

## Methods for public health economic evaluation

Frew, Emma; Breheny, Kathryn

DOI:

[10.1002/hec.3916](https://doi.org/10.1002/hec.3916)

License:

None: All rights reserved

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Frew, E & Breheny, K 2019, 'Methods for public health economic evaluation: a Delphi survey of decision makers in English and Welsh local government', *Health Economics*, vol. 28, no. 8, pp. 1052-1063.  
<https://doi.org/10.1002/hec.3916>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

This is the peer reviewed version of the following article: Frew, E, Breheny, K. Methods for public health economic evaluation: A Delphi survey of decision makers in English and Welsh local government. *Health Economics*. 2019; 28: 1052– 1063.  
<https://doi.org/10.1002/hec.3916>, which has been published in final form at <https://onlinelibrary.wiley.com/doi/full/10.1002/hec.3916>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

## **Methods for public health economic evaluation: a Delphi survey of decision makers in English and Welsh local government**

### **Abstract**

Standard reference-case methods recommended for health technology appraisals do not translate well to a public health setting. This paper reports on a Delphi survey designed to elicit views of public health decision makers in England and Wales, about different methodological elements of economic evaluation. This is important as methods should align with the objective function of decision-makers. The Delphi survey comprised two rounds, with round 1 allowing open-ended recommendations in addition to 5-point Likert scale responses. The final survey comprised 36 questions and levels and strength of agreement were assessed using median values and mean absolute deviation of the median. The Delphi panel (n=66) achieved high levels of agreement for costs, health, wellbeing and productivity impact to be important elements within an economic evaluation. The panel agreed that evaluations should be relevant to the local context and include costs and consequences over a lifetime horizon. There was a call for the transparent reporting of costs and effects for different population subgroups, and for different sectors. Overall, the panel revealed a preference for a flexible approach, understanding that economic evidence fits within a dynamic process of decision making. These results provide empirical evidence to inform guidelines for public health economic evaluation.

**Key words:** Economic evaluation; Public Health; Government Policy

## **1. Introduction**

In England and Wales since 2013, local government authorities have had responsibility for the public health of their local populations. These local authorities provide over 700 services spanning a wide range of functions including maintenance of highways, housing, environmental services, planning and development, children's services, and protective services. Typically, they have an elected council leader who selects a deputy leader and a Cabinet comprising elected councilors, each having responsibility for a portfolio of services. Cabinet members are supported by council officers, including the Director of Public Health, who provide technical advice when formulating policy and when making decisions. In recent times, in response to national austerity measures in the UK, local authority budgets have drastically reduced and this has caused unprecedented financial pressure. . In terms of public health spending, latest figures reveal that total planned spending on public health by English and Welsh local authorities was £3.4 billion in 2017/18 and this represents a 5.2% fall in spending from 2013/14 (UK Government, 2018).

(Dis)investment decisions can be supported by economic evaluation. However, it is now widely accepted that methods for economic evaluation typically applied to health care technologies do not transfer well into a public health context (Kelly, McDaid, Ludbrook, & Powell, 2015; McIntosh & Tudor Edwards, 2018; Payne, McAllister, & Davies, 2012; Weatherly et al., 2009). Suggested reasons are that public health decision making requires evidence on outcomes beyond just a health focus (Whitehead & Shehzad, 2010); the criteria for decision making is not exclusively upon maximizing health outcome; often, decisions involve costs and consequences that span several sectors (Coast, 2004); and, the evaluation of public health investment may require consideration of the systemic effects of multiple interventions that form one strategy thus requiring a systems approach (Frew, Ng, Coast, Hollingsworth, & Smith, 2018). Furthermore, there is a perception that cost-effectiveness is not sufficient to justify public health investment as cost-savings are required (Ferguson, 2016). In recognition of this, decision-making bodies such as the UK National Institute for Health and Care Excellence (NICE) have updated methods guidance for public health appraisal including a societal perspective and consideration of methods such as cost-benefit analysis (National Institute for Health and Care Excellence (NICE), 2014). Although all these issues are relevant to a national context, to have impact at the local level where the responsibility for public health falls, an additional level of understanding is required on the mechanisms for how economic evidence informs decision making. Local authorities are complex organisations (Hunter et al., 2016) and academic research evidence forms only one (often very small) part of decision making that is heavily influenced by complex human behaviour interacting with the political and financial context. Therefore, to have impact at the local level there is a need for academic researchers to better consult with the local stakeholders whose actions they are trying to support as this will help generate guidelines for economic evaluation within public health that are co-produced, rather than academic-led.

