How should we set consumption thresholds for low risk drinking guidelines? Achieving objectivity and transparency using evidence, expert judgement and pragmatism.

John Holmes[1], Colin Angus[1], Petra S Meier[1], Penny Buykx[1], Alan Brennan[1].

Sheffield Alcohol Research Group, School of Health and Related Research (ScHARR), University of Sheffield, 30 Regent Street, Sheffield, S1 4DA, UK

Running head: How should we set drinking guidelines?

Word count: 4,130

Declaration of competing interests: JH and PM were advisors to (and were previously members of) the UK Chief Medical Officers' Guidelines Development Group. Public Health England commissioned all of the authors to provide two epidemiological modelling reports that informed development of the new UK lower risk drinking guidelines.

Abstract

Most high-income nations issue guidelines on low risk drinking to inform individuals' decisions about alcohol consumption. However, leading scientists have criticised the processes for setting the consumption thresholds within these guidelines for a lack of objectivity and transparency. This paper examines how guideline developers should respond to such criticisms and focuses particularly on the balance between epidemiological evidence, expert judgement and pragmatic considerations. Although primarily concerned with alcohol, our discussion is also relevant to those developing guidelines for other health-related behaviours. We make eight recommendations across three areas. First, recommendations on the use of epidemiological evidence: (1) Guideline developers should assess whether the available epidemiological evidence is most appropriately communicated as population-level messages (e.g. suggesting reduced drinking benefits populations rather than individuals); (2) Research funders should prioritise commissioning studies on the acceptability of different alcohol-related risks (e.g. mortality, morbidity, harms to others) to the public and other stakeholders; (3) Guideline developers should request and consider statistical analyses of epidemiological uncertainty. Second, recommendations to improve objectivity and transparency when translating epidemiological evidence into guidelines: (4) Guideline developers should specify and publish their analytical framework to promote clear, consistent and coherent judgements; (5) Guideline developers' decision-making should be supported by numerical and visual techniques which also increase the transparency of judgements to stakeholders. Third, recommendations relating to the diverse use of guidelines: (6) Guideline developers and their commissioners should give meaningful attention to how guidelines are used in settings such as advocacy, health promotion, clinical practice and wider health debates, as well as in risk communication; (7) Guideline developers should make evidence-based judgements that balance epidemiological and pragmatic concerns to maximise the communicability, credibility and general effectiveness of guidelines; (8) As with scientific judgements, pragmatic judgements should be reported transparently.

Summary statement: *Leading scientists have criticised the objectivity and transparency of processes used when setting consumption thresholds for low risk drinking guidelines. Improving epidemiological analysis techniques, the deployment of expert judgement, the incorporation of pragmatic concerns and the balance between these can help to address this problem.*

Introduction

Most high-income countries issue guidelines on low risk drinking to inform individuals' decisions about alcohol consumption [1, 2]. For example, the UK Chief Medical Officers' guideline states that "To keep health risks from alcohol to a low level, it is safest not to drink more than 14 units a week on a regular basis. If you regularly drink as much as 14 units per week, it's best to spread your drinking evenly over three or more days" [3]. An expert committee usually recommends these guidelines after a review of epidemiological evidence; however, there are no standardised methods for committees to follow. This is particularly true when setting the consumption threshold (i.e. the 14 units a week in the above example), which is usually the most visible and contested element of the guidelines [1, 4]. Therefore, this paper aims to examine and make recommendations on the epidemiological and practical challenges faced by guideline developers when setting such thresholds.

The paper comprises five sections. The first describes contemporary drinking guidelines and their development, giving particular attention to the recent use of epidemiological modelling to inform selection of the consumption thresholds [5, 6]. The second section critically examines such modelling and identifies conceptual and methodological priorities for research and practice. The third section move away from epidemiology to assess how guideline developers should deploy expert judgement and make this transparent when translating evidence into practice and, in the fourth section, considers how guideline developers should respond to the practical challenges arising from society's diverse uses of drinking guidelines. The final section summarises our recommendations, discusses the tensions between evidence, judgement and pragmatism, and presents our concluding comments. Throughout the paper, we draw on our experience of providing epidemiological modelling to the 2016 UK Guideline Development Group and our understanding of other guideline development processes as reported by those involved. While drinking guidelines often discuss a diverse set of behaviours, populations and practices (e.g. binge drinking, youth drinking, drinking in the workplace), we focus on the setting of thresholds for daily or weekly alcohol consumption as they are the principal focus of scientific and public debate.

Contemporary drinking guidelines and their development

Drinking guidelines have proliferated internationally since the 1980s [1]; however, the drinkers and types of drinking targeted and the consumption thresholds set differ between countries and over time [2, 7, 8]. This geographic and temporal diversity may be a positive reflection of demographic variation, concern about alcohol-related harm rates, improved epidemiological evidence and distinct drinking cultures. Similarly, it may capture different definitions of a standard drink or intended purposes for guidelines (e.g. Danish guidelines switched focus from a consumption threshold that avoided high risks to one that ensured low risks in 2010) [1]. However, international diversity may also signal practical problems. Leading scientists with experience of developing guidelines have argued that the development processes lack objective methods and transparent documentation [4, 9, 10]. They suggest that guideline developers often "seem to have drawn a deep collective breath

and simply voted for specific cut-off levels" [4: 137], that thresholds for single occasion drinking in particular are "a matter of opaque collective expert opinion" [4: 137] and that decision-making has "an element of mystery" and is "arbitrarily limited to health risk" [9]. Taken together, these accounts suggest there is something unsatisfactory about the way consumption thresholds are set. This problem is not unique to alcohol [11-13], but it is important as drinking guidelines are often key components of national alcohol policies and attract controversy that can be fuelled by apparent subjectivity and opacity [14, 15].

In this context, we welcome recent efforts to subject the setting of consumption thresholds to scientific scrutiny [1, 16, 17]. In particular, guideline development processes in Australia in 2009, Canada in 2011 and the UK in 2016 adopted novel methods that enhanced objectivity and transparency when setting their thresholds [5, 18, 19]. In each case, guideline developers compared one of two *a priori* definitions of 'low risk' against a synthesis of epidemiological evidence to inform selection of the final threshold. We refer to these low risk definitions as the Canadian and Australian approaches. The Canadian approach focuses on relative risks and defines low risk drinking as the consumption level where the all-cause mortality risk is equal to that of abstainers (Figure 1a) [20]. In contrast, the Australian approach focuses on absolute risks and attempts to identify the level of alcohol-related risk that is 'acceptable' to the public. After a review of risk governance literature and evidence from other behaviours, this acceptability threshold is set at the consumption level corresponding to a 1% lifetime risk of alcohol-attributable death [4, 19] (Figure 1b). The Australian approach has proved particularly influential, informing the 2010 Danish, 2016 UK and 2017 French guidelines, as well as the pan-European Reducing Alcohol Related Harm (RARHA) project, which used the 1% threshold to suggest guideline consumption thresholds for seven European countries [1, 21-23].

