PUBLIC HEALTH 128 (2014) 896–903

Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Original Research

Prioritising public health guidance topics in the National Institute for Health and Care Excellence using the Analytic Hierarchy Process

B.P. Reddy [a,*,d], M.P. Kelly [b], P. Thokala [a], S.J. Walters [a], A. Duenas [c]

[a] School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

[b] National Institute for Health and Care Excellence (NICE), 1st Floor, 10 Spring Gardens, London SW1A 2BU, UK

[c] IÉSEG School of Management (LEM-CNRS), socle de la Grande-Arche, 1 parvis de la Défense, 92044 Paris-La Défense Cedex, France

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form
21 May 2014
Accepted 4 July 2014
Available online 17 October 2014

Keywords:
MCDA
Health economics
Analytic Hierarchy Process
NICE
Public health
Problem structuring

ABSTRACT

Objectives: The Centre for Public Health (CPH), at the United Kingdom's National Institute for Health and Care Excellence (NICE) is responsible for producing national guidance relating to the promotion of good health and the prevention and treatment of disease. Given the challenges of developing guidance in this area, choosing the most appropriate topics for further study is of fundamental importance. This paper explores the current prioritisation process and describes how the Analytic Hierarchy Process (AHP), a multi criteria decision analysis (MCDA) technique, might be used to do so.

Study design: A proposed approach is outlined, which was tested in a proof of concept pilot. This consisted of eight participants with experience of related NICE committees building scores for each topic together in a 'decision conference' setting.

Methods: Criteria were identified and subsequently weighted to indicate the relative importance of each. Participants then collaboratively estimated the performance of each topic on each criterion.

Results: Total scores for each topic were calculated, which could be ranked and used as the basis for better informed discussion for prioritising topics to recommend to the Minister for future guidance. Sensitivity analyses of the dataset found it to be robust.

Conclusions: Choosing the right topics for guidance at the earliest possible time is of fundamental importance to public health guidance, and judgement is likely to play an important part in doing so. MCDA techniques offer a potentially useful approach to structuring the problem in a rational and transparent way. NICE should consider carefully whether such an approach might be worth pursuing in the future.

* *Corresponding author*. Tel.: +353 1 4103427; fax: +353 1 4730596.
E-mail addresses: b.reddy@sheffield.ac.uk, reddybr@tcd.ie (B.P. Reddy).
[d] Present address: National Centre for Pharmacoeconomics, St James's Hospital, Dublin, Ireland.

http://dx.doi.org/10.1016/j.puhe.2014.07.003

Introduction

The Centre for Public Health (CPH), the public health division of the National Institute for Health and Care Excellence (NICE), is responsible for producing national guidance relating to the promotion of good health and the prevention and treatment of disease.[1] Because of the inherently political nature of many public health topics,[2] in some cases the Secretary of State for Health or other Ministers in the Department of Health will propose potential topics, for guideline development, directly to the CPH. In other cases, the procedures for prioritising between topics for future guidance are performed by the CPH, though the final decision remains with the relevant government Minister.

Topics for public health guidance are currently appraised at a thrice yearly NICE Topic Advisory Workshop (TAW), a sitting committee made up of experts and lay members of the public. In advance of the meeting, briefing papers are prepared by the CPH and NICE's Information Services team, developing the proposed topic and describing how it would likely work in practice. Participants at the workshop discuss and rank the topics by consensus, bearing in mind a wide variety of concerns, including the potential political constraints.

Multi criteria decision analysis (MCDA) techniques can be used to better structure complex decision problems[3] and increase the transparency of the decision process. Discussions and applications of these approaches have grown in popularity in medical decision making settings at both policy[4–8] and patient level.[9,10] The Analytic Hierarchy Process (AHP), pioneered by Saaty,[11] is one such approach potentially useful in public health settings. Multiple relevant quantitative and qualitative criteria inherent in public health prioritisation are incorporated into the decision process, and combined in a relatively intuitive manner for stakeholders from numerical and non-numerical backgrounds. Total scores are built up for each topic under consideration by a series of pair-wise comparisons. The approach allows decision makers to collaboratively translate independent, subjective judgements into numerical scores in a rational and consistent manner.

