

Response to the appendix to Duffy & Snowdon's report



1. Introduction

- 1.1 In the appendix to their report, Duffy & Snowdon list a series of critiques of the methodology used in the Scottish adaptation of the Sheffield Alcohol Policy Model (SAPM). We respond to each of these points below, however, we would like to make five main points before going into further detail.
- 1.2 Duffy & Snowdon do not identify any evidence (whether datasets, single studies, systematic reviews, or statistical “meta-analyses” which synthesise evidence from several independent studies) that has been omitted from consideration, nor do they make any recommendation for inclusion of alternative evidence within our modelling.
- 1.3 At no point in the appendix do Duffy & Snowdon question the direction of the estimates made. In other words, they are not questioning the conclusion that minimum pricing will lead to reductions in consumption or reductions in rates of harms. The considerable evidence to support those conclusions is described in the main body of this response.
- 1.4 Many of the points made by Duffy & Snowdon are essentially technically-worded commentary on the general limitations of epidemiological evidence. Researchers in the field are fully aware of these limitations and Duffy & Snowdon do not raise any issues which have not previously been considered within our team and with external peer reviewers, researchers, policy experts and other stakeholders.
- 1.5 Although there are always limitations to any piece of scientific research, we are clear that we have followed best practice throughout our work. We are also scrupulous in our documentation and communication of the limitations of our work. For example, our approach to linking rates and risks of harm and levels of alcohol consumption is in line with that used by the world's leading epidemiologists in the WHO's global burden of disease studies [1]. Similarly, we have attempted to incorporate evidence which would be considered 'gold standard' in the epidemiological methods literature wherever possible. Throughout development and dissemination of the model, we have sought advice from leading economists, epidemiologists, policy analysts and other practitioners within our field and have acted upon their suggestions as appropriate. Ensuring we have followed best practice has enabled us to present and publish our work in the most prestigious scientific outlets such as a world leading medical journal, *The Lancet*. This would not have been possible if the rigorous scrutiny that accompanies such dissemination had identified significant methodological flaws in our work. To our knowledge, Duffy & Snowdon have not published any of their points of critique in a peer-reviewed outlet during the four years since our work was first introduced into the public domain. Thus the points made by their critique

have not been subjected to the scientific scrutiny and right of reply that an academic journal would provide.

- 1.6 As with the main body of their critique, the appendix conveys a sense that Duffy & Snowdon reject the use of mathematical models to estimate the potential impact of policy options. We disagree with this and have set out our position earlier in this response.

2. Structure

- 2.1 This response is structured to largely address the critiques raised by Duffy & Snowdon in the order they appear in the appendix of their report and begins by responding to the points around the elasticities, survey data and heavy episodic drinking model, moves on to discuss the epidemiological evidence and sensitivity analyses before briefly dealing with the reiterated points in Duffy & Snowdon's final section.
- 2.2 Readers will note that we expend considerably more words responding to Duffy & Snowdon's critiques than they use outlining them. We believe this is necessary and useful as the briefness of the points made by our critics often fails to do more than scratch the surface of complex scientific issues which we have grappled with in producing our reports and the extensive and rigorous analyses we have conducted to provide the best solutions to arising problems. Therefore, we have attempted to explain the questions at hand in more depth and present a full account of our considerations and why we feel the comments by Duffy & Snowdon do not significantly undermine our work. We hope this provides a balanced and helpful response to the criticisms raised and also furthers understanding of our modelling, the scientific process which underpins it and the challenges and limitations inherent to such work.

3. Economics/consumption data

3.1 *Elasticities*

- 3.1.1 Duffy & Snowdon briefly comment on the English pricing data and price elasticities used within the Scottish adaptation of our model and query whether these English inputs are applicable to a Scottish population. We have been completely transparent that there is insufficient data collected in Scotland to provide quantified estimates of price elasticities for alcohol using the same methods that have been used in our reports for England. Although it is always desirable to have robust evidence from the time period and location in question, this is frequently unavailable. In such circumstances it is standard practice to use the best available international evidence or evidence from a closely-related context. As robust data were available from England in 2001/2-2005/6, we took exactly that approach. Whilst there may be differences in patterns of purchasing, we were also mindful that large elements of the off-trade supply are similar between the two jurisdictions (e.g. through national supermarket chain sales). Recognising that using our basecase price elasticities could be

considered imperfect, we conducted many further scenario analyses using different possible elasticities from UK [2] and also international evidence which includes syntheses of over 100 studies by Wagenaar et al. [3] and Gallet et al. [4]. All of these sensitivity analyses are fully documented in our reports. Duffy & Snowdon make no further specific suggestions for elasticity analyses.