With this aim, a Delphi survey was conducted to measure the level of agreement for different methodological elements within an economic evaluation from the perspective of local public health decision makers. The Delphi method provides a technique to consult with a panel of ‘experts’ and transform individual opinion into a group value on what constitutes helpful, supportive economic evidence (Dalkey, 1969). The method was first developed by the RAND Corporation for technological forecasting and is recommended as a method for when ‘individual judgements must be elicited and combined to arrive at a decision that cannot be taken by one person’ (Cantrill, Sibbald, & Buetow, 1996). Here, the results of a Delphi survey with public health decision makers working within local government authority settings are reported.

## **2. Methods**

### **2.1 The Delphi panel**

The participants were approached using a combination of purposive and snowballing sampling (Polit & Hungler, 1997). First, personal targeted invites were sent via email to participants recommended by health economics and public health colleagues known to the main researcher (n=38) as personal invites have been shown to boost response rates within Delphi surveys (Keeney, Hasson, & McKenna, 2011). Second, a targeted email invite was sent to individuals working within a local authority context with a decision-making role in public health identified from an internet search (n=195); and third, invited participants were asked to forward the invitation email to any colleagues in a decision-making role whom they thought could provide an opinion on economic methods (a snowballing approach) (n=17). All participants were targeted based on their knowledge and experience of public health decision making within a local authority setting. Most participants were located within England and Wales apart from two known contacts to the main researcher who were Directors of Public Health located in Ireland. To keep track of response rates the participants were asked to alert the researcher on how many colleagues whom they had forwarded the survey email invitation. In total, 250 potential participants received the initial invite to participate in the survey.

### **2.2 The Survey**

To determine the acceptability and usefulness of alternative economic evaluation approaches and methods used, a list of initial methodological ‘items’ were generated from two main sources: one-to-one interviews with public health officers working within a local government authority setting reported separately (Frew & Breheny, 2019); and known methodological issues identified from a purposive literature review on the barriers to using economic evaluation methods within public health (Kelly et al., 2015; Payne et al., 2012; Weatherly et al., 2009). The questionnaire was piloted and

refined using internal (academic) health economics colleagues (n=7) and external public health colleagues (n=8) working within one local authority setting. The main feedback from this pilot phase was on providing extra clarification on the meaning of economic terminology, spelling out acronyms, some technical problems with the online survey design, and inserting a progress bar to indicate how far through the survey.

In the first round, participants were asked about their current job role, number of years' experience working within public health, educational attainment and any details of formal health economics training. Participants were asked to report on their current use of economic evidence for decision making by reporting on what sources were most often used to access economic evidence; and then on how often, on average, economic evidence was considered when making decisions. For clarity, the following definition was provided: *'the term economic evidence means formal evidence on the funding required for an intervention (or a service), and how the money invested, compares to the outcomes achieved'*.

The next section of the survey asked participants to rate the importance of different types of economic evaluation evidence (nine items) when making decisions linked to investment, and then when making disinvestment decisions. For each of these, importance was rated on a five-point scale from 1 = very important to 5 = not at all important.

Next, participants were asked an open-ended question to elicit their views on the main barriers to using economic evaluation evidence when making funding decisions. Likert scale items followed, asking their views on different aspects of economic evaluation methodology linked to the generalizability of results (two items); the decision-making criteria (five items); the time period that should be considered (four items); what costs to include and how they should be reported (four items); what outcomes to consider (four items); how the results should be reported (five items); impact on equity (one item); and on how multi-sector budgets should be considered (one item). Round one concluded with an open-ended question for any further suggestions for economic evaluation methods. In all questions, five-point rating scales were either phrased as 1=very important to 5=not important, or from 1=strongly agree to 5=strongly disagree. In total, round one contained questions linked to 26 methodological features (items) of economic evaluation. At the end of round one, the participants were asked if they would be willing to complete a second round by providing a contact email address.

The number of Delphi rounds was fixed at two from the beginning as it was felt that asking the participants to do any more would lead to fatigue and increased attrition (Walker & Selfe, 1996). In round two, the participants received feedback on the descriptive statistics (group median) for each item from round one and asked to consider their individual score in light of this feedback. They did

not receive information on their individual scores. In round two, no items were discarded, but new items were added based on the qualitative comments received in round one. These new items added were linked to the barriers to using economic evaluation evidence (five new items); generalizability (one new item), the time horizon (one new item), what outcomes to include (one new item), and on how economic evidence should be reported (two new items). The round two survey therefore contained 36 items in total and at the end of each sub-section, the participants were provided with the option of an open-ended response. Table 2 lists all items included and a copy of the survey instruments are available within the supplementary appendix. Both rounds were conducted during April and July 2018, respectively. Any non-responders in round two received one follow-up reminder email one week after the initial invite.