Judgement inevitably plays a key role in medical decision making alongside any available evidence, and associated assumptions should be made explicit where possible.[12] Simon[13] characterised decision making as 'problem solving' and pointed out that it is generally carried out more effectively when the 'problem' is well structured.[14] In complex decision problems, ensuring consistent and well thought out choices are made can prove difficult without an appropriate, analytical approach.[9] Decision makers may instead use ad hoc simplifications, thereby losing potentially useful information and producing unnecessarily poorer decisions.[15] NICE makes social value judgements when necessary and uses explicit discussion as a tool for doing so,[16] though it does not go as far as prescribing and weighting the relevant criteria in doing so. By making explicit decision makers' reasoning and assumptions, MCDA techniques may be fairer[17] than more opaque approaches. The technique may also offer improved consistency, transparency and accountability.[5] Explicitness allows for increased public scrutiny and criticism,[18,19] potentially leading to greater public confidence in the decision and process.[5] Nonetheless, others have expressed concern that explicit weightings may lead to increased challenges over decision makers' weightings and scorings.[20] However, it is hoped a two way dialogue in this way would help ensure that the process is better explained to the public and hence better appreciated,[21] and that in turn it may ensure that more representative criteria and weightings are used.

Given the lack of high levels of ambiguity in evidence in many public health settings,[22] expert opinion is utilised throughout the CPH's committees to better inform decisions.[23] MCDA techniques offer a range of approaches to structure this process in a methodical and transparent way. Formal decision analytic techniques, including MCDA, have not been previously applied to public health topic selection within NICE. For a number of reasons, outlined above, they may prove useful in the future. This paper outlines how an AHP approach might be used to help with prioritising topics for selection for the development of public health guidance following a similar approach used to revise HTA processes in Canada,[24] and presents the findings of a piloted workshop testing the technique.

Methods

Format of the pilot

A pilot approach was used in order to discover the unforeseen problems to be mitigated before use in practice[25] which is naturally important given the gravity of NICE decisions.[5] A deliberative, *decision conference* style setting was chosen in order to mimic how AHP could be used in a TAW style meeting. To do so, an impartial facilitator works iteratively with stakeholders to generate an explicit model intended to help those present to think more clearly about the relevant issues.[26] With participants working together to weight criteria and score the topics, the approach allows participants to aid thinking and generate a shared understanding of the issues.[19]

Make up of committee

Alongside the facilitator (BR), eight participants were present: two members of staff from the CPH, three lay members of NICE's Public Health Interventions Advisory Committee (PHIAC), two public health experts and one public health statistician. None were current members of the TAW, as this was seen to be potentially problematic, but all had worked on other NICE committees in the past, were familiar with NICE processes and volunteered to take part after being invited in a group email from the NICE representative (MK). The workshop took place on a single half day, with some remaining topic scores completed later by email.

Identifying criteria

The first stage of the process was the consideration of which criteria might be used in order to best differentiate a 'good topic' from a poor one for the development of public health guidance. As way of an introduction to the concept of criteria, each participant was given three post-it notes and asked to

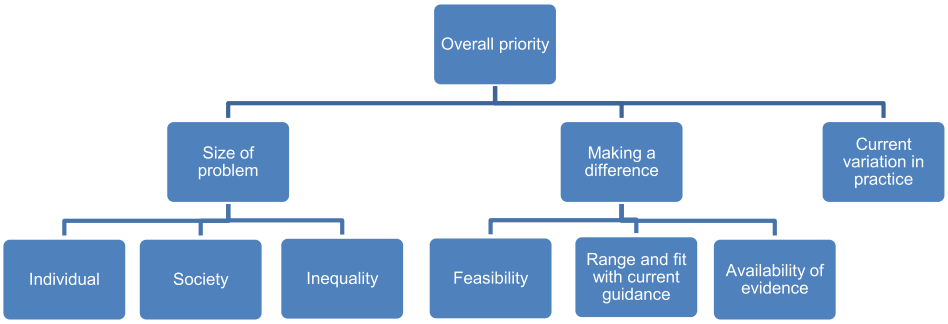

Fig. 1 – Final derived criteria in hierarchy.

write a useful criterion on each, without conferring with others present. Participants were then asked to place these on a wall, positioning them beside any criteria already there that reflected similar themes. As such it was hoped that clusters of relevant criteria might emerge[e] fostering further discussion. Over a number of iterations designed to remove redundant or doubly counted criteria, the participants converged around the hierarchy of criteria shown in Fig. 1.

