Health Policy* 2010.
10. Nobre FF, Trotta LT, Gomes LF: **Multi-criteria decision making—an approach to setting priorities in health care.** *Stat Med* 1999, **18**:3345-3354.
11. Peacock S, Mitton C, Bate A, McCoy B, Donaldson C: **Overcoming barriers to priority setting using interdisciplinary methods.** *Health Policy* 2009, **92**:124-132.
12. Hutton J, Trueman P, Facey K: **Harmonization of evidence requirements for health technology assessment in reimbursement decision making.** *Int J Technol Assess Health Care* 2008, **24**:511-517.
13. Giovagnoni A, Bartolucci L, Manna A, Morbiducci J, Ascoli G: **Health technology assessment: principles, methods and current status.** *Radiol Med* 2009, **114**:673-691.
14. Battista RN, Hodge MJ: **The evolving paradigm of health technology assessment: reflections for the millennium.** *CMAJ* 1999, **160**:1464-1467.
15. **HTA Resources.** [<http://www.inahta.org/HTA/>].
16. Velasco GM, Gerhardus A, Rottingen JA, Busse R: **Developing Health Technology Assessment to address health care system needs.** *Health Policy* 2010, **94**:196-202.
17. Johri M, Lehoux P: **The great escape? Prospects for regulating access to technology through health technology assessment.** *Int J Technol Assess Health Care* 2003, **19**:179-193.
18. Lehoux P, Williams-Jones B: **Mapping the integration of social and ethical issues in health technology assessment.** *Int J Technol Assess Health Care* 2007, **23**:9-16.
19. EUnetHTA work package 4 team: **HTA core model for medical and surgical interventions** 2007.
20. Hailey D: **Toward transparency in health technology assessment: a checklist for HTA reports.** *Int J Technol Assess Health Care* 2003, **19**:1-7.
21. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D: **Evidence and Value: Impact on DEcisionMaking - the EVIDEM framework and potential applications.** *BMC Health Serv Res* 2008, **8**:270.
22. Goetghebeur MM, Wagner M, Khoury H, Rindress D, Gregoire JP, Deal C: **Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decisionmaking framework to growth hormone for Turner syndrome patients.** *Cost Eff Resour Alloc* 2010, **8**:4.
23. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D: **Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): Applying the EVIDEM framework to medicines appraisal (In Press).** *Med Decis Making* 2011.
24. Miot J, Wagner M, Khoury H, Anderson AN, Rindress D, Goetghebeur M: **Field testing of a multi criteria decision analyses (MCDA) framework for coverage of a screening test for cervical cancer in South Africa.** Presented at ISPOR, Paris; 2009.
25. Cepeda MS, Camargo F, Zea C, Valencia L: **Tramadol for osteoarthritis.** *Cochrane Database Syst Rev* 2006, **3**:CD005522.
26. Deshpande A, Furlan A, Mailis-Gagnon A, Atlas S, Turk D: **Opioids for chronic low-back pain.** *Cochrane Database Syst Rev* 2007, **CD004959**.
27. Atkins D: **Creating and synthesizing evidence with decision makers in mind: integrating evidence from clinical trials and other study designs.** *Med Care* 2007, **45**:S16-S22.
28. Berger ML, Mamdani M, Atkins D, Johnson ML: **Good Research Practices for Comparative Effectiveness Research: Defining, Reporting and Interpreting Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report.** *Value Health* 2009, **12**:1044-1052.
29. Busse R, Orvain J, Velasco M, Perleth M: **Best practice in undertaking and reporting health technology assessments. Working group 4 report.** *Int J Technol Assess Health Care* 2002, **18**:361-422.
30. **Multi-criteria analysis manual.** [<http://www.communities.gov.uk/publications/corporate/multicriteriaanalysismanual>].
31. Shrout PE, Fleiss JL: **Intraclass correlations: uses in assessing rater reliability.** *Psychol Bull* 1979, **86**:420-428.
32. **Open access prototypes of the Collaborative registry.** [<http://www.evidem.org/evidem-collaborative.php>].