### **2.3 Determining agreement**

There is little guidance from the literature on how to measure level of agreement within a Delphi study (Hasson, Keeney, & McKenna, 2000). For this study, the level of agreement was measured for each methodological item by the group median and strength of agreement by the mean absolute deviation from the median (MADM) which is the average distance (on a 5-point Likert scale) of individual participant ratings from the panel's median rating (Hutchings, Raine, Sanderson, & Black, 2005; Raine et al., 2015). Following the approach adopted by other published Delphi studies (Raine et al., 2015; Taylor et al., 2016), the statements were then ranked by the median scores with medians of 1 to 2 indicating agreement, 3 no agreement, and 4 to 5 disagreement with the statements. The MADM was then used to measure the strength of agreement, split into thirds according to the MADM (high <0.46, moderate 0.46 to 0.57, low > 0.57).

The survey was administered using the software SurveyMonkey, a web-based survey programme. Ethics approval for conducting the study was obtained from the University of Birmingham Science, Technology, Engineering and Mathematics Ethical Review Committee (ERN\_17-0445).

### **3. Results**

Sixty-six participants completed the round one (R1) survey which represented a 26% response rate overall (66/250 participants initially invited). At the end of the R1 survey, 39 participants provided an email address indicating a willingness to participate in round 2 (R2) and of those, 29 participants completed the round two survey (74% response rate, 29/39 participants). As we asked more detailed questions about characteristics in R1 and did not want to overburden respondents with repeating all of these questions in R2, the characteristics for both samples are shown in Table 1.

[Table 1 here]

The map in Figure 1 shows the geographical spread of the R2 Delphi panel indicating a balanced distribution across England and Wales, and two participants located in Ireland.

**[Figure 1 about here]**

Most of the sample were either public health directors or members of a public health team. There was a broad range of years of experience and a large proportion of the sample had received formal training in health economics but this was limited to either a module within a higher degree course or a one-day short course.

The R1 participants were asked to rank on a 5-point scale (1= most often to 5 =least often) how they would normally access economic evidence to support public health funding decisions. The most common sources cited were the UK National Institute for Health and Care Excellence (NICE) (median 2) and Public Health England reports (median 2); followed in descending order by peer-reviewed academic literature (median 3), a google/internet search (median 4); and lastly, contact with known health economics colleagues (median 5). When asked how often economic evidence is consulted to support funding decisions, 'usually' (40% of R1 sample, n=26/66) was the most common response.

In R1, the decision makers were asked an open-ended question to describe the main barriers to using economic evidence. These comments were then transcribed and summarized and inserted as 'items' in R2, together with a 5-point scale (1 = strongly agree to 5 = strongly disagree) to measure the level of agreement within the R2 panel. There were five items generated that referred to the lack of available evidence. The barriers were an assumption that economic evidence is only required if the investment is deemed high enough; that the evidence is not reliable; not relevant; or too complicated/not accessible. There was high agreement (median 2) with three of the five statements generated from R1 that economic evidence 'tends to not always be available', 'it tends to not be relevant to local context', and 'it is not accessible or too complicated'.

Table 2 provides a summary of all responses to the Delphi survey according to the median response and the MADM score.

[Table 2 here]

With respect to what was considered important when making investment decisions, the R2 panel indicated a high level and strength of agreement for the importance placed on costs, health-related outcomes, wellbeing outcomes and productivity impact (median 1 or 2, MADM <0.42). Furthermore,

for disinvestment decisions, the panel revealed a high level and strength of agreement for information on costs and health-related outcomes only (median 1, MADM < 0.35); wellbeing was seen as important but there was a wider variation in response as indicated by the MADM score (MADM = 0.5) For all other elements, the MADM was > 0.46 which means the panel only achieved moderate strength of agreement on the importance of each item.

### *Generalizability*

With respect to generalizability, the panel revealed some mixed results. A high strength and level of agreement was found for both statements: ‘economic evidence should take account of local context’ and ‘economic evidence should be generalizable to lots of different settings’ (median 2, MADM < 0.42). However, when these two statements were combined to form a third statement: ‘economic evidence should be generalizable but take account of local context’, only a moderate strength of agreement was obtained across the panel (MADM = 0.53).

### *Decision-making criterion*

The panel were asked for their views on alternative decision-making criteria and asked to rank their level of agreement (1= strongly agree to 5 = strongly disagree) to each of the following statements:

Economic evidence should...

...focus exclusively on maximising outcome (however defined)

...focus exclusively on minimising inequalities

...focus equally on maximising outcomes and minimising inequalities

...focus on both maximising outcome and minimising inequality with more weight given to maximising outcome

...focus on both maximising outcome and minimising inequality with more weight given to minimising inequalities

Overall, the panel disagreed with the first two statements (median 4) and the variation in response was moderate (MADM = 0.53) indicating that there were some mixed views across the panel. Similarly, there was only a moderate strength of agreement for economic evaluation to have an equal focus on maximizing outcome and minimizing inequality (median 2, MADM = 0.57). For the other two statements, a median score of 3 was revealed which means the panel could neither agree nor disagree with these statements.