Prioritising the criteria by weighting

In MCDA approaches, criteria are weighted by comparing their relative importance to the decision. AHP uses an intuitive approach similar to Likert scales to do so, building up weightings based on pair-wise comparisons of criteria. Two criteria are selected and decision makers choose which is more important, and by how much, according to the scale in Table 1. The next pair of criteria is then considered. Over time a complete picture of the relative importance of each criterion can be derived, which is checked for consistency in order to ensure a rational and meaningful answer. Explicit numerical weights are then derived for each criterion using the relational matrix's maximal eigenvector. These weights are then considered by the decision makers in order to ensure that they are acceptable and reflect their intuitive feelings of relative importance of each.

In this workshop, participants were initially asked to privately write down whether they felt the *Size of the problem* was more important than *Making a difference*, and by how much according to the scale in Table 1. These were announced, and the geometric mean of these individual results used as a starting point for discussions until a consensus score for the overall group arose. In this *Making a difference* was between moderately and strongly more important (a score of four) and hence the *Size of the problem* was moderately to strongly less important (the reciprocal, hence ¼). *Size of the problem* was then compared with *Current variation in practice*, and so on. Once the matrix was complete, the consistency index of the

Table 1 – Fundamental scale used in AHP comparisons for weighting and scoring, from Saaty (1990).

Intensity of importance on an absolute scale	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance of one over another	Experience and judgement strongly favour one activity over another
5	Essential or strong importance	Experience and judgement strongly favour one activity over another
7	Very strong importance	An activity is strongly favoured and its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2,4,6,8	Intermediate values between two adjacent judgements	When compromise is needed
Reciprocals	If activity *i* has one of the above numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	

[e] Five clusters of themes initially emerged: 1. Number of people (population) affected; Burden of disease [two participants]; Significant burden of disease; Topic will improve the health of individuals significantly. 2. Is there harmful practice that needs to be challenged by presenting the evidence in a NICE guideline?; Equity; Potential to impact on inequalities in health; Topic will help reduce postcode lottery; What is the need in terms of reducing health inequalities?; Equity of access to information and services; Does it affect rural or urban areas?; Access; Screening in primary care; Implementable-equitably, affordably, practically. 3. Information gathering of intelligence; Topic will tackle an increasing health issue; Minimum of adverse side-effects or potential for unwanted consequences; Significant evidence for missed opportunities for effective interventions; Evidence of population based health improvement; The need for evidence to improve practice (is it likely to have a positive impact). 4. Cost effectiveness [two participants]. 5. Policy.

Table 2 – Implied weights of criterion headings.

	Size of problem	Making a difference	Variation of practice	Size of problem	Making a difference	Variation of practice	Geometric mean	Normalised weights
Size of problem	*Equal*	Moderately-strongly less important	Very moderately more important	1	0.25	2	0.79	0.22
Making a difference	Moderately-strongly more important	*Equal*	Moderately more important	4	1	3	2.29	0.63
Current Variation of practice	Very moderately less important	Moderately less important	*Equal*	0.5	0.33	1	0.55	0.15

relationships between criteria was tested using the standard approach.[27] The final results of the criterion headings, the second level of Fig. 1, are shown below in Table 2. The normalised geometric mean of each criterion is calculated, giving the matrix's maximal eigenvector, containing the weights of the criteria. For example, the geometric mean of the *Current variation in practice* criterion is calculated [((0.5*0.333*1)(1/3)) = 0.55] and then this is normalised in relation to the other means to calculate the criterion weight [0.55/(0.79 + 2.29 + 0.55) = 0.15]. Using this approach, the most important of the criterion headings was *Making a difference*, with a weight of 0.63.

