

**REVIEW**

# The development of scientific evidence for health policies for obesity: why and how?

MB Richardson<sup>1,2</sup>, MS Williams<sup>3</sup>, KR Fontaine<sup>2,3</sup> and DB Allison<sup>2</sup>

Potential obesity-related policy approaches have recently been receiving more attention. Although some have been implemented and others only proposed, few have been formally evaluated. We discuss the relevance, and in some cases irrelevance, of some of the types of evidence that are often brought to bear in considering obesity-related policy decisions. We discuss major methods used to generate such evidence, emphasizing study design and the varying quality of the evidence obtained. Third, we consider what the standards of evidence should be in various contexts, who ought to set those standards, as well as the inherent subjectivity involved in making policy decisions. Finally, we suggest greater transparency from both academics and policymakers in the acknowledgment of subjectivities so they can distinguish and communicate the roles of empirical evidence and subjective values in the formulation of policy.

*International Journal of Obesity* (2017) 41, 840–848; doi:10.1038/ijo.2017.71

**INTRODUCTION**

Proposals to use policy measures such as taxing persons with obesity as ways to raise revenue and discourage poor health behaviors, including high levels of consumption, existed at least as early as 1904.<sup>1</sup>

However, it was largely in the mid-1990s that the academic and professional dialog around obesity shifted from one dominated by basic science and clinical research to involve a third branch, namely, public health approaches. Inspired in part by the successful efforts to curtail cigarette smoking, potential obesity-related policy approaches began receiving more attention. A selection of such policies include, but are not limited to, providing information (for example, labeling restaurant menus with nutritional facts), marketing ideas to inspire behavior change (for example, placing public health posters in subway systems to discourage or encourage certain food or activity behaviors), mandating the measurement and reporting of the body mass index of schoolchildren to parents, enacting worksite economic contingencies, changing food offerings for schoolchildren, zoning of allowable restaurants, banning the sale of certain portion sizes, taxing or subsidizing certain foods, and providing economic incentives and disincentives through insurance charges. Some of these have been implemented and some have only been proposed. Few have been rigorously evaluated and fewer still have unequivocal evidence demonstrating efficacy in stabilizing or reducing body weight.

Because the implementation of such policies typically involves at least some of the following: money, limitations on the freedom of businesses to engage in certain types of commerce, limitations on personal freedom and opportunity cost with regard to time and attention; it is not surprising that obesity-related policy proposals often provoke heated debate. Moreover, the debate frequently focuses on moral issues, sometimes involving the balance between autonomy and beneficence or individual fairness

and societal benefits. Because these issues revolve around morals and values, they are difficult to reconcile. As such, they are repeatedly deferred while the dialog jumps to questions of judging the quality of evidence. Yet even here, disagreements abound as to the strength of evidence and whether it supports a particular position on a proposed policy. Equally important and sometimes debated, but often simply glossed over, are questions such as: (i) What type of evidence is needed and appropriate for a particular situation? (ii) How can such evidence be generated? and (iii) Is evidence even needed at all to justify the implementation or rejection of a particular proposed policy?

In this article, we address three macro-level questions. First, concerning evidence, we raise questions about the relevance of some types of evidence that are often brought to bear in policy dialogs. Second, we discuss the major methods used to generate such evidence, with particular focus on the fact that there are a range of study designs (that is, ordinary association tests (OATs) to pure randomized controlled trials (RCTs)) that yield evidence of varying quality and varying ability to support causality. Third, we consider what the standards of evidence should be in various contexts, as well as who ought to set those standards, and emphasize the inherent subjectivities involved in making policy decisions. We conclude by noting that it would be beneficial if both academics and policymakers were transparent in recognizing and conveying those subjectivities while taking care to both understand and distinguish the roles of empirical evidence and subjective values.