### *Time horizon*

The panel were asked about the appropriate time horizon for economic evaluation and asked specifically about budgetary and political cycles to get a sense of how important that was for how

economic evidence should be presented. Overall, the panel were neither in agreement nor disagreement with taking account of budgetary or political cycles (median 3, MADM > 0.6) and instead revealed a high level and strength of agreement for economic evaluation to adopt a time horizon of sufficient length that captures all costs and benefits, or that accounts for lifetime costs and benefits (median 2, MADM < 0.46).

#### *What costs to include and how to present them?*

Relevant to the choice of perspective for the economic evaluation, the panel were asked about what costs should be included. Overall, this was the section of the survey that achieved the highest level and strength of agreement as indicated by the median scores and the MADM. There was a high level of agreement for economic evaluation to present costs by population subgroup and by sector (median 2, MADM < 0.42), and for economic evidence to take account of non-health sector budgets e.g. housing or education budgets (median 2, MADM = 0.46). However, when asked about including set up costs and including all costs required to maintain outcome effects, only a moderate strength of agreement was obtained overall (MADM = 0.57).

#### *Outcomes*

To gauge the level of agreement for capturing different outcomes the panel were asked to respond to each of the following statements:

Economic evidence should...

- ...use money to represent both costs and outcomes
- ...use a single outcome that captures both length of life and health-related quality of life
- ...use a single outcome that captures wellbeing gain
- ...capture a broader set of outcomes beyond health or wellbeing
- ...use multiple outcomes to capture health, wellbeing, and broader outcomes

This was the section of the survey that revealed the greatest variation in response. None of the results were in a form that could lead to a recommendation for methodology. The only statement that achieved a low MADM score (0.42) was on having a single outcome that captures wellbeing gain but this had a median score of 3 which meant most of the panel could neither agree nor disagree with this statement. All other statements had MADM scores > 0.6.

#### *Dissemination guidance*

The panel were asked about alternative methods for disseminating economic evidence. This was to get a sense from the panel on how economic evidence should be reported to maximize impact at the

local level. Very high levels of agreement with low variation were obtained for the use of local reports (median 2, MADM 0.19), and for a four-page briefing report (median 2, MADM 0.46). Moderate variation in response was obtained for using national reports, and 2-10-25 reports (median 2, MADM 0.5). Lower levels of consensus were obtained for having a central repository and for making use of infographics and stories (MADM > 0.61). Perhaps the most interesting finding was that a very high variation in response was obtained for the use of peer-reviewed articles (MADM 0.84) which means that some of the panel either strongly agreed or strongly disagreed with this statement.

### *Equity*

In addition to the statements about the assumed decision criteria within economic evaluation, the panel were asked about methods for considering distributional consequences. The panel achieved a high level and strength of agreement for a formal weighting of outcomes by population subgroup (median 2, MADM 0.46).

## **4. Discussion**

The main aim of this study was to elicit the views of public health decision makers working in local government for different aspects of economic evaluation methodology. This was to provide evidence from the decision maker perspective on what is valued, recognizing that standard methods commonly applied to health technology assessment are not transferable to a public health setting.

Overall, the Delphi panel achieved high levels and strength of agreement for costs, health and wellbeing effects, and productivity impact to be important pieces of information when making resource allocation decisions; and for economic evaluation to be relevant to the local context, to be disseminated using local reports; for a time-horizon that captures all costs and benefits or a lifetime horizon; and for analysis that transparently reports costs and effects for different population subgroups, and by different sectors; and for economic evaluation that formally weights outcomes by population subgroup. Where there was a variation in response, so MADM scores were > 0.57, we explored the distribution of the responses further and found no evidence of an extreme response at the item level nor an individual giving consistently different responses.

Taking each of these findings in turn, with respect to wellbeing effects, there have been methodological advances in recent years for broadening the scope of economic evaluation to include wellbeing effects, with the development of instruments such as the ICECAP (Grewal et al., 2006) and ASCOT (Netton, Burge, & Malley, 2012). There is also increasing literature on how to include

spillover effects although disagreement remains on the normative judgement to include these spillover effects within economic evaluation – see Brouwer, 2018 and McCabe, 2018 for a full discussion (Brouwer, 2018; McCabe, 2018).