[Figures 1a and 1b about here]

Developing the use of epidemiological evidence when setting consumption thresholds

The relative merits of the Canadian and Australian approaches (hereafter the exemplar approaches) are debated elsewhere [4, 9, 17, 20, 24], but the epidemiological modelling that underpins them requires critical examination. It has considerable strengths including tailoring the models to local populations by synthesising international evidence on alcohol-related health risks with national data on alcohol consumption and health outcomes [5, 9]. This means the resulting evidence provides nationally-specific estimates of alcohol-related health risks. It can also indicate the appropriateness of alternative consumption thresholds, aid comparison of risks across population subgroups and inform judgements on how new evidence might change previous decisions. However, there are important limitations and we use three key questions below to illustrate these, consider their implications and make recommendations for further research.

Question 1: A low risk for whom?

The rationale for the exemplar approaches implies a given individual drinking at the consumption threshold faces a specified lifetime alcohol-attributable mortality risk. For example, the UK guidelines suggest an individual drinking 14 units per week will accumulate across their lifetime an approximately 1% risk of dying due to alcohol [25: 3]. However, such claims are typically derived from an estimate of the average risk for the population (or for males and females separately). This

assumes implicitly that the population's average risk is equal to each individual's risk. Although commonplace, this assumption is problematic for at least three reasons. First, the population's average risk aggregates many individual-level risks that vary substantially by age, socioeconomic status, genetic profile, wider health-related behaviours and experiences of inequalities [23, 26, 27]. For example, drinkers of lower socioeconomic status incur greater risks than counterparts of higher socioeconomic status for each unit of alcohol consumed [26]. Second, *differences* in the average risks associated with lighter and heavier drinking populations do not necessarily equate to *changes* in risk for individuals who move between these consumption groups, in part because previous drinking will continue to influence each individual's health outcomes [28]. Third, the population-average risk is a synthesis of risks for diverse health outcomes (e.g. heart disease, cancer, injuries) which have differently shaped risk relationships with alcohol consumption and are more or less relevant to different sociodemographic groups [29]. This means that an individual's characteristics will not only determine the level of risk they face, but also the type of risk and how that risk changes in line with their alcohol consumption. For these reasons, the exemplar approaches may provide a population-level rationale for setting the consumption threshold at a particular level, but they do not provide a robust explanation for why that threshold is appropriate for any given individual.

We suggest guideline developers might consider two responses to this problem. First, health authorities may wish to produce more personalised guidance on alcohol consumption, akin to the cardiovascular QRISK score (which excludes alcohol use) [30] or equivalents within popular health monitoring technologies [31]. We do not favour this response at this stage in the development of the evidence base because the capacity of epidemiology to identify who will benefit from behavioural change is disputed [32, 33], the associated evidence is inherently imprecise (see below) and a meta-analysis of studies evaluating the effectiveness of communicating personalised genetic risks did not change behaviour or behavioural motivations [34]. Ethical challenges may also arise including the risk of blaming individuals who are subject to structural risks factors for their failure to achieve good health [35-37] and the need to decide whether increased alcohol-related health risks for those in lower socioeconomic groups mean the poor should be advised to drink less than the wealthy [26, 27].

A second response is to design population-level risk messages. Rather than suggesting that individuals who drink at the consumption threshold face a 1% lifetime risk of dying due to alcohol, a population message might read, 'If everyone drinks within the guideline consumption level, no more than 1 in 100 will die due to alcohol.' The use and effectiveness of population-level messages is under-researched and they require evaluation to assess whether they have sufficient personal salience to affect behaviour. Nonetheless, they do describe the evidence provided by the exemplar approaches more accurately than individualised counterparts and, in line with the prevention paradox [38], they avoid implying that any given individual will benefit meaningfully from small reductions in risk.

Question 2: How much risk is acceptable?

Consideration of risk acceptability within the Australian approach usefully moves beyond epidemiology and addresses people's interactions with risk. However, the acceptability threshold has limited evidential support as no study has examined the specific acceptability of alcohol-related risks. Instead, alcohol researchers have drawn heavily on a 1969 analysis by Starr. Starr used data

on fatalities from selected activities, alongside the time and money spent on them, to conclude that the public will accept voluntary risks from behaviours such as drinking that are 1,000 times greater than involuntary risks, for which acceptability standards can be found in environmental hazard regulations [39, 40]. Yet, Starr's work has a number of limitations. The author viewed his empirical analysis as 'exploratory' and his conceptualisation of risk as incomplete [39, 41]. Others have since identified dimensions beyond voluntariness that influence risk acceptability, including the risk's potential to cause immediate, catastrophic or severe harm, whether it evokes feelings of dread, its perceived controllability or newness, and associated public or scientific knowledge [42]. Also problematic is Starr's reliance on revealed preferences (i.e. observed behaviours), which equate risk-taking to risk acceptance, assume that observed behaviours do not result from unobserved constraints (e.g. family or work commitments) and presume that the observer values risks and benefits similarly to the observed [41]. Further limitations of the Australian approach to assessing the acceptability of alcohol-related risks include focusing only on mortality as an outcome, rather than on morbidity, dependence, negative experiences or harming others [43], and a failure to weigh the risks of drinking against its benefits, particularly when comparing the acceptability of alcohol-related risks to those associated with other behaviours

Given these limitations, research priorities to improve understanding of alcohol-related risk acceptability include using surveys to identify how the acceptability of different alcohol-related outcomes varies on the above dimensions of risk acceptability, using discrete choice experiments to identify risk acceptability thresholds for alcohol-related outcomes and assessing how these are affected by knowledge of the consumption levels and patterns associated with those thresholds, and using qualitative methods to understand how risk acceptability informs drinking practices. Where feasible, each of the above should attend to variation in risk acceptability across key population subgroups.

Question 3: How certain can we be?

The limitations of alcohol epidemiology mean that risk estimates involve a significant degree of imprecision and the exemplar approaches do not engage with this. Example limitations include the substantial underestimation of alcohol consumption within epidemiological surveys, which leads to overestimation of risks for a given level of alcohol consumption [44], and commonplace biases in observational studies (e.g. in sample selection, definition of abstainers and controls for confounding) which are widely discussed in relation to the putative cardioprotective effects of moderate drinking, but apply also to many other health conditions to an unknown degree and in potentially differing directions [45, 46]. Other uncertainties arise from weak measurement of lifetime drinking trajectories, patterns and practices [28, 47, 48] and necessary assumptions embedded within epidemiological models (e.g. regarding the relationships between alcohol consumption and risks of injuries[5]). In practice, these uncertainties mean the consumption level corresponding to a given definition of 'low risk' may be identifiable only as a range of plausible values.