33. Boulanger A, Clark AJ, Squire P, Cui E, Horbay GL: **Chronic pain in Canada: have we improved our management of chronic noncancer pain?** *Pain Res Manag* 2007, **12**:39-47.
34. Rosenberg MT: **The role of tramadol ER in the treatment of chronic pain.** *Int J Clin Pract* 2009, **63**:1531-1543.
35. Sunshine A: **A comparison of the newer COX-2 drugs and older nonnarcotic oral analgesics.** *J Pain* 2000, **1**:10-13.
36. Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, et al: **Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group.** *N Engl J Med* 2000, **343**:1520-8, 2.
37. FitzGerald GA, Patrono C: **The coxibs, selective inhibitors of cyclooxygenase-2.** *N Engl J Med* 2001, **345**:433-442.
38. Tambllyn R, Berkson L, Dauphinee WD, Gayton D, Grad R, Huang A, Isaac L, McLeod P, Snell L: **Unnecessary prescribing of NSAIDs and the management of NSAID-related gastropathy in medical practice.** *Ann Intern Med* 1997, **127**:429-438.
39. Whelton A: **Nephrotoxicity of nonsteroidal anti-inflammatory drugs: physiologic foundations and clinical implications.** *Am J Med* 1999, **106**:135-245.
40. Wolfe MM, Lichtenstein DR, Singh G: **Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs.** *N Engl J Med* 1999, **340**:1888-1899.
41. Ahmad SR, Kortepeter C, Brinker A, Chen M, Beitz J: **Renal failure associated with the use of celecoxib and rofecoxib.** *Drug Saf* 2002, **25**:537-544.
42. Aronson MD: **Nonsteroidal anti-inflammatory drugs, traditional opioids, and tramadol: contrasting therapies for the treatment of chronic pain.** *Clin Ther* 1997, **19**:420-432.
43. Dworkin RH, O'Connor AB, Backonja M, Farrar JT, Finnerup NB, Jensen TS, Kalso EA, Loeser JD, Miaskowski C, Nurmikko TJ, et al: **Pharmacologic management of neuropathic pain: evidence-based recommendations.** *Pain* 2007, **132**:237-251.
44. Hansen GR: **Management of chronic pain in the acute care setting.** *Emerg Med Clin North Am* 2005, **23**:307-338.
45. Benyamin R, Trescot AM, Datta S, Buenaventura R, Adlaka R, Sehgal N, Glaser SE, Vallejo R: **Opioid complications and side effects.** *Pain Physician* 2008, **11**:S105-S120.
46. Lynch ME, Watson CP: **The pharmacotherapy of chronic pain: a review.** *Pain Res Manag* 2006, **11**:11-38.
47. Beaulieu AD, Peloso PM, Haraoui B, Bensen W, Thomson G, Wade J, Quigley P, Eisenhoffer J, Harsanyi Z, Darke AC: **Once-daily, controlled-release tramadol and sustained-release diclofenac relieve chronic pain due to osteoarthritis: a randomized controlled trial.** *Pain Res Manag* 2008, **13**:103-110.
48. Biovail Laboratories I: **Double-blind, randomized, dose-ranging, parallel-group comparison of the efficacy and safety of extended release Tramadol Hydrochloride (Tramadol HCl ER) 100 mg, 200 mg and 300 mg, Celecoxib 200 mg and placebo in the treatment of osteoarthritis of the knee and/or hip** 2003.
49. Mullican WS, Lacy JR: **Tramadol/acetaminophen combination tablets and codeine/acetaminophen combination capsules for the management of chronic pain: a comparative trial.** *Clin Ther* 2001, **23**:1429-1445.
50. Pavelka K, Peliskova Z, Stehlikova H, Ratcliffe S, Repas C: **Intraindividual differences in pain relief and functional improvement in osteoarthritis with diclofenac or tramadol.** *Clin Drug Investig* 1998, **16**:421-429.
51. Rauck RL, Raj PP, Knarr DC, Denson DD, Speight KL: **Comparison of tramadol and acetaminophen with codeine for long-term pain management in elderly patients.** *Current Therapeutic Research* 1994, **55**:1417-1431.
52. Biovail Pharmaceuticals Canada: **Product monograph. Ralivia** 2009.
53. Labopharm Inc.: **Product monograph. Tridural** 2008.
54. Purdue Pharma: **Product monograph. Zytam** 2009.
55. Malonne H, Coffiner M, Fontaine D, Sonet B, Sereno A, Peretz A, Vanderbist F: **Long-term tolerability of tramadol LP, a new once-daily formulation, in patients with osteoarthritis or low back pain.** *J Clin Pharm Ther* 2005, **30**:113-120.