3.2 *Survey data*

- 3.2.1 Duffy & Snowdon make several points arguing that the methods used to collect the underlying survey data in the model contains various biases which may affect the accuracy of the data. These include chance variation in who is selected to participate in the survey (sampling variability), particular types of people being more likely to refuse to participate (non-response bias), deliberate misreporting of behaviour (response bias) and biases arising from the purchasing diary data, used to measure spending on alcohol, being collected only within a two-week period (as these may be atypical purchasing weeks for some and non-purchasing weeks for infrequent purchasers).
- 3.2.2 Our response here focuses on the alcohol purchasing survey, the Living Cost and Food (LCF) survey (formerly the Expenditure and Food Survey - EFS). Individuals participating in this survey keep a purchasing diary for two weeks within which they record information about purchases of a range of products, including alcohol.
- 3.2.3 Before we address the critiques directly, it is worth re-iterating that the estimates of price elasticities from which we take the values used in the modelling are based on analysis of data from the LCF. These elasticity estimates are broadly in line with studies of the best international evidence which review research using a range of research designs and data collection methods [3, 4]. Although not definitive, this provided ourselves, policy makers and previous peer reviewers with reassurances that the data does not contain major biases.
- 3.2.4 In responding to the critique made of survey data, we would again reiterate that we ourselves have pointed out the limitations of the data sources.
- 3.2.5 Duffy & Snowdon appear to believe that the Sheffield researchers are unaware of the fact that the LCF's two week diary data appears to overestimate the numbers of people with very high or zero purchasing levels. This is in no way the case. We have been very careful to seek to mitigate this effect of the data collection window by conducting our analyses using subgroups with a large enough sample of people. One way we have achieved this is by pooling the annual data from 2001/2 to 2005/6 into a single dataset to increase the sample size and thus reduce uncertainty. As such, we have taken care to ensure that mean levels of purchasing recorded in the LCF's diary data are representative of the true underlying mean levels of purchasing in the population from which the data is collected.
- 3.2.6 We are already aware of the limitations of survey datasets in terms of both non-response bias and response bias and have done work within our reports for the Scottish government to examine the effects of both of these which Duffy & Snowdon do not mention. In our most recent report to the Scottish Government [5], we describe extensive sensitivity

analyses around the survey data on alcohol consumption. These analyses take account of non-response bias by making evidence-based modifications to the survey data to address under-representation of key population subgroups in the survey sample (e.g. students, dependent drinkers). The analyses also estimated the effect of accounting for response bias in the form of under-recording of alcohol consumption, which may occur due to a wide range of potential biases which are discussed at length in the epidemiological literature and not just 'deliberate misreporting' as implied by Duffy & Snowdon [6-9].

- 3.2.7 To summarise, we are fully aware of the limitations of different kinds of survey data and have carried out a range of analyses to assess the robustness of the data which is available for use in our modelling. We are aware the data has limitations and we have sought to conduct analyses to quantify these and have included the results within our reports to the policy makers and wider stakeholders so that they also have been able to consider them.

3.3 *Peak consumption regression model*

- 3.3.1 Duffy & Snowdon raise two main points regarding SAPM's analysis of heavy episodic (i.e. binge) drinking. First, they query the measure of heavy episodic consumption used and, second, they argue the model linking average consumption to heavy episodic consumption may not be reliable.
- 3.3.2 Heavy episodic consumption is measured by asking, for each of a range of different beverage types (e.g. beer, strong beer, spirits, fortified wines), how much the respondent drank on their heaviest drinking day in the last week. The total for each beverage can be converted into units of alcohol using standard assumptions about the strengths of different beverages [10] and adding the units for all beverages together. This measure of heavy episodic consumption is referred to as 'peak daily consumption'.
- 3.3.3 Duffy & Snowdon argue peak daily consumption measures underestimate heavy episodic consumption as the questions only address drinking in the last week. The limitations of consumption measures used in alcohol epidemiology are well-known by researchers in this area and have been for many decades [7, 11]. It is correct to state that measures which ask respondents to recall recent consumption tend to lead to over-representation of the extremes of heavy and zero consumption; however, they also tend to provide estimates of per capita consumption which are a closer match to more accurate sales and taxation data than the main alternative which is asking about how much people have usually drunk over the last year or a similarly long period [12].
- 3.3.4 Survey data have limitations and, to date, no feasible survey measure has been developed which can be shown to measure alcohol consumption with perfect accuracy. Unsurprisingly, given the difficulties, Duffy & Snowdon make no suggestions as to how this might be achieved. Throughout their critique of the measures of consumption used in our modelling, Duffy & Snowdon do not attempt to balance the strengths and weaknesses of different measures and, more importantly, give no consideration as to what measures are actually available for use in our work within large, representative contemporary UK surveys. All of