Guideline developers should request sensitivity analyses using alternative valid assumptions about the epidemiological evidence to quantify the extent of uncertainty. Figure 2 summarises selected results from the sensitivity analyses that we provided to the UK Guideline Development Group [5]. It shows that the consumption threshold implied by the Australian approach for men varies between 21g and 159g per week depending on the assumptions made. The Group also considered further

unquantified uncertainty relating to underestimation of alcohol consumption. We envisage that future guideline developers will face similarly imprecise evidence as, despite methodological advances [e.g. 44, 47, 49, 50-53], many of the above limitations are inherent to epidemiology [5, 54, 55].

[Figure 2 about here]

The three discussions above demonstrate that while the exemplar approaches advance methods for setting consumption thresholds, they can only take us so far. An imperfect evidence base means that conceptually clear-cut methods still only lead to a range of broadly appropriate values from which guideline developers can select their preferred consumption threshold. Although the supporting evidence is more robust, that range is, arguably, not markedly different to those considered under previous, simpler methods. Two RAHRA reports reflect this gap between theory and practice. The first stresses, "The chosen methodology does not assume any expert judgements, once the acceptable risk thresholds are set. [The guideline] is based solely on evidence" [6: 15]. The second concedes, "Such summarising of quantitative data does not replace expert judgement but provides a transparent approach for justifying experts' choices" [1: 7]. In the remainder of this paper, we discuss how to make the expert judgements necessary to select a consumption threshold from the range of appropriate values and how to make those judgements transparent.

Improving the objectivity and transparency of judgements on epidemiological evidence

Criticisms of the way guideline consumption thresholds are set do not target the necessary presence of expert judgements per se, but focus instead on the subjectivity and opacity of those judgements [4]. Subjectivity and opacity can manifest in several ways. For example, a lack of clarity, consistency or coherence may emerge when making multiple judgements in an unstructured manner over a lengthy process or report. This can mean important decisions are unexplained, comparisons are distorted by switching between relative and absolute risk metrics, priority is given to condition-specific risks rather than overall risks, and uncertainties are glossed over inappropriately or given undue prominence [6, 56, 57]. Below, we draw on previous guidance to recommend two linked ways of improving the execution and communication of expert judgement [12, 58, 59].

First, we recommend that guideline developers publish and consult upon an explicit analytical framework that details the questions, evidence, metrics, procedures and considerations in play across all of their judgements. This should occur at the start of the process and the final framework should record and explain any revisions. Developing an internationally-standardised framework would be a useful future exercise but, at present, we suggest including the following: the questions to be answered, the types and quality of evidence to be used, the outcomes of interest (e.g. mortality, QALYs, harms to others, inequalities), any new analyses to be undertaken and the planned approach to uncertainty (e.g. extrapolation, sensitivity analyses). With regard to the handling of epidemiological evidence, the framework might specify whether the focus is on overall- or condition-specific risks, how to conceptualise the relationship between population- and individual-level risks and which risk metrics to use. [57].

Second, we recommend that guideline developers use systematic and transparent methods to reach consensus on each judgement. Detailed guidance on consensus methods is available elsewhere [60-65] and we limit our discussion to highlighting numerical and visual approaches, noting that these

are not mutually exclusive. Numerical approaches offer excellent transparency by assigning scores to alternative decision options, as in some forms of multi-criteria decision analysis [65]. For example, where sensitivity analyses point to a range of possible consumption thresholds, developers can score each analysis on metrics such as the extent to which it changes their prior position and the plausibility of any assumptions made. They can then use these scores to calculate a weighted average score for each of the candidate thresholds. An alternative numerical approach, as described by Dawson et al., is to assess candidate consumption thresholds for their specificity and sensitivity when predicting harmful alcohol-related outcomes [16].

Visual approaches can communicate accessibly and rapidly the nature, coherence and consistency of the key judgements made. We have developed a visualisation in Figure 3 that demonstrates this. It draws on questions that arose when setting the UK consumption threshold, although it is necessarily hypothetical as we were not involved in the final decision-making. The boxes in Figure 3 describe key evidence and considerations, the x-axis shows the extent to which each consideration suggested a higher or lower threshold than was ultimately selected and the y-axis shows the weight given to each consideration by developers when setting the threshold. For example, the box labelled 'SA2: No protective effects' is positioned to indicate that sensitivity analyses two suggested the appropriate threshold would be much lower if there were no cardioprotective effects and that this evidence played a substantial role in the final decision (i.e. the chosen threshold would have been much lower if other judgements had not pulled in the opposite direction). In contrast, the underestimation of alcohol consumption within epidemiological surveys points towards a higher threshold but this was not given much weight in reaching the final decision (i.e. the threshold would be in much the same place even if this was the only consideration). The value of the diagram is that it makes visible to the public the key judgements made by guideline developers and the *potential* and *actual* impact of those judgements on the final threshold. If developers exclude considerations, prioritise them in decision-making or find that they have little influence on overall risk estimates, then this is immediately apparent. The diagram can be supported in a layered way with brief explanations and more detailed commentary within supporting documentation.

[Figure 3 about here]

Figure 3 is deliberately unbalanced and suggests an incoherent decision-making process. This is because it omits considerations beyond epidemiological evidence, such as the acceptability, credibility and public interpretation of guidelines. The final section of this paper focuses on these pragmatic concerns (and Figure S1 in the appendix presents a balanced version of the diagram that incorporates them).

Making pragmatic considerations meaningful and visible

Scientific debate on guideline consumption thresholds focuses primarily on epidemiological matters. This aligns with the view that drinking guidelines exist to communicate alcohol-related health risks to the public. Yet, guidelines serve a much wider range of purposes. Practitioners embed them in in behaviour change interventions, health promotion campaigns and clinical practice. Advocates leverage the consumption thresholds in public debate to illustrate levels of excess alcohol consumption and the need for intervention. Policy analysts use the same thresholds as benchmarks for measuring intervention success or when categorising populations. The public and commentators also discuss and satirise guidelines. All of this also means that drinking guidelines inevitably

contribute to definitions and understandings of normative concepts such as responsible, irresponsible and binge drinking. As such, they are not neutral expressions of epidemiological data. They are social objects with a history, a literal and symbolic meaning, and a position within political, scientific and public practices.