56. Nossol S, Schwarzbald M, Stadler T: Treatment of pain with sustained-release tramadol 100, 150, 200 mg: results of a post-marketing surveillance study. *Int J Clin Pract* 1998, **52**:115-121.
57. Cicero TJ, Inciardi JA, Adams EH, Geller A, Senay EC, Woody GE, Munoz A: Rates of abuse of tramadol remain unchanged with the introduction of new branded and generic products: results of an abuse monitoring system, 1994-2004. *Pharmacoepidemiol Drug Saf* 2005, **14**:851-859.
58. Cicero TJ, Wong G, Tian Y, Lynskey M, Todorov A, Isenberg K: Co-morbidity and utilization of medical services by pain patients receiving opioid medications: data from an insurance claims database. *Pain* 2009, **144**:20-27.
59. Liedgens H, Nuijten MJ, Nautrup BP: Economic evaluation of tramadol/paracetamol combination tablets for osteoarthritis pain in the Netherlands. *Clin Drug Investig* 2005, **25**:785-802.
60. Reddy BS: The epidemic of unrelieved chronic pain. The ethical, societal, and regulatory barriers facing opioid prescribing physicians. *J Leg Med* 2006, **27**:427-442.
61. Brennan F, Carr DB, Cousins M: Pain management: a fundamental human right. *Anesth Analg* 2007, **105**:205-221.
62. Health Canada meeting Re: scheduling of Tramadol. [http://canadianpainsociety.ca/Tramadol/Tramadol_JoveyPresentationHealthCanada.doc].
63. Ballantyne JC, Mao J: Opioid therapy for chronic pain. *N Engl J Med* 2003, **349**:1943-1953.
64. Raffa RB: Basic pharmacology relevant to drug abuse assessment: tramadol as example. *J Clin Pharm Ther* 2008, **33**:101-108.
65. Proposal to schedule Tramadol under the CDSA. [http://www.canadianpainsociety.ca/Tramadol/Tramadol_CPS_HealthCanada_Proposal.pdf].
66. A question of balance: The impact of scheduling on pain management In Canada. [http://www.canadianpainsociety.ca/Tramadol/Tramadol_brochure.pdf].
67. WHO Expert Committee on Drug Dependence: thirty-fourth report. (WHO technical report series; no. 942). [http://whqlibdoc.who.int/trs/WHO_TRS_942_eng.pdf].
68. Saarni SI, Hofmann B, Lampe K, Luhmann D, Makela M, Velasco-Garrido M, Autti-Ramo I: Ethical analysis to improve decision-making on health technologies. *Bull World Health Organ* 2008, **86**:617-623.
69. EVIDEM Collaboration. [http://www.evidem.org].
70. Straus SE, Tetroe JM, Graham ID: Knowledge translation is the use of knowledge in health care decision making. *J Clin Epidemiol* 2009.
71. Robeson P, Dobbins M, DeCorby K, Tirilis D: Facilitating access to pre-processed research evidence in public health. *BMC Public Health* 2010, **10**:95-105.
72. Dolan JG: Multi-Criteria clinical decision support. A primer on the use of multiple-criteria decision-making methods to promote evidence-based, patient-centered healthcare. *Patient* 2010, **3**:229-248.
73. Felli JC, Noel RA, Cavazzoni PA: A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med Decis Making* 2009, **29**:104-115.
74. Lampe K, Makela M, Garrido MV, Anttila H, Autti-Ramo I, Hicks NJ, Hofmann B, Koivisto J, Kunz R, Karki P, et al: The HTA core model: a novel method for producing and reporting health technology assessments. *Int J Technol Assess Health Care* 2009, **25**(Suppl 2):9-20.
75. Peacock SJ, Richardson JR, Carter R, Edwards D: Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). *Soc Sci Med* 2007, **64**:897-910.
76. Mitton C, Donaldson C: Health care priority setting: principles, practice and challenges. *Cost Eff Resour Alloc* 2004, **2**:3.
77. Sibbald SL, Singer PA, Upshur R, Martin DK: Priority setting: what constitutes success? A conceptual framework for successful priority setting. *BMC Health Serv Res* 2009, **9**:43.
78. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM: Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess* 2001, **5**:1-186.
79. Daniels N: Decisions about access to health care and accountability for reasonableness. *J Urban Health* 1999, **76**:176-191.
80. Daniels N: Justice, health, and healthcare. *Am J Bioeth* 2001, **1**:2-16.

Pre-publication history

The pre-publication history for this paper can be accessed here:
<http://www.biomedcentral.com/1472-6963/11/329/prepub>

doi:10.1186/1472-6963-11-329

Cite this article as: Tony et al.: Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. *BMC Health Services Research* 2011 **11**:329.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