the measures that we have used are considered acceptable and are widely used in alcohol epidemiology. As described above, our most recent report to the Scottish Government described our considerable efforts to assess and account for the impact of the limitations of survey measures, and the under-reporting of alcohol they produce, on our estimates of policy impact. This work will soon be published within a peer-reviewed academic journal [13].

- 3.3.5 Turning to how our model of peak day consumption; to estimate effects of pricing policies on acute harms (e.g. alcohol-related violence, road traffic accidents) it is necessary to estimate the change in expected levels of intoxication (which we have measured using the proxy 'peak daily consumption'). As the data on prices paid for alcohol does not allow us to directly calculate how levels of consumption within single heavy drinking occasions change, this involves going through two steps. First, we estimate the expected percentage change in mean weekly consumption following a price change using the price elasticities discussed earlier, and second, we estimate the impact on peak consumption by utilising the statistical relationship between expected mean weekly consumption and expected peak daily consumption. In statistical terms, we do this latter analysis by population subgroup (e.g. male drinkers aged 45-54), so that we can reflect both the price elasticities and the fact that the extent of heavy drinking varies within population subgroups.
- 3.3.6 Duffy & Snowdon query the reliability of the statistical model which provides this estimate of change in peak daily consumption, in particular based on the low value of R^2 for the model provided (para.12). The reader should note that R^2 is a measure of how well a statistical model fits the data and values range from 0 to 1 and indicate the proportion of the variability across individuals in peak daily consumption which is accounted for by the statistical model.
- 3.3.7 We disagree with Duffy & Snowdon's conclusion that our method is unreliable. We are aware of no published evidence on the relationship between mean weekly and peak daily consumption and, as such, we were very keen to investigate it empirically. Having done so, we fitted a parsimonious statistical model which estimates the relationship between mean weekly and peak daily consumption, while accounting for the age and sex of respondents. As an example, for moderate drinking males aged 45-54, the statistical model estimated that the relationship was as follows:

$$\text{Peak Daily Consumption} = 2.802307 * \text{Mean Consumption} + 1.2923073$$

- 3.3.8 This can be interpreted as saying, if a male age 45-54 drinks one more unit of alcohol per week, then, on average, one would expect his peak daily consumption to be 2.802307 units higher. Importantly, when we make operational use of this relationship we do not use the model to directly predict respondents' peak daily consumption; instead, we follow the two step process described above. We know for each individual in our consumption data, their mean weekly consumption and peak daily consumption. We model a pricing policy and

estimate, using the price elasticities, the percentage change in mean weekly consumption. For each individual we use the statistical model to calculate two values for the expected (i.e. population subgroup average) peak daily consumption level, the first using the baseline mean weekly consumption and the second using the newly estimated mean weekly consumption level after the simulated price change. Finally, we then use the percentage change between these two values to adjust each individual within the population subgroup's baseline recorded peak daily consumption.

- 3.3.9 In statistical terms, what Duffy & Snowden term as a 'low' R^2 for a statistical model like this which uses individual-level data is extremely likely to occur because there is substantial individual-level heterogeneity in peak daily consumption behaviour which is not explained by the level of mean weekly consumption, age and sex. In other words, it is to be expected that people of the same gender and age group who drink the same amount in an average week will vary considerably in how much they drink on their heaviest drinking day. As most statisticians and econometricians would agree, a 'low' R^2 for statistical models of this kind using individual-level data does not invalidate the estimated relationship between expected peak daily consumption and expected mean weekly consumption if the relationship is statistically significant in the models, as is exactly the case in our statistical models.
- 3.3.10 We do take account of heterogeneity by using individual-level data from the consumption survey for baseline peak daily consumption; but we do not account for all heterogeneity in response. In other words, we assume the percentage change in individuals' baseline observed peak daily consumption after policy implementation is the same for all members of a given age-sex subgroup.
- 3.3.10 This reflects a wider point which Duffy & Snowden do not clarify, namely that SAPM essentially works by estimating mean effects of price increases on population subgroups (e.g. 18-24 year-old male moderate drinkers). It is important to understand that we have not undertaken a fully individualised approach to modelling heterogeneity in policy impact for different people who are in the same population subgroup. Again, our method is a standard and often necessary approach within epidemiological and health economic modelling [14-16].