**WHAT DO WE WANT EVIDENCE ABOUT?**

Evidence regarding plausibility

When considering a potential policy, the first evidence-oriented question we might ask is, 'Is there evidence that the policy will

<sup>1</sup>Department of Population Health Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; <sup>2</sup>Nutrition Obesity Research Center, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA and <sup>3</sup>School of Nursing, Auburn University, Auburn, AL, USA. Correspondence: Dr DB Allison, Nutrition Obesity Research Center, School of Public Health, University of Alabama at Birmingham, Ryals Building, Room 140J, 1665 University Boulevard, Birmingham, AL 35294, USA. E-mail: dallison@uab.edu

Received 23 September 2016; revised 8 February 2017; accepted 4 March 2017; accepted article preview online 15 March 2017; advance online publication, 18 April 2017

plausibly be effective?' That is, is there reason to speculate that the policy will work? Of course, beyond simply saying we cannot prove the contrary, the plausibility of a proposition is subjective, but one's reasons for declaring something plausible or implausible can be specified. At the most superficial level, many obesity policies can be deemed plausible on the basis of the simple concept of energy imbalance as a cause of obesity. Any policy directed at either increasing energy expenditure or decreasing energy intake might thus be assessed as plausible by some. In some cases, this general plausibility is all that is needed to initiate a policy. For example, when considering calorie-labeling of restaurant menus, the US District Judge Richard J Holwell ruled that:

'The Court agrees with Dr Allison that one cannot conclude with scientific certainty from the available evidence that a regulation of this type will ultimately be successful in combating obesity. But even if there are no data demonstrating conclusively that Regulation 81.50 will be effective, conclusive proof is not required to establish a reasonable relationship between Regulation 81.50 and the City's interest in reducing obesity. Based on the evidence presented by the City, as well as common sense, it seems reasonable to expect that some consumers will use the information disclosed pursuant to Regulation 81.50 to select lower calorie meals when eating at covered restaurants and that these choices will lead to a lower incidence of obesity.'<sup>2</sup>

In contrast, empiricists (or Bayesians) might state that the existing evidence indicates that no proposed public health approach to obesity has been convincingly shown to work or, at best, that no approach has more than very modest effects when it has been applied or tested.<sup>3,4</sup> Therefore, the *a priori* expectation is that the next proposed policy will have little to no effect. By analogy, this rationale is similar to the statistically minded high school guidance counselor who advises the basketball star to study academics because, while the counselor cannot rule out that this player will be the one to get drafted to the NBA or WNBA, it is unlikely.

Plausibility may also be low in some people's minds for policies that aim to affect one component of energy balance in one context while leaving other components of energy balance untouched. Such policies, even if effective in altering the one component of energy balance in the desired direction, will only be effective if this alteration is not compensated (or is at least incompletely compensated) for by alterations in other components of energy balance. Empirical, experimental evidence indicates that such compensation does indeed occur, although the compensation is usually incomplete.<sup>5</sup> This suggests that the plausible effects of policies that work through proposed alterations in one component of energy balance should not be based on models that assume no compensation (c.f., The Caloric Calculator, which estimates average caloric impact, which predicts effect sizes for childhood obesity interventions<sup>6</sup>), as such models will likely markedly overestimate plausible effects.

The plausible benefit of many proposed policy approaches also rests on the assumption of additivity—a small effect coupled with several other slight effects will collectively produce a larger response in the outcome. This is particularly applicable to the category of 'nudge,' a term introduced by Thaler and Sunstein<sup>7</sup> to describe multiple, minor, likely unnoticeable changes to alter one's behavior. Rozin *et al.*<sup>8</sup> showed that multiple modest changes, or nudges, affecting food accessibility (location of ingredients at a salad bar and size and type of serving utensils) in a cafeteria setting reduced the calories purchased during single meals without removing choices. They predicted that the reduced purchasing would translate to a cumulative benefit of weight loss