For taking account of inequalities, the Delphi panel obtained moderate agreement that economic evaluation should not focus exclusively on maximizing outcome or on minimizing inequality and should take a more balanced approach. This is interesting as distributional consequences are not often formally considered within economic evaluations for health technology assessment (McCabe, 2018) although some methodological advances for inclusion of equity effects have been made over recent years (Cookson et al., 2017). When asked about the methods for how economic evaluation should consider distributional effects, the panel were in high agreement for a formal weighting of outcomes by population subgroup however there was a cautionary note revealed by one participant who felt the weighting needs to be flexible and transparent: *‘perhaps the tools could suggest a weighting based on consensus and also allow the user to vary this to their local setting’*.

There have been three main academic contributions within the literature that have shaped thinking on the design of economic evaluation for supporting public health decision making (Edwards, Charles, & Lloyd-Williams, 2013). The challenges noted by Kelly et al. outlined the need for multiple outcomes and the difficulties with attributing causal effect to public health intervention (Kelly et al., 2015); Weatherley et al. suggested outcome measures beyond the Quality Adjusted Life Year, the need to consider wider-sector costs and consequences, and to take account of inequalities (Weatherly et al., 2009); and Payne et al. noted the complex nature of many public health interventions and a need to broaden the focus of outcomes to wider non-health effects (Payne et al., 2012). The results from this Delphi survey provide empirical evidence that local decision makers agree with these points. However, it appears that although the panel achieved high agreement for factors important to consider for investment decisions such as productivity, wellbeing and health outcomes, when asked about the mechanisms for how this could be done, only low levels of agreement were obtained. For example, one particular aspect of economic evaluation that the panel struggled to agree on was with how health, wellbeing and non-health outcomes should be captured. Only low strength of agreement was obtained for the statements that economic evaluation should use multiple outcomes; should use money to value outcomes; should have a set of outcomes that go beyond health and wellbeing; and should have single outcomes to capture either health or wellbeing.

It is widely acknowledged that the process of policy making is influenced by many complex factors and academic evidence forms only one of the many inputs (Oxman, Lavis, Lewin, & Fretheim, 2009). Only one previous study was identified that conducted interviews with Directors of Public Health (DPH’s) (n=16) located within the Southern region of the UK on how they make the case for

investment in public health (Willmott, Womack, Hollingsworth, & Campbell, 2015). The findings highlight the force of the political context and that DPH's are having to negotiate influence within highly politicized environments; and a recognition that while making the case for evidence-based decision making, DPH's are also adapting practice to include 'forms of knowledge perceived legitimate in local government decision making'. With respect to the politicized environment, the Delphi panel in this study found only low strength of agreement for economic evaluation methods to explicitly consider the political time cycle, instead the panel found high strength of agreement for consideration of all costs and benefits relevant to the decision problem, and for a lifetime horizon, all to inform a narrative to help justify investment when working with local politicians.

Although there are guidance documents available for the conducting and reporting of public health economic evaluations (Edwards et al., 2013) there does seem to be a continued lack of availability of economic evidence for public health investment and for those that do exist, a heterogeneity in methods applied (Edwards et al., 2013; Hill, Vale, Hunter, Henderson, & Oluboyede, 2017; Lung et al., 2017). The UK NICE in 2014 updated its guidance for the evaluation of public health interventions recommending both a public sector and a societal approach that means that all relevant costs and benefits should be included. However, in a recent review of interventions for alcohol prevention for example, it was reported that the methodological challenge of including multi-sectorial costs and effects is still prevalent and not yet resolved due to difficulties of consistently and accurately measuring non-health cost (Hill et al., 2017).

Edwards is very clear at the end of their narrative synthesis of UK guidance documents for public health evaluation that health economists need to "think multidisciplinary and approach public health interventions as 'public policy economists' rather than trying to stretch the medical model"(Edwards et al., 2013). The findings of this Delphi survey here further support this conclusion that local government public health decision makers value economic evidence that clearly reflects the societal context for decision making and broadens the framework to include non-health costs and effects and distributional consequences, and that this needs to be transparently reported to support engagement across local-government functions. The methodological challenge remains however on how opportunity cost is valued. Nonetheless, there is a clear message from this Delphi panel that 'bundling up' cost-effectiveness results into one cost-effectiveness ratio is not helpful, rather the panel revealed a preference for the transparent reporting of how costs and effects are distributed by sector and by population subgroup. This form of reporting will not lead to policy recommendations based on established thresholds of cost-effectiveness, but instead provide information to support the process of decision making and enable a discussion between decision makers and local Councillors on the implications of resource investment and impact on local budgets, offset against local priorities and needs within local populations. The panel favoured a flexible approach, with the ability for decision

makers to value the outcomes that are relevant to their local populations. This finding fits very closely with the ‘impact inventory’ advocated by the US 2<sup>nd</sup> Panel on Cost-Effectiveness that lists comprehensively all costs and effects resulting from a resource allocation decision using a societal perspective (Sanders et al., 2016), and with the UK NICE guidelines to consider approaches such as a cost-consequence analysis when evaluating public health interventions. Perhaps this might offer a starting point for the ‘how to’ capture health, wellbeing and productivity outcomes within economic evaluation that the local decision makers struggled to agree on.