It is unclear whether and how guideline developers reflect these realities within their decision-making. Do consumption thresholds need to be credible with the public and other stakeholders? Should public preconceptions, such as those about the excess risk from alcohol faced by women, feature in decision-making? How can consumption thresholds communicate epidemiological uncertainty? Is it better to maintain the *status quo* to avoid guidelines becoming mired in criticism from stakeholders? These questions mainly put guideline developers in a defensive position, adjusting their decisions to limit the adverse consequences for themselves or the public of an epidemiologically-driven process. However, a recent study of the development, communication and interpretation of Danish dietary guidance suggests that pragmatic decision-making can form part of a positive and strategic approach to health promotion [66]. The Danish guidance recommended eating six portions of fruit and vegetables a day but the number 'six' was not based solely on epidemiology. Among other considerations was the closeness between the Danish words for 'six' and 'sex'. This subsequently featured in a highly successful campaign that prioritised pictures of amorous vegetables over communicating epidemiology. Interviews with the public revealed that, while the number six gave the guideline legitimacy and communicated a need for dietary discipline and self-monitoring, they rarely treated as a literal or important target. Instead, people drew on lay health knowledge, interpreted the guideline in the context of their pre-existing health practices and responded in diverse, but often positive, ways. The lesson is that best practice in setting guideline consumption thresholds does not equate solely to robust epidemiological analysis and judgement. Instead, it involves meaningful consideration by guideline developers, from the outset, of the different ways in which end-users including the public, health professionals, advocates and industry will communicate, interpret and put into the practice the guidelines.

To date, there is little research to support such considerations. Surveys suggest that few drinkers use guidelines to monitor their consumption [67, 68] and qualitative studies report that people typically disregard consumption thresholds as they are difficult to accommodate within existing drinking practices, rely on an understanding of units that many drinkers do not have, and align poorly with more common, embodied ways of monitoring consumption [69-71]. Therefore, priorities include further assessment of the credibility and acceptability of guidelines to the public and other stakeholders, exploring how the public accommodate guidelines into their existing health practices and lay understandings of alcohol-related risks and benefits, and examining how guidelines are used by different stakeholders within communications and interventions.

Recommendations

The above discussion considers the role of evidence, expert judgement and pragmatism in developing consumption thresholds for use within drinking guidelines. We argue that while epidemiological evidence should be central to selecting these thresholds, expert judgements remain essential and both must be balanced, deliberately and transparently, against pragmatic considerations. Table 1 summarises our recommendations for achieving that balance.

[Table 1 about here]

We recognise that our recommendations invite criticisms that 'the science' should be paramount when developing drinking guidelines. In our view, arguments for epidemiological purity are unconvincing as presenting a single consumption threshold containing a single number is itself a necessarily pragmatic response to the need to communicate complex and imprecise evidence in a simple and digestible form. More generally, suggestions that policy must reflect science beyond all else are flawed and give insufficient weight to the importance of democratic politics, moral hazard and scientific fallibility [72]. Without rejecting science's valuable contribution, we suggest that incorporating pragmatic decisions may produce guidelines that are more effective and, indeed, this may already be occurring to some degree, for example, when guideline developers elect not to communicate uncertainty. In our view, the pressing need is for researchers to support guidelines developers by facilitating evidence-based pragmatic judgements. Prioritising research that addresses the relevant questions is, therefore, a priority.

For the sake of brevity and focus, we have not discussed a number of important matters related to the development of consumption thresholds. These include the role of drinking patterns and contexts, public misunderstanding of units or standard drinks, the focus on gender over other sociodemographic characteristics, the possibility of setting multiple consumption thresholds to indicate that risks increase gradually and the use of non-epidemiological evidence on the health consequences of alcohol (e.g. clinical trials or mechanistic studies). These omissions are not tacit approvals of the status quo. On the contrary, we encourage debate on each of these points as, in different ways, they are likely to shape the effectiveness of drinking guidelines. Although we do not speak directly to those debates, the need to balance evidence, judgement and pragmatism is likely to be equally as relevant in those discussions as it is in the present debate.

Conclusions

The development of best practice for setting consumption thresholds within drinking guidelines remains a work in progress. Challenges remain relating to the analysis of epidemiological data, the application of expert judgement and the pragmatic consideration of how guidelines may function after they are set. Our eight recommendations can help guideline developers, practitioners, researchers and policy-makers internationally to produce drinking guidelines that are epidemiologically justified, transparently developed, perceived as legitimate by the public, and effective when promoted.

References

1. Broholm K., Galluzzo L., Gandin C., Ghirini S., Ghiselli A., Jones L. et al. Good practice principles for low risk drinking guidelines. Helsinki, Finland: National Institute for Health and Welfare (THL); 2016 accessed
2. Kalinowski A., Humphreys K. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries, Addiction 2016.

3. Department of Health. UK Chief Medical Officers' Low Risk Drinking Guidelines. London: Department of Health; 2016 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/545937/UK_CMOs__report.pdf accessed 14th February 2018.
4. Room R., Rehm J. Clear criteria based on absolute risk: Reforming the basis of guidelines on low-risk drinking, Drug Alcohol Rev 2012: 31: 135-140.
5. Holmes J., Angus C., Buykx P., Ally A., Stone T., Meier P. Mortality and morbidity risks from alcohol consumption in the UK: Analyses using the Sheffield Alcohol Policy Model (v.2.7) to inform the UK Chief Medical Officers' review of the UK lower risk drinking guidelines. Sheffield: ScHARR, University of Sheffield; 2016 https://www.sheffield.ac.uk/polopoly_fs/1.538671!/file/Drinking_Guidelines_Final_Report_Published.pdf accessed 14th February 2018.
6. Rehm J., Gmel G., Probst C., Shield K. D. Lifetime-risk of alcohol-attributable mortality based on different levels of alcohol consumption in seven European countries. Implications for low-risk drinking guidelines. Toronto, ON, Canada: Centre for Addiction and Mental Health; 2015 http://www.camh.ca/en/research/news_and_publications/reports_and_books/Documents/Lifetime%20Risk%20of%20Alcohol-Attributable%20Mortality.pdf accessed 14th January 2015.
7. Scafato E., Galluzzo L., Ghirini S., Gandin C., Martire S. Low risk drinking guidelines in Europe: Overview of RARHA survey results; 2016 http://www.rarha.eu/Resources/Guidelines/Pages/details.aspx?itemId=2&lista=Guidelines&bkUrl=Resources/Guidelines/ accessed 1st December 2017.
8. House of Commons Science and Technology Committee. Alcohol Guidelines, Eleventh Report of Session 2010-12 (HC 1536). London: The Stationery Office; 2012 accessed
9. Rehm J., Room R., Taylor B. Method for moderation: measuring lifetime risk of alcohol-attributable mortality as a basis for drinking guidelines, Int J Methods Psychiatr Res 2008: 17: 141-151.
10. Smith R. A row plucked out of the air: The Guardian; 2007.
11. The National Academies of Sciences E. a. M. Optimising the process for establishing the dietary guidelines for Americans: The selection process. Washington, DC: The National Academies Press; 2017 accessed
12. Alonso-Coello P., Schünemann H. J., Moberg J., Brignardello-Petersen R., Akl E. A., Davoli M. et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction, BMJ 2016: 353.
13. Woodruff T. J., Sutton P., The Navigation Guide Work Group. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences, Health Aff (Millwood) 2011: 30: 931-937.
14. The Times. Cut in drink limit ignored advice; 2017.
15. The Times. Anti-drink lobby drew up official safety limits; 2016.
16. Dawson D. A., Smith S. M., Pickering R. P., Grant B. F. An empirical approach to evaluating the validity of alternative low-risk drinking guidelines, Drug Alcohol Rev 2012: 31: 141-150.
17. Stockwell T. I. M., Room R. Constructing and responding to low-risk drinking guidelines: Conceptualisation, evidence and reception, Drug Alcohol Rev 2012: 31: 121-125.
18. Butt P., Beirness D., Stockwell T., Gilksman D., Paradis C. Alcohol and Health in Canada: A summary of evidence and guidelines for low-risk drinking. Ottawa, ON: Canadian Centre on Substance Abuse; 2011 accessed
19. National Health and Medical Research Council. Australian guidelines to reduce health risks from drinking alcohol; 2009 https://www.nhmrc.gov.au/guidelines/publications/ds10 accessed 22nd December 2014.