The subsequent criteria in the branches on the bottom level of Fig. 1 were compared similarly, and their absolute weights calculated by multiplying the derived relative weight by the weight of their associated criterion heading above.

Estimating the performance of the topics on each of the criteria

The next stage of the process required that each topic under consideration be scored on each of the criteria. This could only be partially completed at the workshop, and scores on the remaining criteria were completed by email. The process follows a similar pattern to the weighting process explained in the previous section. Participants were asked to describe which of a pair of topics performed better on a given criterion, and by how much, using the fundamental AHP scale. By continuing this process for all such pairs, the scores for how each topic performs on that criterion can be derived. Topics are then compared on the next criterion, resulting in scores for each topic. The results for the *Current variation in practice* criterion completed at the workshop are shown in Table 3.

Results

Given scores for each potential topic on each criterion and the associated weight of these criteria, a 'total score' for each topic can be derived using a weighted sum approach. These can be ranked and used as the basis for better informed discussion for prioritising topics for recommendation for future guidance to the Minister. These are shown in Table 4. The scores for *Current Variation in Practice*, previously calculated in Table 3 are shown in the fourth column, alongside its weight as calculated in Table 2. As both weights and criterion scores are normalised, the total scores also sum to one.

Sensitivity analyses

Health care decisions are by their nature often highly uncertain – Arrow stated in a seminal paper[28] that 'all the special

Table 3 – Performance of each topic on 'Current variation in practice' criterion.

	Sickle cell screening	Substance misuse	Tackling smoking through the media	Fluoridation of water	Pain as a public health problem	Score for *Current variation in practice*
Sickle cell screening	*Equal (1)*	Equal (1)	Moderately-strongly more variation (4)	Moderately less variation (1/3)	Moderately less variation (1/3)	0.14
Substance misuse	Equal (1)	*Equal (1)*	Moderately more variation (3)	Moderately less variation (1/3)	Very moderately less variation (1/2)	0.14
Tackling smoking through the media	Moderately-strongly less variation (1/4)	Moderately less variation (1/3)	*Equal (1)*	Strongly-v strongly less variation (1/6)	Strongly less variation (1/5)	0.05
Fluoridation of water	Moderately more variation (3)	Moderately more variation	Strongly-v strongly more variation (6)	*Equal (1)*	Equal (1)	0.36
Pain as a public health problem	Moderately more variation (3)	Very moderately more variation (2)	Strongly more variation (5)	Equal (1)	*Equal (1)*	0.32

Table 4 – Topics ranked by total weighted score.

Absolute weights	0.29	0.28	0.15	0.11	0.07	0.06	0.04	1.00
	Making a difference – Feasibility	Making a difference – Evidence available	Current variation in practice	Size of problem – Societal	Size of problem – inequality	Making a difference – range and fit	Size of problem – individually	Total score
Fluoridation of water	0.23	0.25	0.36	0.09	0.18	0.22	0.10	0.23
Tackling smoking through the media	0.22	0.29	0.05	0.36	0.16	0.16	0.13	0.22
Substance misuse	0.24	0.18	0.14	0.26	0.27	0.18	0.27	0.21
Pain as a public health problem	0.13	0.10	0.32	0.23	0.18	0.25	0.22	0.17
Sickle Cell Screening	0.18	0.17	0.14	0.06	0.21	0.19	0.29	0.17

features of this industry, in fact, stem from the prevalence of uncertainty'. Sensitivity analysis is therefore a vital part of such decision processes. While AHP approaches required the use of subjective judgement, MCDA approaches in general are however remarkably insensitive to imprecision in scoring,[29] and the decisions makers' preferences reflected in such orderings tend to be exceptionally robust.