4. Health risks and other forms of harm

4.1 *Epidemiological evidence and risk functions*

- 4.1.1 In order to respond fully to Duffy & Snowden's critique of our use of the epidemiological evidence linking levels of consumption to harmful outcomes, it is necessary to present a brief overview of this evidence and how the aspects used in our model are obtained. A key part of the epidemiological evidence is known as risk functions which describe the relationship between the amount of alcohol individuals have consumed and the consequent risk of experiencing a particular outcome. Outcomes may be health-related (e.g. liver cirrhosis, ischaemic stroke, colorectal cancer) or social (crime, work absenteeism) and our model includes both. Outcomes can be divided into those which would never occur in the

absence of alcohol and are thus classified as wholly attributable to alcohol (e.g. alcoholic neuropathy) and those which are partially attributable to alcohol and partially attributable to other causes such as smoking, poor quality diet, air pollution or the behaviour of others (e.g. ischaemic stroke, road traffic accidents).

- 4.1.2 In our model, risk functions for partially attributable health conditions are largely taken from the best-available published research. This research systematically reviews the epidemiological evidence and synthesises all of the studies identified using a technique known as meta-analysis. Meta-analysis essentially calculates the average relationship between alcohol consumption and the outcome across all of the available evidence, taking account of the relative uncertainty of estimates within each study. Studies with a higher degree of uncertainty are given less weight compared to studies with more certainty. As meta-analyses are based on a far greater weight of evidence than any individual study could achieve, evidence from them is widely regarded as the gold standard of epidemiological evidence even though it often merges results from different times and places [17]. In the context of our work, a risk function using specific Scottish data (if one existed) would probably be considered less of a gold standard than a meta-analysis of many harm studies from across many developed countries.
- 4.1.3 It is worth noting that meta-analyses of the kind used in our modelling only synthesise evidence which is taken from the case control or cohort studies¹, which we assume Duffy & Snowden agree are the most appropriate research designs for deriving risk functions [18].
- 4.1.4 When published risk functions were unavailable, risk functions were instead derived using calibration techniques which calculate (or 'fit') a risk function for the available data². For some of the harms we consider, particularly acute harms or those harms wholly attributable to alcohol, there is no alternative but to estimate a calibrated risk function as there is no risk function or suitable analyses from which to derive one available in the published evidence.

¹ Case control studies and cohort studies are two approaches to collecting data on individual's circumstances and behaviours at multiple points in their lives. A case control study identifies individuals (cases) with a condition and matches each of them to 'controls' who are similar on a wide range of relevant characteristics but do not have the condition. However, the cases and controls are not matched on the exposure measure (in this case alcohol consumption). Data on the exposure is collected retrospectively from the cases and controls. Thus the underlying logic is that differences in earlier levels of alcohol consumption can be investigated as possible explanations for why some individuals experience the condition and others do not whilst simultaneously ruling out potential influences from the characteristics the participants are matched on. In contrast, a cohort study follows a sample of a population prospectively over time, typically surveying them at intervals and comparing their behaviours (e.g. alcohol consumption) and characteristics to assess how these contribute to any outcomes they experience. Both types of study are particularly useful in epidemiological research as it is easier to judge whether a behaviour or exposure is linked to an outcome if it can be shown that the exposure preceded the outcome and that changes in the levels of exposure over time have an effect on the likelihood of experiencing the outcome.

² One may question why this is the case for alcoholic liver cirrhosis where much epidemiological evidence is available on its relationship with alcohol consumption. Liver cirrhosis can either be reported in administrative or survey data as alcoholic or various non-alcoholic forms. Typically epidemiological studies calculate risk functions for all forms of liver cirrhosis as recording within subcategories may not be accurate [e.g. 18]. As the structure of other data sources meant we needed to separate alcoholic and non-alcoholic liver cirrhosis and thus required a risk function specific to each, we judged the published evidence was not sufficiently robust to provide a valid risk function and instead we calibrated one appropriate to the Scottish context.