20. Stockwell T., Butt P., Beirness D., Gliksman L., Paradis C. The basis for Canada's new low-risk drinking guidelines: A relative risk approach to estimating hazardous levels and patterns of alcohol use, Drug Alcohol Rev 2012: 31: 126-134.
21. Department of Health. Alcohol Guidelines Review: Report from the Guidelines development group to the UK Chief Medical Officers; 2016 https://www.gov.uk/government/consultations/health-risks-from-alcohol-new-guidelines accessed 11th May 2016.
22. Santé publique France, Institut National du Cancer. Avis d'experts relatif à l'évolution du discours public en matière de consommation d'alcool en France. Saint-Maurice: Santé publique France; 2017 accessed
23. Shield K. S., Gmel G., Gmel G., Mäkelä P., Probst C., Room R. et al. Lifetime risk of mortality due to different levels of alcohol consumption in seven European countries: implications for low-risk drinking guidelines, Addiction 2017.
24. Dawson D. A. Low risk drinking limits: absolute versus relative risk, Addiction 2009: 104: 1303-1304.
25. Department of Health. UK Chief Medical Officers' Alcohol Guidelines Review: Summary of the proposed new guidelines; 2016 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/489795/summary.pdf accessed 14th February 2018.
26. Mäkelä P., Paljärvi T. Do consequences of a given pattern of drinking vary by socioeconomic status? A mortality and hospitalisation follow-up for alcohol-related causes of the Finnish Drinking Habits Survey, J Epidemiol Community Health 2008: 62: 728-733.
27. Bellis M. A., Hughes K., Nicholls J. The alcohol harm paradox: using a national survey to explore how alcohol may disproportionately impact health in deprived individuals, BMC Public Health 2016: 16.
28. Britton A., Marmot M., Shipley M. J. How does variability in alcohol consumption over time affect the relationship with mortality and coronary heart disease?, Addiction 2010: 105: 639-645.
29. Rehm J., Gmel G. E., Gmel G., Hasan O. S. M., Imtiaz S., Popova S. et al. The relationship between different dimensions of alcohol use and the burden of disease—an update, Addiction 2017: 112: 968-1001.
30. Hippisley-Cox J., Coupland C., Vinogradova Y., Robson J., May M., Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study, BMJ 2007: 335: 136.
31. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies, Critical Public Health 2013: 23: 393-403.
32. Rockhill B. Theorizing About Causes at the Individual Level While Estimating Effects at the Population Level: Implications for Prevention, Epidemiology 2005: 16: 124-129.
33. Wemrell M., Merlo J., Mulinari S., A-C H. Contemporary Epidemiology: A review of critical discussions within the discipline and a call for further dialogue with social theory, Sociology Compass 2016: 10: 153-171.
34. Hollands G. J., French D. P., Griffin S. J., Prevost A. T., Sutton S., King S. et al. The impact of communicating genetic risks of disease on risk-reducing health behaviour: systematic review with meta-analysis, BMJ 2016: 352.
35. Bissell P., Peacock M., Blackburn J., Smith C. The discordant pleasures of everyday eating: Reflections on the social gradient in obesity under neo-liberalism, Soc Sci Med 2016: 159: 14-21.
36. Lupton D. M-health and health promotion: The digital cyborg and surveillance society, Social Theory & Health 2012: 10: 229-244.
37. Ayo N. Understanding health promotion in a neoliberal climate and the making of health conscious citizens, Critical Public Health 2012: 22: 99-105.