A main strength of this paper is the use of the Delphi framework to elicit the views of local public decision makers on what their preferences are for economic evaluation methods. The Delphi method was chosen because with other consensus group methods such as the nominal group technique there is potential for one individual to dominate the group. The Delphi avoids this as participants are approached by email so are anonymous to other group members. It also gives participants time to reflect on their responses (Evans & Crawford, 2000). The participants were not selected randomly but purposively selected based on their expertise and knowledge of the decision-making context. If there is agreement that the design of an economic evaluation should reflect the goals of the decision maker, then the results from this Delphi provide a strong steer towards the use of a societal perspective and for transparent reporting of costs and effects disseminated using a range of tools such as local and national reports, and study briefings. Interestingly, only low strength of agreement was obtained for disseminating through peer-reviewed journal articles – the most common means of academic dissemination. In terms of limitations of the study, definitions were not provided for all concepts/terms used and a basic understanding of these terms was assumed. Some respondents might have therefore interpreted these concepts differently.

This survey was not designed to achieve consensus but rather to measure the level and strength of agreement across the panel on which items of economic evidence are most supportive for public health decision making. It was not the intention to generate a set of guidelines for economic evaluation as consideration also needs to be given to what is theoretically plausible and methodologically sound. It was intended as a study to formally elicit stakeholder opinion on what they would like to see generated that would genuinely support public health decision making in a local setting, considering the practical, financial, organisational and political contexts that these stakeholders operate within. With that in mind the panel revealed a preference for an evaluative framework that is broadened to capture non-health costs and outcomes and that more consideration needs to be given to the local context and understanding of the infrastructure that exists to support the implementation of policies. Furthermore, the Panel revealed a preference for a flexible approach as economic evidence fits within a dynamic process of decision making and is the catalyst that brings decision makers from across different government departments to work together. As a way forward,

we have the following three recommendations. First, that research design explicitly incorporates stakeholder feedback as an iterative process to ‘sense-check’ methods development. Second, that methodological research continues on the ‘how to’ incorporate wellbeing impact and wider societal outcomes within an economic framework. Third, to have more applied examples of economic evaluations that are ‘pushing the boundaries’ using societal perspectives, thus sharing learning by doing.

## References

- Brouwer, W. (2018). The inclusion of spillover effects in economic evaluations: not an optional extra. *Pharmacoeconomics*, <https://doi.org/10.1007/s40273-018-0730-6>. doi:10.1007/s40273-018-0730-6
- Cantrill, J., Sibbald, B., & Buetow, S. (1996). The Delphi and nominal group techniques in health services research. *International Journal of Pharmacy Practice*, *1*, 67-71.
- Coast, J. (2004). Is economic evaluation in touch with society's values? *British Medical Journal*, *329*(7476), 1233-1236.
- Cookson, R., Mirelman, A., Griffin, S., Asaria, M., Dawkins, B., Norheim, O., . . . Culyer, A. (2017). Using cost-effectiveness analysis to address health equity concerns. *Value in Health*, *20*(2), 206-212. doi:10.1016/j.val.2016.11.027
- Dalkey, N. (1969). *The Delphi Method: An experimental study of group opinion*. Santa Monica: Rand Corp.,
- Edwards, R., Charles, J., & Lloyd-Williams, H. (2013). Public health economics: a systematic review of guidance for the economic evaluation of public health interventions and discussion of key methodological issues. *BMC Public Health*, *13*(1001).
- Evans, C., & Crawford, B. (2000). Expert judgement in pharmacoeconomic studies. *Pharmacoeconomics*, *17*(6), 545-553.
- Ferguson, B. (2016). Investing in prevention: counting the cost. Retrieved from <https://publichealthmatters.blog.gov.uk/2016/03/07/investing-in-prevention-counting-the-cost/>
- Frew, E., & Breheny, K. (2019). Health economic methods for public health resource allocation: a qualitative interview study of decision makers from an English local authority. *Health Economics, Policy and Law*, *in press*.
- Frew, E., Ng, S., Coast, J., Hollingsworth, B., & Smith, R. (2018). How can economics help tackle obesity? *Obesity*, *26*(7), 1112-1113.
- Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science & Medicine*, *62*, 1891-1901.
- Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008-1015.
- Hill, S., Vale, L., Hunter, D., Henderson, E., & Oluboyede, Y. (2017). Economic evaluations of alcohol prevention interventions: Is evidence sufficient? A review of methodological challenges. *Health Policy*, *121*, 1249-1262.
- Hunter, D., Marks, L., Brown, J., Scalabrini, S., Salway, S., Vale, L., . . . Payne, N. (2016). The potential value of priority-setting methods in public health investment decisions: qualitative findings from three English local authorities. *Critical Public Health*, *26*(5), 578-587.
- Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2005). An experimental study of determinants of the extent of disagreement within clinical guideline development groups. *BMJ Quality & Safety*, *14*(4), 240-245.
- Keeney, S., Hasson, F., & McKenna, H. (2011). *The Delphi technique in nursing and health research*. Oxford: Wiley-Blackwell.