Sensitivity analyses are shown in Table 5, displaying the required increase (or decrease) in each criterion's relative weighting to change the final ordering of the topics. The most sensitive of the rankings was between first and second place; for 'Tackling smoking through the media' to overtake 'Fluoridation of water' either the *Current variation in practice* criterion's relative weighting would have to be reduced by 19% (from 0.151 to 0.122) or *Societal size of problem* be increased in importance by 30%. Given the broad consensus achieved between participants, and consistency found for each criterion, such changes seem unlikely. Other changes seem increasingly

Table 5 – Relative percentage increase required in criterion weightings for Topic B to get a higher total score than Topic A.

	Current weightings	0.29	0.28	0.15	0.11	0.07	0.06	0.04
Topic A	Topic B	Making a difference – feasibility	Making a difference – evidence available	Current variation in practice	Size of problem – societal	Size of problem – inequality	Making a difference – range and fit	Size of problem – individually
Fluoridation of water	Tackling smoking through the media	*Infeasible*	71%	−19%	30%	*Infeasible*	*Infeasible*	775%
Fluoridation of water	Substance misuse	*Infeasible*	*Infeasible*	−65%	117%	369%	*Infeasible*	319%
Fluoridation of water	Pain as a public health problem	*Infeasible*	*Infeasible*	*Infeasible*	381%	*Infeasible*	*Infeasible*	1175%
Fluoridation of water	Sickle Cell Screening	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	849%
Tackling smoking through the media	Substance misuse	*Infeasible*	−39%	91%	*Infeasible*	181%	936%	224%
Tackling smoking through the media	Pain as a public health problem	*Infeasible*	−88%	115%	*Infeasible*	*Infeasible*	831%	1303%
Tackling smoking through the media	Sickle Cell Screening	*Infeasible*	*Infeasible*	411%	*Infeasible*	*Infeasible*	*Infeasible*	863%
Substance misuse	Pain as a public health problem	*Infeasible*	*Infeasible*	127%	*Infeasible*	*Infeasible*	798%	*Infeasible*
Substance misuse	Sickle Cell Screening	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*	*Infeasible*
Pain as a public health problem	Sickle Cell Screening	45%	39%	−28%	−41%	337%	*Infeasible*	282%

improbable, and for 40 of the 70 possible weighting changes, changes in rankings are infeasible regardless of the size of the increase or decrease. The results appear to show a robust dataset. If presented in practice as part of a TAW meeting and results are found that are more sensitive to change, further discussion could ensure that the issue can be further investigated and borne in mind in ultimate decision making.

Discussion

Interpretation of results

According to the total scores shown in Table 4, 'Fluoridation of Water' is the best performing topic based upon the selected criteria, and 'Sickle Cell Screening' is the worst. However, in practice further discussion is required to ensure that recommendations reflect the broader concerns of the health services, NICE and the public, as extensive discussion is typically required to transform health care prioritisation decisions into a simple Yes or No answer.[30] This will also likely provide the opportunity to increase understanding and the range of perspectives included and 'provide a basis for thoughtful and informed moral choices'.[6] It is therefore envisaged that the TAW (or some other similar) committee would continue to play a key role and this stage to address issues related to uncertainty. The AHP approach outlined does not give the 'correct answer' or relieve decision makers of their responsibility to use appropriate judgement; it can only inform this process.[31] However it is likely to provide an initial snapshot of how the topics perform on key criteria and may highlight clear 'winner' and 'dud' topics, making subsequent discussion not just better informed, but calmer, easier and quicker.

Is the approach appropriate?

As previously stated, explicit weighting of the criteria and scores of the proposed topics is not necessarily without risks, particularly in a public body such as NICE. However, MCDA techniques' explicitness afford increased legitimacy and further public commentary arising out of this process may be healthy and could be used to inform future updates to criteria and their weights. Their criteria were intended to be illustrative of how the process might work, but for use in practice clearer definitions would be required to better foster understanding and consistency over time. It is also likely that some version of the 'Feasibility' criterion would be required. Currently topic selection is the only phase of NICE's decision process at which 'resource impact' is explicitly considered. After this point, NICE is expected to recommend all cost effective approaches to increasing public health care, regardless of the likely impact on the overall health budget. However, it would be potentially damaging to recommend any such approaches that are unlikely to prove affordable, not least in public health, where the borders of what constitutes a 'health' intervention can be vague.