over 1 year. Again, this type of study relies on several assumptions: that fewer calories purchased translates to fewer calories consumed; that 'all else is equal,' that is, no compensation occurs; that short-term effects persist in the long term; that multiple interventions have additive effects; and that effects of interventions work equally well when subjects are fully aware of the interventions (as in ordinary commerce) as when the interventions are not disclosed (as in many studies). Such a study also brings up questions of whether patrons would purchase fewer calories in an ordinary setting such as a store and that would result in weight loss. For example, Wansink and colleagues found that increasing the cost of soda resulted in reduced soda purchased but was associated with increased sales of beer.<sup>9</sup> These nudges also may elicit a different response when persons are made aware of the interventions or with repeated long-term exposure (that is, daily or weekly grocery shopping).<sup>10</sup> The nudge approach has also been criticized on several other grounds,<sup>11,12</sup> and such criticism highlights that what seems plausible to one person may not seem so to another.

#### Evidence regarding postulated intermediaries

Evidence of the effectiveness of obesity policy may also rest on evidence regarding presumed mediating variables. An example is a proposed policy for an action intended to increase fruit and vegetable consumption, with the main assumption being that increased intake of fruits and vegetables (the mediating factor) will decrease adiposity or promote less weight gain. Empirical support for the policy may include a demonstration that the proposed action does indeed lead to increased fruit and vegetable consumption. However, such support can only be suggestive because it does not necessarily follow that increases in fruit and vegetable intake will actually decrease or prevent adiposity or lessen weight gain, and the same argument applies for other postulated intermediaries.<sup>13,14</sup>

#### Evidence from analog studies

Analog studies attempt to represent key aspects of ordinary life while controlling or limiting external factors, which increases internal validity and can yield key insights<sup>15</sup> yet potentially decreases external validity.<sup>16</sup> An example of an analog study was conducted by Epstein *et al.*<sup>17</sup> to compare the effects of taxation versus subsidization on food purchases. They found that using taxes on foods with low nutrient density but also high caloric content was successful at reducing caloric intake, whereas subsidizing low-calorie foods increased caloric intake. This type of evidence supports the plausibility, but not necessarily the effectiveness, of a policy for decreasing obesity. One area of opportunity is the use of pragmatic RCTs, which emphasize rigorous methods in real-world contexts.<sup>18</sup>

#### Direct evidence regarding effectiveness

Of course, the key evidence desired is evidence of a policy's effectiveness on the ultimate outcome: decreased levels of obesity. Although optimal, such evidence is often difficult to obtain. Ultimately, an ideal study would bear direct evidence of effectiveness, under actual conditions of use, during extended periods of time and would be of a nature to allow strong inference of cause and effect. These would be randomized studies of actual policy or of extremely close proximity. There is no question that these studies would be difficult, expensive, time-consuming and in some cases potentially unethical. We do not advocate a lack of action without this type of evidence; however, there should be a clear understanding that without such evidence, statements about the effects of a policy remain speculative.

### Evidence regarding unintended consequences

It is important to keep in mind that implementation of any policy often brings with it unintended and undesirable consequences. Many of these consequences have been previously highlighted.<sup>19,20</sup> Such consequences can include, but are not limited to, inequitable distribution of the costs to implement the policy, encroaching on individual freedoms, over-consumption or increased purchasing of certain foods, stigmatization, depression and avoidance of doctor appointments.<sup>19,20</sup> One author contends that the emphasis on body weight has led to weight-based bullying, increased disordered eating, body dissatisfaction, extreme dieting and complications from obesity surgery, among others.<sup>21</sup> Although some evidence exists on potential unintended and undesirable consequences, it is fairly limited as this field has not been fully investigated. Again, fear of unintended negative consequences should not paralyze us into inaction, but should lead us to practice humility about the potential value of our proposals, to think things through carefully, and to vigilantly monitor implemented policies for any potential unintended consequences.

### Evidence regarding public opinion

Reports of the results of public opinion surveys on the desirability of particular obesity-related policies have proliferated in recent years.<sup>22</sup> By implication, this suggests that if a large portion of the population supports a proposed policy, then implementing the proposal is merited. Is such a conclusion reasonable? Should evidence of public opinion about the desirability of policies be considered?