38. Rose G. Sick individuals and sick populations, Int J Epidemiol 2001: 30: 427-432.
39. Starr C. Social benefit versus technological risk, Science 1969: 165: 1232-1238.
40. Rehm J., Lachenmeier D. W., Room R. Why does society accept a higher risk for alcohol than for other voluntary or involuntary risks?, BMC Med 2014: 12.
41. Fischhoff B., Kadvany J. Risk: A very short introduction Oxford: Oxford University Press; 2011.
42. Fischhoff B., Slovic P., Lichtenstein S., Read S., Combs B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits, Policy Sciences 1978: 9: 127-152.
43. Rehm J., Patra J. Different guidelines for different countries? On the scientific basis of low-risk drinking guidelines and their implications, Drug Alcohol Rev 2012: 31: 156-161.
44. Meier P. S., Meng Y., Holmes J., Baumberg B., Purshouse R., Hill-MacManus D. et al. Adjusting for unrecorded consumption in survey and per capita sales data: Quantification of impact on gender- and age-specific alcohol attributable fractions for oral and pharyngeal cancers in Great Britain, Alcohol Alcohol 2013: 48: 241-249.
45. Bergmann M. M., Rehm J., Klipstein-Grobusch K., Boeing H., Schütze M., Drogan D. et al. The association of pattern of lifetime alcohol use and cause of death in the European Prospective Investigation into Cancer and Nutrition (EPIC) study, Int J Epidemiol 2013: 42: 1772-1790.
46. Naimi T. S., Stockwell T., Zhao J., Xuan Z., Dangardt F., Saitz R. et al. Selection biases in observational studies affect associations between 'moderate' alcohol consumption and mortality, Addiction 2017: 112: 207-214.
47. Britton A., Ben-Shlomo Y., Benzeval M., Kuh D., Bell S. Life course trajectories of alcohol consumption in the United Kingdom using longitudinal data from nine cohort studies, BMC Med 2015: 13.
48. Ally A., Lovatt M., Meier P., Brennan A., Holmes J. Developing a social practice-based typology of British drinking culture in 2009-2011: Implications for alcohol policy analysis, Addiction 2016: 111: 1568-1579.
49. Holmes M. V., Dale C. E., Zuccolo L., Silverwood R. J., Guo Y., Ye Z. et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data, BMJ 2014: 349.
50. Rehm J., Kehoe T., Gmel G., Stinson F., Grant B., Gmel G. Statistical modeling of volume of alcohol exposure for epidemiological studies of population health: the US example, Popul Health Metr 2010: 8: 3.
51. Parish W. J., Aldridge A., Allaire B., Ekwueme D. U., Poehler D., Guy G. P. et al. A new methodological approach to adjust alcohol exposure distributions to improve the estimation of alcohol-attributable fractions, Addiction 2017: 112: 2053-2063.
52. Rehm J., Shield K. D., Roerecke M., Gmel G. Modelling the impact of alcohol consumption on cardiovascular disease mortality for comparative risk assessments: an overview, BMC Public Health 2016: 16: 363.
53. Cherpitel C. J., Ye Y., Kerr W. C. Risk of past year injury related to hours of exposure to an elevated BAC and average monthly alcohol volume: Data from four national alcohol surveys (2000-2015), Alcohol Clin Exp Res 2017: n/a-n/a.
54. Gmel R., Rehm J. Measuring Alcohol Consumption, Contemporary Drug Problems 2004: 31: 467-540.
55. Roerecke M. On bias in alcohol epidemiology and the search for the perfect study, Addiction 2017: 112: 217-218.
56. Department of Health. New alcohol guidelines show increased risk of cancer; 2016.
57. Spiegelhalter D. Risk and Uncertainty Communication, Annual Review of Statistics and Its Application 2017: 4: 31-60.
58. NICE. Social value judgements: Prinicples for the development of NICE guidance; 2008 https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0089902/ accessed 14th February 2018.

59. NICE. Developing NICE guidelines: the manual; 2014 https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview accessed 14th February 2018.
60. Jones J., Hunter D. Consensus methods for medical and health services research, BMJ : British Medical Journal 1995: 311: 376-380.
61. Fink A., Kosecoff J., Chassin M., Brook B. H. Consensus methods: characteristics and guidelines for us, Am J Public Health 1984: 74: 979-983.
62. Triantaphyllou E. Multi-Criteria Decision Making Methods. Multi-criteria Decision Making Methods: A Comparative Study, Boston, MA: Springer US; 2000, p. 5-21.
63. Abo-Hamad W., Arisha A. Multi-criteria approach using simulation-based balanced scorecard for supporting decisions in health-care facilities: an emergency department case study, Health Systems 2014: 3.
64. Dolan J. Multi-criteria clinical decision support: A primer on the use ofmultiple criteria decision making methods to promote evidence based, patient-centered healthcare, Patient 2010: 3.
65. Department for Communities and Local Government. Multi-criteria analysis: A manual. London: Department for Communities and Local Government; 2009 accessed
66. Kristensen D. B., Jacobsen C. B., Pihl-Thingvad S. Perception and translation of numbers: the case of a health campaign in Denmark, Critical Public Health 2017: 1-12.
67. Stevely A. K., Buykx P., Brown J., Beard E., Michie S., Meier P. S. et al. Exposure to revised drinking guidelines and 'COM-B' determinants of behaviour change: descriptive analysis of a monthly cross-sectional survey in England, BMC Public Health 2018: 18: 251.
68. Furtwängler N., de Visser R. O. Motivation to adhere to unit-based guidelines for alcohol consumption and ability to do so is limited among university students, Drug: Educ Prev Polic 2017: 24: 418-425.
69. Lovatt M., Eadie D., Meier P. S., Li J., Bauld L., Hastings G. et al. Lay epidemiology and the interpretation of low risk drinking guidelines by adults in the United Kingdom, Addiction 2015.
70. Jayne M V., G and Holloway S.L. What use are units? Critical geographies of alcohol policy, Antipode 2012: 44: 828-846.
71. Furtwängler N. A. F., de Visser R. O. University students' beliefs about unit-based guidelines: A qualitative study, J Health Psychol 2017: 22: 1701-1711.
72. Humphreys K., Piot P. Scientific evidence alone is not sufficient basis for health policy, BMJ 2012: 344.

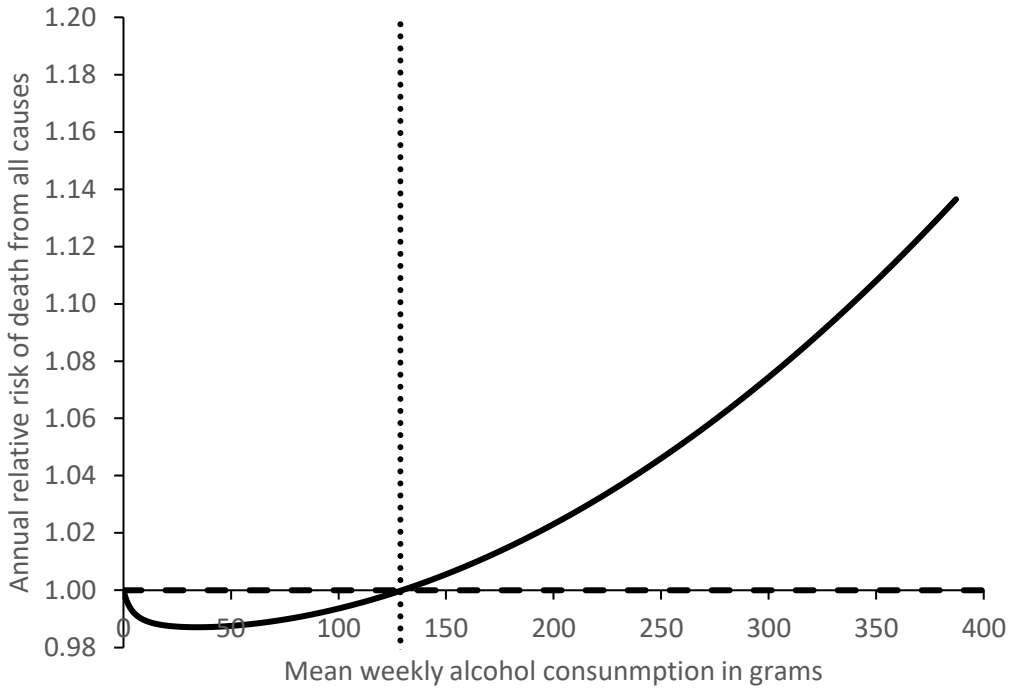

Figure 1a: The Canadian approach to setting drinking guidelines as used in our report to the UK Guideline Development Group [5]