- 4.1.5 Duffy & Snowden raise questions around these different kinds of risk functions and the evidence sources which underpin them. Some of these points merit discussion but several appear built on incorrect understanding of our work. We address four points in turn.
- 4.1.6 Firstly, the critique argues that the partially-attributable health risk functions used in the model are not based on observation of alcohol consumption and outcomes within individuals participating in case control or cohort studies. This is not the case. As we describe above, these risk functions are largely taken from meta-analyses of exactly this kind of study [19].
- 4.1.7 Secondly, Duffy & Snowden argue the risk functions taken from meta-analyses are not based on Scottish data and are thus not applicable to Scotland. As detailed by Duffy & Snowden themselves and described in the 'Survey data' section of this document, individual studies are subject to a range of known and unknown sampling variations. This is why meta-analyses, which average effects across multiple studies and give greater weight to those studies subject to least sampling variation, are regarded as the gold standard of epidemiological evidence. Given using such evidence addresses their concerns regarding the precision of estimates from a single study, it is surprising that Duffy & Snowden offer no explanation as to why their concerns about precision should be disregarded in favour of their concerns about using data from the population and period in question. It is our view that using gold standard evidence where available, appropriate and feasible is the best approach. Although important, this point is essentially moot as the contemporary Scottish studies Duffy & Snowden wish us to use do not exist and Duffy & Snowden have not suggested any more appropriate studies on which they would prefer our estimates to be based.
- 4.1.8 Thirdly, the critique notes there is less strong evidence for some of the non-health risk functions and that these risk functions are based on self-attribution of the outcome to alcohol rather than objective measurement. We acknowledge here and in our reports that there are limitations to these particular risk functions. One should be clear that these points refer only to risk functions for non-health harms. Duffy & Snowden are correct to note that the crime risk functions are based on anonymous survey respondents' attributions of their own self-reported offending to their alcohol consumption and this is similarly true for absenteeism. However, this is the best evidence available. The difference in robustness between the health and non-health epidemiological evidence and the specific limitations of the evidence we use are described in detail in our reports (e.g. p43-53 of our report to NICE [20]) and we present a clear account for policy makers of how we balanced the relative strengths and weaknesses of different evidence sources. In discussing Figure 3.11 (Duffy & Snowden incorrectly cite Figure 3.10), which shows the sizeable impact of using alternative assumption regarding the proportion of crimes attributable to perpetrators alcohol consumption, Duffy & Snowden make the point that this demonstrates the weakness of our evidence. In one sense we completely agree, and Duffy & Snowden make the point for us. That is, we have been very keen to transparently highlight for all readers of our work, and especially for policy makers, the particular areas of our analysis where greater uncertainty is present and how applying alternative assumptions can affect results.

4.1.9 Fourth, Duffy & Snowdon question our approach to updating the calibrated risk functions in our successive reports to the Scottish Government and point to perceived inconsistencies in the results which emerge. When updating the risk functions, the method we chose was to use the published evidence on the proportion of each harm which is attributable to alcohol (the attributable fraction) as the gold standard and to fit our risk functions to this. Duffy & Snowdon raise the alternative possibility of utilising the risk function itself as the gold standard evidence and computing revised attributable fractions; however, as discussed above, there is no consensus on what this risk function should look like and any choice would be open to criticism. We have provided three reports to the Scottish Government containing three consecutive different estimates of the calibrated risk functions as new evidence on absolute levels of harm and levels of exposure to alcohol in Scotland have emerged over time. In doing so, we have essentially met one of Duffy & Snowdon's other criticisms, which is that there is not enough reflection of uncertainty of the model parameters in the results. The three updated estimates are transparent and both the Scottish Government and any independent third party can examine them and indeed make re-estimations using alternative risk functions should they wish to do so.

4.1.10 In summary, we have utilised the best available evidence for risk functions including published international meta-analyses. We have sought advice from leading researchers internationally and been transparent in the evidence used and its strengths and limitations. Duffy & Snowdon are somewhat inconsistent in wanting both gold standard evidence with minimal uncertainty and, at the same time, direct contemporary Scottish evidence on every aspect of risk of alcohol-related harm. Further, they do not acknowledge the limitations of the available data in Scotland. We have tried to balance these two issues by blending published international work with up to date Scottish data. Duffy & Snowdon make no suggestions for other evidence either internationally or from Scotland that could be brought to bear. We believe we have transparently presented the evidence base available to policy makers "warts and all" and we have enabled them to ask for sensitivity analyses and to recommend and have us incorporate Scottish data sources wherever it has been possible.