- Kelly, M., McDaid, D., Ludbrook, K., & Powell, J. (2015). *Economic appraisal of public health interventions*. Retrieved from London: [http://www.cawt.com/Site/11/Documents/Publications/Population%20Health/Economics%20of%20Health%20Improvement/Economic appraisal of public health interventions.pdf](http://www.cawt.com/Site/11/Documents/Publications/Population%20Health/Economics%20of%20Health%20Improvement/Economic%20appraisal%20of%20public%20health%20interventions.pdf)
- Lung, T., Muhunthan, J., Laba, T.-L., Shiell, A., Milat, A., & Jan, S. (2017). Making guidelines for economic evaluations relevant to public health in Australia *Australian and New Zealand Journal of Public Health*, *41*, 115-117.
- McCabe, C. (2018). Expanding the scope of costs and benefits for economic evaluations in health: some words of caution. *Pharmacoeconomics*, <https://doi.org/10.1007/s40273-018-0729-z>.
- McIntosh, E., & Tudor Edwards, R. (2018). *Applied Health Economics for Public Health Practice and Research*. Oxford University Press: Oxford.
- National Institute for Health and Care Excellence (NICE). (2014). *Developing NICE guidelines*. Retrieved from London: <https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf>
- Netton, A., Burge, P., & Malley, J. (2012). Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment*, *16*(16), 1-165.
- Oxman, A., Lavis, J., Lewin, S., & Fretheim, A. (2009). SUPPORT Tools for evidence-informed Policymaking (STP) 1: What is evidence-informed policymaking? *Health Research Policy and Systems*, *7 Suppl 1*(S1).
- Payne, K., McAllister, M., & Davies, L. (2012). Valuing the economic benefits of complex interventions when maximising health is not sufficient. *Health Economics*, *22*(3), 258-271.
- Polit, D., & Hungler, B. (1997). *Essentials of Nursing Research: Methods, Appraisal and Utilisation*. New York: Lippincott.
- Raine, R., Nic a Bhaird, C., Xanthopoulou, P., Wallace, I., Ardron, D., Harris, M., . . . Livingston, G. (2015). Use of a formal consensus development technique to produce recommendations for improving the effectiveness of adult mental health multidisciplinary team meetings. *BMC Psychiatry*, *15*, 143. doi:10.1186/s12888-015-0534-6
- Sanders, G., Newmann, P., Basu, A., Brock, D., Feeny, D., Krahn, M., & et al. (2016). Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analysis. *JAMA*, *316*(1093). doi:10.1001/jama.2016.12195
- Taylor, R., Feltbower, R., Aslam, N., Rosalind, R., Whelan, J., & Gibson, F. (2016). Modified international e-Delphi survey to define healthcare professional competencies for working with teenagers and young adults with cancer. *BMJ Open*, *6*(e011361). doi:10.1136/bmjopen-2016-011361
- UK Government. (2018). Local authority revenue expenditure and financing. Retrieved from <https://www.gov.uk/government/collections/local-authority-revenue-expenditure-and-financing>
- Walker, A., & Selfe, J. (1996). The Delphi technique: a useful tool for the allied health researcher. *British Journal of Therapy and Rehabilitation*, *3*, 677-680.
- Weatherly, H., Drummond, M., Claxton, K., Cookson, R., Ferguson, B., Godfrey, C., . . . Sowden, A. (2009). Methods for assessing the cost effectiveness of public health interventions: key challenges and recommendations. *Health Policy*, *93*, 85-92.

Whitehead, S., & Shehzad, A. (2010). Health outcomes in economic evaluation: the QALYs and utilities. *British Medical Bulletin*, 96(1), 5-21.