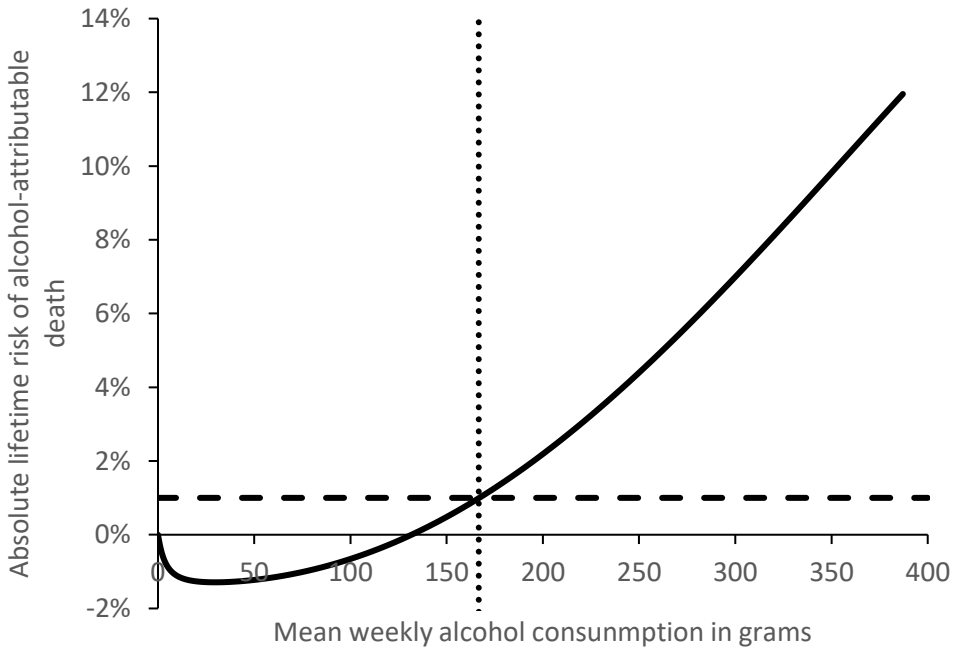

Figure 1b: The Australian approach to setting drinking guidelines as used in our report to the UK Guideline Development Group [5]

Note: The y-axis in Figure 1b describes the estimated percentage of deaths in the modelled year which would be attributable to alcohol if everyone consumed at the level on the x-axis. The putative protective effect of moderate drinking introduces a complication as this curve must either: (a) have negative absolute risks or (b) use the nadir of the curve as the zero absolute risk point and assign a risk of 'alcohol-attributable' death to

abstainers. We present the first option as our analyses for the UK Guideline Development Group required us to compare risk curves for different population groups and different drinking patterns within those groups. This comparison would have been impractical if the zero absolute risk point (i.e. the x-axis) were set at the nadir as each of the curves has a different nadir.

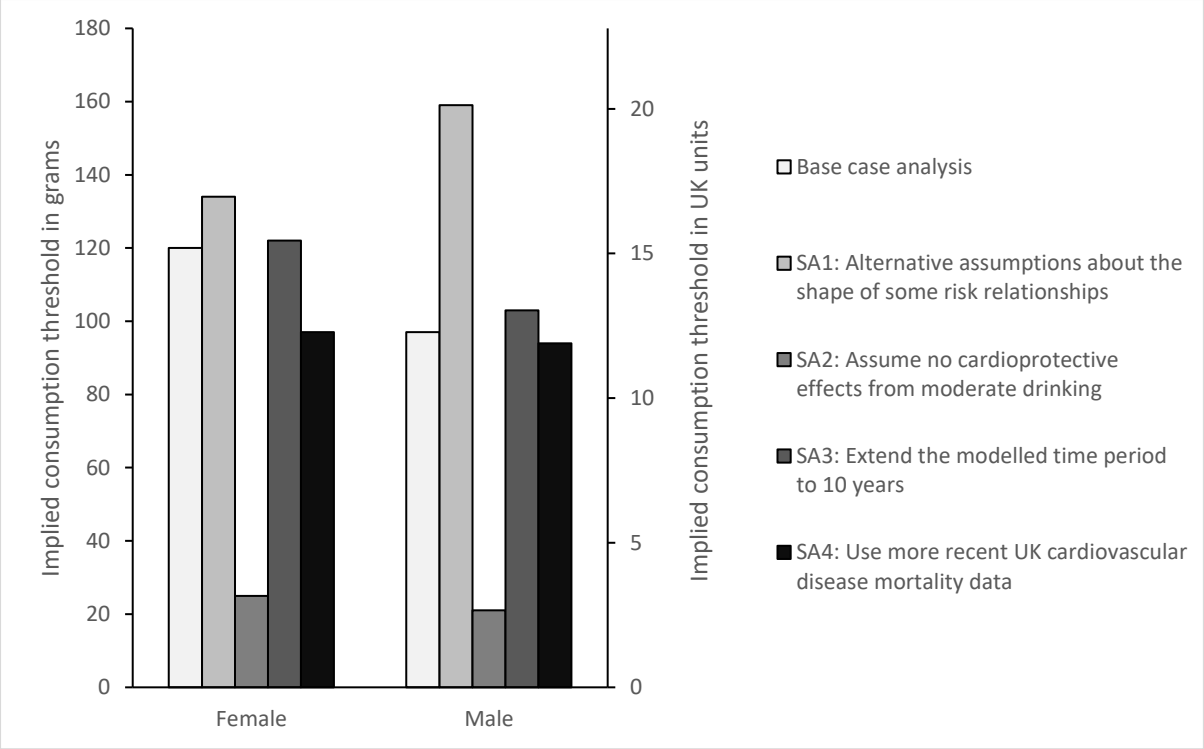

Figure 2: Implied female and male UK guideline thresholds under the Australian approach in the University of Sheffield's base case model and in sensitivity analyses

Note: SA=Sensitivity Analysis. Figures assume that drinkers spread their consumption evenly across four days. For the Canadian approach, other consumption patterns and full details of sensitivity analyses, please see the modelling report [5]. Note that some results imply a higher guideline for females than males. This counter-intuitive finding is due to males facing greater risks of acute harm and, again, is discussed in more detail in the modelling report.