MCDA approaches typically require extra resources and effort than standard approaches and are not always worthwhile.[19] It is unclear whether this would be the case in this setting. Nonetheless, the pilot showed AHP to be rather intensive, vindicating the decision to consider only five topics at this workshop. The authors could not get all scores completed in the allotted time and the process was clearly draining for all involved. Participants may become quicker at assigning scores with practice, but nonetheless this approach is unlikely to be feasible for the 10–15 topics currently typically considered at TAW meetings. If an AHP technique is to be used, further changes to current processes may therefore be required to accommodate such an approach, such as more regular TAW meetings to consider smaller batches of topics can be considered. Other MCDA approaches may allow for less collaboration and resources on the day, especially by assembling the performance of topics on criteria before the meeting by literature review.[32] However, given relative lack of clear evidence available in public health settings, this approach may not be suitable in practice, which was why the judgement based AHP was used for this pilot. If used in practice, NICE's Information Services team might continue to provide briefing documents (as it currently does) to the participants, but scores could be derived after discussion.

However, one potential concern with the AHP approach is that it will be open to criticism of the technique itself. While AHP is quite accessible to non-experts and mathematically elegant, this comes at a cost. It has been found that by introducing a new topic for consideration [if, in this example, after the meeting a sixth topic was introduced and scoring calculations revised], the rankings of the topics previously considered may change, a concept is known as rank reversal.[33] Given the flaws associated with any such prioritisation approach – not least ad hoc discussion – there remains a need to match the most appropriate approach with the given decision problem. Ideally these approaches should also be compensatory,[15] to allow rankings of topics, be easy to understand for participants and so on. AHP broadly met these criteria, but its limitations must be borne in mind.

Was it a success?

Topics were successfully ranked using this approach, so it can be shown to work on this level. Participants had very little practical information about the topics ranked and no subsequent discussion or refinement was possible, so while not directly comparable, it is instructive to investigate the differences between the predictions and findings of the actual TAW committee. 'Fluoridation of water' and 'Pain as public health problem' (ranked first and fourth respectively) were both rejected, while final decisions related to the other approaches remain unclear at this time. It is possible that other criteria may have been used to prioritise topics, or that there may have been specific concerns given the topics under consideration are not captured by the general criteria used for this pilot. Further discussion may have ultimately changed the final rankings, but the very uncertainty over what happened at the TAW highlights the lack of transparency inherent in the current prioritisation process.

This was an extremely demanding exercise for those taking part. It is unclear if the approach would have made the process easier than the status quo, though there are reasons to suspect it might over time. It appears that smaller batches

of five or so potential topics may prove more feasible to compare at TAWs using AHP than current numbers of 10–15 topics. It may be worthwhile to include the highest scoring rejected topic from the previous meeting to decrease the impact of variation of topic standards between batches and act as a better benchmark across time. For any potential MCDA approach under consideration in this setting, it would clearly be preferable if the criteria and their weightings were used consistently over time rather than changed at each meeting. This would free up time for scoring and discussion, but more importantly would better ensure that results are comparable and consistent over time. Sensitivity analyses should be carried out as part of the meeting so participants can discuss their likely implications. Furthermore, participants are likely to become more comfortable and faster using the approach over time, though this familiarity could potentially lead to increased risk of gaming. The requirement to discuss and refine the model's rankings afterwards may help to reduce this bias. Whether or not NICE would be willing to publicise explicit criteria to be used at this stage at the current time is unclear. Given any scores used will be supplemented by further discussion, the ultimate decision making process will still be undertaken confidentially if this is a perceived requirement.

Judgement is at the heart of the AHP approach used, from the early stages until final revisions. Given the lack of hard evidence available a priori in public health settings, the use of 'subjective' measures such as these is inevitable. It may be better therefore to instead embrace them, better understand them and apply them in the most appropriate way possible.