Suh et al.<sup>23</sup> suggest that public opinion should be solicited to 'better understand the public mindset about relevant policy strategies, and to identify attitudes among different subsets of the population towards specific legal measures that can increase protections for individuals affected by obesity'. Pollard et al.<sup>24</sup> also contend that it is important to survey public opinion or community perception, especially when the policy in question involves what may be thought of as government 'interference' in issues concerning food (labeling, advertising and supply of environmentally friendly food). But are such opinions always important? When assessing public opinion is warranted, which methodologic issues are involved? And, are there actually circumstances when assessing public opinion would be quite inappropriate? Because this article is primarily about evidence for effectiveness, we consider these questions only briefly here.

*Are scientific assessments of public opinions about policies always important?* Throughout the history of the United States, political leaders have wrestled with the pursuit of what seems morally right based on fundamental principles and doing what is popular. One such example is the famous Lincoln–Douglas debates about slavery. In one of the debates, Lincoln famously said, 'In this and like communities, public sentiment is everything. With public sentiment, nothing can fail; without it nothing can succeed. Consequently he who moulds public sentiment, goes deeper than he who enacts statutes or pronounces decisions. He makes statutes and decisions possible or impossible to be executed'.<sup>25</sup> It is noteworthy that Lincoln, like some modern day authors interested in obesity policy,<sup>26</sup> is talking about 'moulding' public opinion to enable what one has already determined is right and just, and not assessing public opinion to determine what is right and just.

If Lincoln had conducted a public opinion poll and found that most pre-Civil War Americans favored retaining slavery in the United States, would he have judged that pertinent evidence as to whether the practice should be abolished? Would we? The answer is evidently no. When something is judged to be morally wrong, it is wrong and should be 'off the table' for discussion regardless of its popularity. Consider the recent posting from Ted Kyle on a 'UK

Proposal for Explicit Weight Discrimination in Healthcare'.<sup>27</sup> Kyle argues that a proposed policy was a grossly unjust form of discrimination against persons with obesity in terms of health care access. Or, consider proposed policies that entail institutionalized 'fat shaming'<sup>28</sup> or a failed/withdrawn Mississippi bill to limit access of persons with obesity to restaurants.<sup>29</sup> Many, including the current authors, would consider such proposals morally indefensible, and if one adopts such a position, then no public opinion polls are needed. If moral opinion has superior authority relative to public opinion, this invites important questions of who or how many determine the moral authority and on what basis.

*When public opinion assessment is warranted, which methodologic issues are involved?* The above notwithstanding, situations certainly exist where public opinion is important, such as to determine whether a policy which is neither morally indefensible nor a moral imperative is desired by the citizenry. In such situations, it will be important to rely on good principles of designing and interpreting opinion surveys and to keep in mind that who is surveyed<sup>30</sup> and how questions are worded<sup>31</sup> can both be used to manipulate the answers one receives. Extensive discussions on these and other methodologic points are covered in standard textbooks on survey and sampling methods.

*Are there actually circumstances when assessing public opinion would be quite inappropriate?* Finally, we suggest that in some circumstances, assessing public opinion is not only unnecessary, but inappropriate. Specifically, in situations where a proposed approach is morally indefensible, to admit the value of public opinion surveys on determining whether a policy should be enacted invites a 'tyranny of the majority'.<sup>32,33</sup> An interesting corollary of this is that empirical evidence on the harm or lack of benefit for some morally indefensible practice might also be seen as not only unnecessary, but counterproductive, because the very act of considering the empirical evidence implies that the practice under consideration might be worthy of adoption if the evidence came out a particular way. For example, consider this headline from an internet posting: 'Science Says Fat Shaming Backfires—So Can We Finally Stop It?'.<sup>34</sup> The article seems to be referencing an observational study<sup>35</sup> that is interpreted to show that perceived weight discrimination leads to greater future obesity in the person experiencing the discrimination. The answer to the headline's rhetorical query, 'So Can We Finally Stop It?', in our opinion is that we unequivocally should stop fat shaming, but not because of this (or any other) study but rather because it is wrong. Even if one accepts our view that fat shaming is wrong *a priori*, might one ask where the harm is of buttressing the position with some empirical support. The harm is that the empirical support, like all empirical support, is subject to differential interpretation, criticism and being overturned. In the observational study in this example, it would be easy to point out many limits, most notably that the study cannot show cause and effect. This may lead others to conclude, 'Well, if the wrongness of fat shaming depended in part on the empirical evidence and the empirical evidence has holes in it, I guess fat shaming may not be wrong after all'. If this example is not stark enough, we can ask ourselves would we take seriously the need for studies to show deleterious effects of policies that institutionalized racial or religious discrimination as justifications for eliminating such heinous policies?