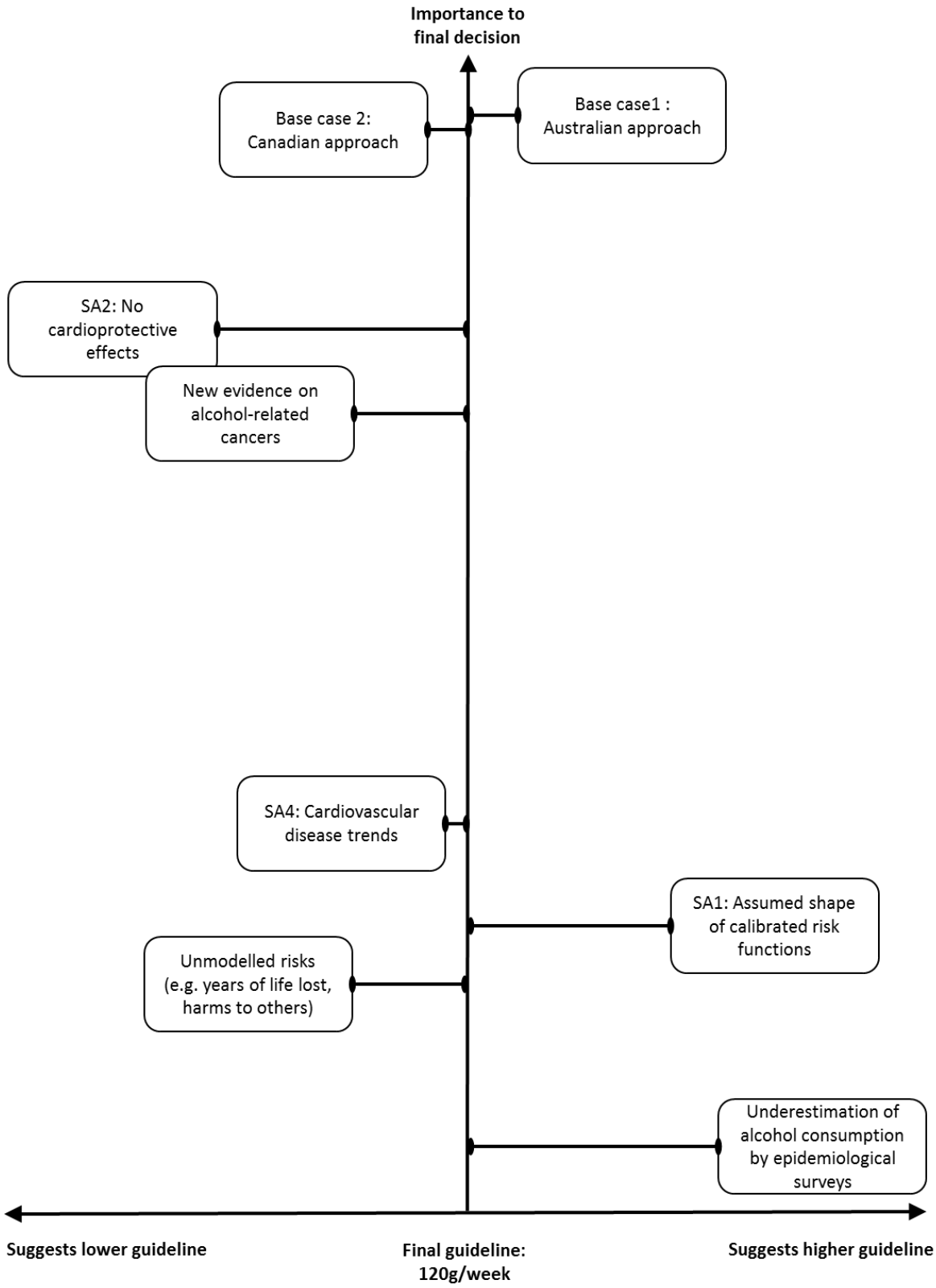

Figure 3: Visual representation of the impact and importance of epidemiological judgements taken in a hypothetical male guideline development process.

Note: SA=Sensitivity Analysis

Table 1: Recommendations and suggested actions to improve the development of drinking guidelines

Recommendation	**Suggested actions**
Developing the use of epidemiological evidence when setting consumption thresholds	
1. Appropriately reflect the relationship between estimates of population-level and individual-level risks.	a. Make transparent how this has been conceptualised within the analytical framework (see Recommendation 4). b. Consider developing population-level health promotion messages.
2. Develop understanding of the acceptability of alcohol-related risks.	a. Use survey research to understand the position of different alcohol-related outcomes on dimensions of risk acceptability [42]. b. Use discrete choice experiments to identify risk acceptability thresholds for alcohol-related outcomes and test how revealing the associated consumption levels and patterns affects these. c. Use qualitative focus groups and/or interviews to understand how risk acceptability informs drinking practices.
3. Ensure appropriate uncertainty analyses	a. Request that epidemiological modellers undertake sensitivity analyses on major areas of uncertainty where feasible. b. Draw on previous uncertainty analyses in this area. c. Where uncertainty is large and important, give this appropriate prominence within decision-making.
Improving the objectivity and transparency of judgements on epidemiological evidence	
4. Ensure judgements on the evidence are clear, consistent and coherent.	a. Develop at an early stage an analytical framework to be used across the guideline development process, potentially detailing the questions, evidence, metrics, procedures and considerations in play across all judgements. b. Publish and consult upon this framework prior to key decision-making.
5. Make judgements transparent and accessible.	b. Use numerical techniques such as multi-criteria decision-making or visual techniques (e.g. Figure 3) to communicate the importance and impact of judgements.
Making pragmatic considerations meaningful and visible	

6. Give attention to purposes of drinking guidelines beyond communicating risk information.	a. Assess the credibility and acceptability of alternative guidelines to relevant stakeholders and examine the factors which affect this. b. Explore how the public accommodate guidelines into lay understandings of risks and benefits as well as existing health practices. c. Examine how stakeholders in multiple sectors translate and communicate guidelines to the public. d. Evaluate and experiment with guidelines and associated messages to assess their effectiveness across their full range of purposes.
7. Make strategic and positive judgements which allow guidelines to function effectively.	a. Draw on research evidence (see Recommendation 6) to select guidelines which balance epidemiological concerns with the concerns of communicators and others who must make the guidelines work in practice.
8. Make judgements on these wider considerations transparent and accessible.	a. Adapt the methods outlined in Recommendation 5 (e.g. Figure 6).