Limitations

The fact that the workshop had to take place on a single half day led to a number of limitations to the pilot study. The necessarily illustrative nature of the criteria was also potentially a problem, and ideally the criterion weighting and scoring would have taken place over separate days to allow greater reflection on the nature of the criteria and reduce the cognitive burden placed on participants. The lack of time available for reflection meant that participants were at times confused as to the purpose of each criterion, despite having chosen them themselves a couple of hours previously. Further scoring and comments were conducted by email, removing the possibility for face to face discussions and consensus building at this stage. Geometric means of these scores were used in their place, as advised by Saaty,[34] but this would not always be appropriate in practice.

If the process were to be used recurringly over a longer period of time, this may throw up further issues not captured in a one off pilot, such as the potential for increased gamesmanship. This would have to be managed effectively, such as ensuring that experienced participants do not dominate discussions. The approach may also have consequences outside the control of NICE, such as whether ministers would avoid the TAW process completely by fast tracking chosen topics. Such issues are hard to predict and could not been investigated in the pilot.

Due to the lack of available briefing papers and background information, the pilot could not directly mimic the approach likely to be used. This impacted on participants' scoring of topics on the criteria and likely therefore on their total scores. The very fact it was a pilot, rather than the sitting TAW, meant that the findings could not be directly put into practice. However it has highlighted a number of potential stumbling blocks, as intended, hopefully ensuring that a suitable version of this work can be applied in future.

Conclusions

Given the relative lack of availability of firm evidence in public health interventions, choosing the right topics for guidance at the earliest possible time is of fundamental importance, and judgement is likely to play an important part in doing so. MCDA techniques offer a potentially useful approach to structuring the problem in a rational way, along with the opportunity to make explicit the judgements used as part of the decision making model. While there may be some issues with doing so publicly, such explicitness adds to the legitimacy of the approach and may ultimately increase the public's faith in and engagement with NICE's decision making process. Many decisions in public health will remain debatable, and any rankings produced by the AHP model must be refined by further discussion, but this approach may help make some decisions to accept or reject topics easier and more consistently. This paper does not answer all potentially relevant questions and further research may be required, but it is clear that there is room for improvement in the current topic selection process. NICE should consider carefully whether such an approach might be worth pursuing in the future.

Author statements

The authors would like to thank the participants, who gave their time voluntarily and anonymously, and without which this work would not have been possible; Natalie Bartle who helped coordinate the workshop; and others at NICE who supported the project throughout. The anonymous referees made a number of useful comments on a prior draft. The research was funded by ESRC CASE award ES/I026088/1, entitled 'Multi criteria approaches to public health decision making', with NICE as the non-academic partner. The ESRC had no role in preparing this document. Full ethics approval for the project was given in advance by the School of Health and Related Research (ScHARR) ethics committee in the University of Sheffield. The authors have no competing interests.

REFERENCES

1. Kelly MP. Public health guidance and the role of new NICE. *Public Health* 2005;**119**(11):960–8.
2. Turnock BJ. *Essentials of public health*. Jones Bartlett Learn; 2007.
3. Franco LA, Montibeller G. *Problem structuring for multicriteria decision analysis interventions. Wiley Encyclopaedia of operations*