### HOW MIGHT EVIDENCE FOR OBESITY POLICIES BE GENERATED?

We now turn from the question of what evidence we want to the question of how such evidence can be generated. In doing so, we emphasize that we are focusing in this section on questions regarding the effects of potential policies on outcomes and do not consider questions about assessing other things such as public

opinion about policies. In considering the generation of evidence regarding the effects of potential policies, we are considering questions of cause and effect and readers may find the videos available from an annual short course on this topic of interest (see: <http://www.norc.uab.edu/courses/shortcourse>).

Here, we divide the types of research to be considered into three categories: (i) research that can be determinative of the causal effects of policies; (ii) research that can contribute to an overall assessment of the causal effects of policies, but cannot on its own determine causation; and (iii) research that formally synthesizes multiple sources of information to estimate the causal effects of policies.

## RESEARCH THAT CAN BE DETERMINATIVE OF THE CAUSAL EFFECTS OF POLICIES

### Role of randomized controlled trials

Empirical evidence derived from RCTs aimed at identifying factors that increase or decrease the risk or magnitude of obesity can provide the strongest evidence to guide the development of obesity policies.<sup>36</sup> RCTs are regarded as the gold standard in the hierarchy of research designs because they are the most reliable method for determining causality.<sup>37</sup> Evidence generated from RCTs has been used to guide the development of several types of obesity policies such as dietary recommendations, sugar-sweetened beverage taxes and food pricing.<sup>38–40</sup> Despite the acknowledgment that RCTs offer the strongest inferences about cause and effect, several arguments are commonly offered against reliance on RCTs for causal inference in policy research. We very briefly review these arguments here.

*RCTs are imperfect.* Some authors note that RCTs are imperfect. They can be designed and executed with flaws. Like all empirical studies, they are subject to stochastic variation. Finally, they often entail subject selection criteria and/or study conditions that limit generalizability of the results owing to the broader population and more ‘real-life’ circumstances. These are all legitimate criticisms, but two things are noteworthy. First, these weaknesses are all surmountable. RCTs can be designed and executed well and can be executed in large enough samples and tested with small enough nominal type 1 error levels to minimize stochastic errors. Finally, pragmatic controlled trials offer investigators the ability to conduct a study that examines the effectiveness and efficacy of an intervention in the real world by allowing for the inclusion of a diverse sample of the population and by enabling the intervention to be adapted to local settings.<sup>41</sup> For example, the Moving to Opportunity study found that certain social programs involving housing vouchers providing the ‘opportunity to move from a neighborhood with a high level of poverty to one with a lower level of poverty was associated with (caused) modest but potentially important reductions in the prevalence of extreme obesity and diabetes’.<sup>42</sup>

*RCTs are sometimes impractical or impossible.* We agree that RCTs are sometimes impractical or impossible, but this has no bearing on the extent to which RCTs and other designs can or cannot provide strong inferences or causation. The argument that (i) RCTs are sometimes impractical or impossible, (ii) such that if we relied on only them for strong causal inferences we would not be able to make strong causal inferences in some situations in which we wished to make strong causal inferences and (iii) therefore we should not make strong causal inferences solely from RCTs, is simply a special case of *Argumentum ad Consequentiam*.<sup>43</sup>

*There are no RCTs showing that parachutes work.* It is sometimes noted that we accept many propositions as true on the basis of some evidence and intuitive obviousness such as that smoking

causes lung cancer or that parachutes save lives among skydivers.<sup>44</sup> This is an example of argument by analogy.<sup>45</sup> Arguments by analogy can be useful foils to provoke thought, but in and of themselves prove or disprove nothing.