4.2 *Sensitivity analyses*

4.2.1 Duffy & Snowdon argue that we do not allow for sampling and other forms of statistical error when quantifying our estimates of potential policy impact. We take this to mean that they would like us to provide the explicit probabilities that our estimated outcome will fall within a certain range (e.g. there is a 95% chance that, with a 50p minimum price, the number of premature deaths avoided per year would be between 100 and 300). From the text of the main body of their critique and the line of argument in the appendix, we understand that Duffy & Snowdon would like to see sensitivity analyses undertaken which simultaneously take account of all possible sources of uncertainty across all of the inputs to our model. Below we refer to this as conducting a full probabilistic sensitivity analysis.

4.2.2 We agree with Duffy & Snowdon that policy makers should be as aware as possible of the uncertainty in estimates of potential policy impact. Indeed, our School at the University of Sheffield is one of the leading proponents of the argument that full probabilistic sensitivity

analyses are useful in health economic research. For policy makers to have as full as possible an understanding of uncertainty, we have continued to write comprehensive and extremely transparent reports on our methodology and results.

- 4.2.3 We have, firstly, and contrary to Duffy & Snowdon's claims, undertaken and reported sensitivity and scenario analyses using alternative data sources or assumptions to test the impact of these on our estimates. Secondly, we have updated our original analyses for Scotland twice as new evidence has become available; a process which furthers some understanding of uncertainty in model outputs. Thirdly, we have, wherever possible, confirmed that our results are in line with existing evidence and, where evidence is weaker or absent, that our assumptions and results are in line with the theory-based expectations of experts within our field. Fourthly, throughout the process of consultation with stakeholders and policy makers we have further developed the sensitivity analyses undertaken in response to both academic researchers' critique and wider policy maker and stakeholder questions.
- 4.2.4 So why have we not undertaken the full simultaneous uncertainty analysis suggested by Duffy & Snowdon? There are four main reasons. Firstly, our modelling has utilised many sources of evidence which do not report uncertainty in their estimates. For example, one of our principal data sources for estimating proportions of harmful health outcomes attributable to alcohol (known as alcohol attributable fractions) includes no confidence intervals [21]. This problem makes full probabilistic sensitivity analysis a very difficult exercise. In this case, it would potentially involve a further larger set of assumptions made by analysts.
- 4.2.5 Secondly, many of our data sources have very large sample sizes and hence are subject to very little statistical uncertainty (e.g. the Expenditure and Food Survey (n=44,150), the Scottish Health Survey (n=7,099) and the baseline mortality and hospitalisation data which are obtained from data on the whole population of Scotland (n≈5.2m).
- 4.2.6 Thirdly, we, the peer reviewers and many other commentators felt that it was much more important to undertake one way sensitivity analyses which test alternative assumptions or data sources, rather than quantifying statistical uncertainty around every one of the values inputted into our model. This is in line with guidance on best practice in such model appraisals (e.g. 5.72 in The Green Book [22] published by HM Treasury). For example, due to many conditions resulting from many years of drinking, there is uncertainty about how long it takes for changes in a population's drinking behaviour to result in changes in rates of alcohol-related diseases [23]. Therefore, we have re-run the model assuming a 10-year time lag between changes in drinking and the full impact on individuals' risk of experiencing such conditions and then again with a 5-year time lag. This allows policy makers to see exactly how much difference that particular model parameter makes to results, rather than imposing an analyst-assumed confidence interval on the time lag and developing a spurious full probabilistic analysis.
- 4.2.7 Fourthly, Duffy & Snowdon's focus on intensive analysis of numerical data as a route to eliminating and quantifying uncertainty suggests a world in which every issue is an empirical one which can be solved by more analysis of more data. In reality it is also important to

consider fully the mechanisms of cause and effect. In other words, it is important to consider how theory relates to analysis and findings from raw data. Prioritising evermore data analysis risks misleading policy makers as alternative theoretical propositions underpinning that data analysis are not tested, leading to unwarranted confidence in seemingly precise findings. These points are crucial to the process of policy appraisal and to our understanding of uncertainty. Therefore, more important for our analyses is what health economists would call 'structural uncertainty analysis.' This involves changing different components of the model structure to test alternative assumptions or alternative beliefs about the appropriateness of the evidence sources. These assumptions or beliefs are then investigated as scenarios [24]. In particular we have investigated those issues which we identified as potentially having important effects. Specific examples of this include alternative price elasticities from the literature which are underpinned by different theoretical approaches and different proportions of crime being attributable to alcohol based on alternative views about which sources of data to believe most. The full set of sensitivity analyses can be seen in our reports but includes testing of:

- An alternative elasticity matrix for hazardous and harmful drinkers which assumes they are one-third less responsive than moderate drinkers [25];
- Alternative preferences for off-trade consumption based on market research data purchased from Nielsen;
- An alternative risk function for ischaemic heart disease which shows greater protective effects of alcohol;
- Alternative estimates of alcohol attributable fractions for crime using the Offending, Crime and Justice Survey;
- Alternative baseline consumption patterns which take into account limitations of the alcohol consumption survey data which are discussed below.

4.2.8 Finally, it should be pointed out that Duffy & Snowdon make a general critique on uncertainty but do not at any point address the challenges or weakness of further analyses in practice. They do not give specific suggestions for specific analyses, nor do they point to evidence on uncertainty in specific elements of the model which we have overlooked or that could be used directly. Most importantly, they fail to acknowledge how important the structural uncertainty analyses we have undertaken are in helping policy makers to understand how the model results are affected by the different uncertain elements within the existing evidence.

4.2.9 In summary, we believe that the process used to consider uncertainty is well-justified and reasonable as it accounts for statistical uncertainty using one-way and multi-way sensitivity analyses when possible, but further considers structural uncertainty on key aspects of the model. Hence, it enables policy makers and wider stakeholders to understand which features of the model 'drive the results' and to what extent. This approach has evolved over four years of feedback, peer review and public consultation.

5. Limitations admitted and not admitted

- 5.1 The final section of Duffy & Snowdon's appendix largely restates points raised earlier. We address the points regarding sampling variation, consumption measures, Scotland-specific crime risk functions and uncertainty above. The limitations of the methods employed to calculate price elasticities are simply quoted from our own report and are documented therein. We do not consider our transparency regarding limitations of our analyses to be a flaw in our work.

6 Summary

- 6.1 Duffy & Snowdon do not identify any evidence which has been omitted from consideration or make any recommendation for inclusion of alternative evidence within the modelling.
- 6.2 The conclusions drawn from our work are in line with those found in widely accepted reviews of the relevant evidence and policy effectiveness literature. Duffy & Snowdon do not dispute this at any point.
- 6.3 Many of the points made by Duffy & Snowdon are essentially a commentary on the limitations of epidemiological evidence and they have been considered previously within our team, with academic researchers under peer review and with wider stakeholders under government and public consultation.
- 6.4 Throughout our work we have followed best practice and clearly documented our methods and the limitations of our estimates. Our work has been submitted to the highest levels of scientific scrutiny over several years.
- 6.5 There is a substantial set of guidance and good practice in regard to the necessary process of appraisal and impact assessment prior to policy implementation. We have engaged fully in these processes and undertaken our work in this context. The place of scientific evidence within this process is crucial and it is the role of the model to enable a synthesis of all of the available scientific evidence and make estimates of the potential impact of a policy which has yet to be implemented. It is important to note that many government policy decisions are made with substantially less evidence-based analysis than has been undertaken for minimum pricing. Until implementation happens, the model remains an estimate of potential effects. The judgement as to whether the collection and synthesis of evidence within the modelling is reliable enough for policy makers to use for decision-making will rightly follow a complex public process of debate. We fully appreciate the Duffy & Snowdon critique as part of that process.
- 6.6 In conclusion, whilst Duffy & Snowdon raise several relevant points concerning the modelling, there is nothing within the critique that has not been previously heard, considered, and discussed within our team and with policy makers, peer-reviewers and wider stakeholders. Duffy & Snowdon do not make any specific, feasible recommendations for improving the model assumptions or estimates, they do not identify any issues with the process of transparent presentation of the evidence which is our core objective in

supporting policy makers decisions and nor do they identify omissions we have made in our considerations of the substantial evidence base on alcohol pricing, consumption and harm.