research and management science. New York: John Wiley and Sons; 2012.
4. Orr S, Wolff J, Morris S. *What values should count in HTA for new medicines under value based pricing in the UK?*. Available at: www.ucl.ac.uk/cpjh/mcda.pdf; 2011.
5. Devlin NJ, Sussex J. *Incorporating multiple criteria in HTA: methods and processes*. London: Office of Health Economics; 2011.
6. Airoldi M, Morton A, Smith J, Bevan G. *Healthcare prioritisation at the local level: a socio-technical approach. Priority-setting for population health working paper series, Working Paper No 7*. London: Department of Management; 2011.
7. Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. *J Public Health*; 2012::14.
8. Youngkong S, Teerawattananon, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. *Health Res Policy Syst* 2012;**10**:6.
9. Dolan JG. Involving patients in decisions regarding preventive health interventions using the analytic hierarchy process. *Health Expect* 2000;**3**(1):37–45.
10. Cunich M, Salkeld G, Dowie J, Henderson J, Bayram C, Britt H, Howard K. Integrating evidence and individual preferences using a web-based multi-criteria decision analytic tool: an application to prostate cancer screening. *Patient: Patient-Centered Outcomes Res* 2011;**4**(3):153–62.
11. Saaty TL. *The analytic hierarchy process: planning, priority setting, resource allocation*. New York: McGraw-Hill; 1980.
12. Dowie J. 'Evidence-based', 'cost-effective' and 'preference-driven' medicine: decision analysis based medical decision making is the pre-requisite. *J Health Serv Res Policy* 1996;**1**(2):104–13.
13. Simon HA. *Models of man: social and rational*. New York: Wiley; 1957 [from Williams & Bryan 2007].
14. Williams IP, Bryan S. Cost-effectiveness analysis and formulary decision making in England: findings from research. *Soc Sci Med* 2007;**65**(10):2116–29.
15. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Allocation* 2006;**4**:14.
16. Rawlins MD, Culyer AJ. National institute for clinical excellence and its value judgements. *Br Med J* 2004;**329**:224–7.
17. Daniels N. Accountability for reasonableness. *Br Med J* 2000;**321**:1300–1.
18. Sen A. *Development as freedom*. New York: Alfred A. Knopf; 1999.
19. Mussen F, Salek S, Walker S. *Benefit risk appraisal of medicines: a systematic approach to decision making*. UK: John Wiley and Sons; 2009.
20. Kennedy I. *Appraising the value of innovation and other benefits, a short study for NICE*. NICE. Available at: www.nice.org.uk/media/98F/5C/KennedyStudyFinalReport.pdf; 2009.
21. Baltussen R, Youngkong S, Paolucci F, Niessen L. Multi-criteria decision analysis to prioritize health interventions: capitalizing on first experiences. *Health Policy* 2010;**96**(3):262–4.
22. Owen L, Morgan A, Fischer A, Ellis S, Hoy A, Kelly MP. The cost effectiveness of public health interventions. *J Public Health*. 2012;**34**(1):37–45.
23. NICE. *Methods for the development of NICE public health guidance*. 3rd ed. NICE. Available at: http://publications.nice.org.uk/methods-for-the-development-of-nice-public-health-guidance-third-edition-pmg4/; 2012.
24. Husereau D, Boucher M, Noorani H. Priority setting for health technology assessment at CADTH. *Int J Technol Assess Health Care* 2010;**26**(3):341–7.
25. Sanders WB, Pinhey TK. *The conduct of social research*. New York: The Dryden Press; 1983.
26. Phillips LD, Bana e Costa C. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Ann Oper Res* 2007;**154**(1):51–68.
27. Saaty TL. How to make a decision: the analytic hierarchy process. *Eur J Oper Res* 1990;**48**:9–26.
28. Arrow K. Uncertainty and the welfare economics of medical care. *Am Econ Rev* 1963;**53**(5):951–73.
29. Von Winterfeldt D, Edwards W. *Decision analysis and behavioural research*. Cambridge: Cambridge University Press; 1986.
30. Phillips LD, Fasolo B, Zafiropoulos N, Beyer A. Is quantitative benefit-risk modelling of drugs desirable or possible. *Drug Discov Today Technol* 2011;**8**(1):E3–10.
31. Wilson ECF, Rees J, Fordham RJ. Developing a prioritisation framework in an English Primary Care Trust. *Cost Eff Resour Allocation* 2006;**4**:3.
32. Matrix Insight. *Prioritising investments in preventative health*. Health England. Available at: www.healthengland.org/publications/HealthEnglandReportNo5.pdf; 2009.
33. Belton V, Geare T. On a short-coming of Saaty's method of analytic hierarchies. *Omega* 1983;**11**(3):143–4.
34. Saaty TL. Decision making with the analytic hierarchy process. *Int J Serv Sci* 2008;**1**:1.