Willmott, M., Womack, J., Hollingworth, W., & Campbell, R. (2015). Making the case for investment in public health: experiences of Directors of Public Health in English local government. *Journal of Public Health*. doi:10.1093/pubmed/fdv035

<b>Table 1: Characteristics of decision makers participating in Delphi survey and by Round</b>				
<b>Characteristics</b>	<b>R1</b>	<b>R2</b>	<b>Characteristics</b>	<b>R1</b>
	n=66	n=29		n=66
Current role n (%):			Health economics training n(%):	
Commissioner	5 (8)	2 (7)	MSc (equivalent)	3 (5)
PH director	14 (21)	9 (31)	Health economics module	39 (64)
PH lead	21 (32)	6 (21)	3-day short course	8 (13)
PH team member	16 (24)	8 (26)	1-day short course	18 (30)
Elected member	1(1)	4 (14)	None	6 (10)
Other	9 (14)		Other	5 (8)
Years of work experience n(%):			Highest qualification n(%):	
0-2	17 (26)	5 (17)	PhD	11 (18)
3-5	22 (33)	12 (42)	MSc/MPH	43 (72)
6-10	11 (17)	3 (10)	BA	4 (7)
>10	16 (24)	9 (31)	Higher national diploma	2 (3)

<b>Table 2 Features of economic evaluation considered most important* by Delphi panel</b>	<b>Median</b>	<b>MADM</b>
<b>What economic evidence is considered <i>most</i> important when making:</b>		
<b>Public health investment decisions:</b>		
Costs	1	0.21
Wellbeing outcomes	1	0.39
Health related outcomes	1	0.28
Political pressures	2	0.5
Colleagues' experience	2	0.46
Own experience	2	0.60
Productivity	2	0.42
Media	3	0.67
Education	3	0.64
<b>Public health disinvestment decisions:</b>		
Costs	1	0.10
Wellbeing outcomes	1	0.5
Health-related outcomes	1	0.35
Political pressures	2	0.64
Media reaction	2	0.53
Colleagues' experience	2	0.71
Productivity outcomes	2	0.5
Own experience	3	0.78
Education	3	0.46
<b>Statements relevant to methods for economic evaluation</b>		
<b>Decision-making criteria</b>		
Economic evidence should focus equally on maximizing outcomes and minimizing inequality	2	0.57
Economic evidence should focus on both but with more weight given to maximizing outcomes	3	0.42
Economic evidence should focus on both but with more weight given to minimizing inequality	3	0.69
Economic evidence should focus exclusively on maximizing outcomes	4	0.53
Economic evidence should focus exclusively on minimizing inequalities	4	0.53
<b>Generalizability:</b>		
Economic evidence should be generalizable but take account of local evidence	1	0.53
Economic evidence should take account of local context	2	0.38
Economic evaluation should be generalizable to lots of different settings	2	0.42
<b>Time period to be considered:</b>		

Economic evaluation should consider a range of both short-term and long-term costs and effects	1	0.5
Economic evaluation should consider a time period that is of sufficient length to capture all costs and benefits	2	0.46
Economic evaluation should consider a time period that accounts for a persons' lifetime	2	0.42
Economic evaluation should consider a time period that reflects annual budget cycles	3	0.69
Economic evaluation should consider a time period that reflects political cycles	3	0.61
<b>What costs to consider:</b>		
Economic evidence should take account of all the future costs required to maintain outcome effects	1	0.57
Economic evidence should take account of all set up costs required	1.5	0.57
Economic evidence should present a breakdown of costs for each sector e.g. housing, education	2	0.42
Economic evidence should present costs by population subgroup e.g. more deprived groups	2	0.38
<b>What outcomes to consider:</b>		
Economic evidence should use multiple outcomes to capture health, wellbeing and broader outcomes	1	0.65
Economic evidence should use money to represent both costs and outcomes	2	0.65
Economic evidence should capture a broader set of outcomes beyond health or wellbeing	2	0.61
Economic evidence should use a single outcome that captures length of life and quality of life e.g. QALY's	3	0.61
Economic evidence should use a single outcome that captures wellbeing gain	3	0.42
<b>Consideration of budgets:</b>		
Economic evidence should take account of multisectoral budgets available	2	0.46
<b>Distributional concerns:</b>		
Economic evaluation should formally weight outcomes by population subgroup	2	0.46
<b>Dissemination:</b>		
To best impact policy, economic evidence should be communicated via local reports	2	0.19
To best impact policy, economic evidence should be communicated via national reports	2	0.5
To best impact policy, economic evidence should be communicated in an accessible form using infographics and stories	2	0.61
To best impact policy, economic evidence should be stored in a central repository	2	0.65
To best impact policy, economic evidence should be communicated via 4-page reports	2	0.46
To best impact policy, economic evidence should be communicated via 2-10-25 page reports	2	0.5
To best impact policy, economic evidence should be communicated via peer-reviewed journal articles	3	0.84

\*Items were rated on a Likert scale from 1= strongly agree to 5=strongly disagree; median  $\leq$  2 indicated high agreement. \*\*MADM, mean absolute deviation from the median