References

1. Rehm J., Baliunas D., Borges G.L.G., et al. The relation between different dimensions of alcohol consumption and burden of disease: an overview. *Addiction* 2010; **105**: 817-43
2. Huang C-D. *Econometric Models of Alcohol Demand in the United Kingdom*. London: HM Customs and Excise; 2003 May.
3. Wagenaar A.C., Salois M.J., Komro K.A. Effects of beverage alcohol price and tax levels on drinking: a meta-analysis of 1003 estimates from 112 studies. *Addiction* 2009; **104**: 179-90
4. Gallet C.A. The demand for alcohol: a meta-analysis of elasticities. *Aust J Agric Resour Eco* 2007; **51**: 121-35
5. Meng Y, Hill-MacManus D, Brennan A, Meier P. Model-based appraisal of alcohol minimum pricing and off-licensed trade discount bans in Scotland using the Sheffield Alcohol Policy Model (v 2): Second update based on newly available data. SchARR, University of Sheffield; 2012.
6. Midanik L.T. Perspectives on the validity of self-reported alcohol use. *Br J Addict* 1989 Dec; **84**: 1419-23
7. Midanik L.T. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* 1988; **83**: 1019-30
8. Duffy J.C., Waterton J.J. Under-reporting of alcohol consumption in sample surveys: the effect of computer interviewing in fieldwork. *Br J Addict* 1984 Sep; **79**: 303-8
9. Gmel R., Rehm J. Measuring Alcohol Consumption. *Contemp Drug Probl* 2004; **31**: 467-540
10. Goddard E. Estimating alcohol consumption from survey data: updated method of converting columes to units. London: Office for National Statistics; 2007.
11. Stockwell T., Donath S., Cooper-Stanbury M., et al. Under-reporting of alcohol consumption in household surveys: a comparison of quantity-frequency, graduated-frequency and recent recall. *Addiction* 2004; **99**: 1024-33
12. Stockwell T., Zhao J.H., Chikritzhs T., Greenfield T.K. What did you drink yesterday? Public health relevance of a recent recall method used in the 2004 Australian National Drug Strategy Household Survey. *Addiction* 2008; **103**: 919-28
13. Meier P.S., Meng Y., Holmes J., 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 ALCOHOLISM* 2013;
14. Brennan A., Chick S.E., Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ* 2006; **15**: 1295-310

15. Gunningschepers L. The Health Benefits of Prevention - A Simulation Approach. Health Policy 1989; **12**: 1-255
16. Briggs A., Sculpher M., Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.
17. Greenhalgh T. How to read a paper: getting your bearings (deciding what the paper is about). Brit Med J 1997; **315**: 243
18. Rehm J., Taylor B., Mohapatra S., et al. Alcohol as a risk factor for liver cirrhosis: A systematic review and meta-analysis. Drug Alcohol Rev 2010; **29**: 437-45
19. Corrao G., Bagnardi V., Zambon A., La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. Prev Med 2004; **38**: 613-9
20. Purshouse, R., Brennan, A., Lattimer, N., Meng, Y., Rafia, R., Jackson, R., and Meier, P. Modelling to assess the effectiveness and cost-effectiveness of Public health related strategies and interventions to Reduce alcohol attributable harm in England using the Sheffield Alcohol Policy Model version 2.0. Report for NICE. 2009
21. Jones L, Bellis MA, Dedman D, Sumnall H, Tocque K. Alcohol-attributable fractions for England: alcohol-attributable mortality and hospital admissions. Liverpool: Liverpool John Moores University, Centre for Public Health; 2008 Jun.
22. HM Treasury. The Green Book: Appraisal and Evaluation in Central Government. London: TSO; 2011.
23. Holmes J., Meier P.S., Booth A., Guo Y., Brennan A. The temporal relationship between per capita alcohol consumption and harm: A systematic review of time lag specifications in aggregate time series analyses. DRUG ALCOHOL DEPEND 2011; **123**: 7-14
24. Bojke L., Claxton K., Schulpheer M., Palmer S. Characterising structural uncertainty in decision-analysis models: a review and application of methods. Value in Health 2009; **12**: 739-49
25. Chisholm D., Rehm J., Van Ommeren M., Monteiro M. Reducing the global burden of hazardous alcohol use: A comparative cost-effectiveness analysis. J Stud Alcohol 2004; **65**: 782-93