*We cannot wait for perfect data.* It is sometimes argued that we cannot (or more aptly should not) wait for perfect data to take certain actions, such as enact certain policies. We agree with this proposition. However, the statement ‘we cannot (or more aptly should not) wait for perfect data to take certain actions’ is not equivalent to ‘we cannot (or more aptly should not) wait for perfect data to draw strong conclusions about causation’. Taking actions and drawing causal conclusions are distinct processes and the need and justification to take prudent action in the face of uncertainty is not a justification for denying that the uncertainty exists.<sup>19,46,47</sup>

*Inadvertently promoting a false dichotomy.* Majumdar and Soumerai<sup>48</sup> have cogently noted that ‘some contend that only randomized controlled trials produce trustworthy evidence’. Unfortunately, such a position discounts valid nonrandomized or quasi-experimental study designs, even though health policy randomized controlled trials are rarely feasible. Such a constrained view inappropriately lumps together valid evidence from strong nonrandomized designs (that is, before–after studies with concurrent controls or the interrupted time series study in which a policy causes a sudden, visible change in trend) with evidence from weak designs that permit little causal inference (that is, the commonly conducted cross-sectional analysis that looks at outcomes only after a policy has been implemented). We agree that there is a continuum of non-RCT designs that vary in the strength of causal inferences they justify. We also agree that the stronger designs are underutilized as we discuss later in this article. However, these recognitions do not affect the validity of propositions that randomization is key to valid causal inference.<sup>49</sup> If we accepted otherwise, we would again be engaging in *Argumentum ad Consequentiam*.<sup>43</sup>

## RESEARCH THAT CAN CONTRIBUTE TO AN OVERALL ASSESSMENT OF THE CAUSAL EFFECTS OF POLICIES

Having emphasized the critical role of RCTs in humans on the policy under question and for the outcomes under question in drawing strong causal inference, we also note that with such information often unavailable and sometimes unattainable, it is frequently necessary to make decisions about actions without drawing firm conclusions about causation. In doing so, we must commonly integrate multiple sources of information, none of which alone is unequivocally dispositive about a conclusion of causation, to make informed decisions about what might reasonably be expected to work. Several sources of evidence can contribute to such decisions.

### Role of model organism evidence

Model organisms are used to generate information regarding casual relationships that cannot be derived through human studies. For example, exposure to environmental obesogens, such as endocrine-disrupting chemicals, has been identified as a possible factor that increases the risk of obesity.<sup>50,51</sup> Such studies are vital in policy decisions, for example, to approve or disapprove use of a food additive, but cannot offer unequivocal conclusions about causation in humans because of the possible heterogeneity of effects across species.<sup>52</sup>

**Table 1.** Examples of EATs in obesity research

<i>Topic under study</i>	<i>Design</i>	<i>Finding</i>	<i>References</i>
Effects of breastfeeding on offspring obesity	Co-sib control	Mixed	80,81
Effects of socioeconomic status of rearing parents on child BMI	Adoption study	Consistent with a causal effect, but only half of ordinary association	82
Effects of menu labeling on calories purchased	Quasi-experiment	No support	58
Effects of migrating from Tonga to New Zealand	Natural experiment	Some evidence for BMI increase in some ages	83
Effects of roommate characteristics on freshman weight gain	Packet randomized experiment	Association suggesting that being assigned to higher BMI roommate leads to less weight gain	84
Effect of education on food choice	Quasi-experiment	No support	85
Effect of casinos (as economic boosters) on child obesity	Quasi-experiment	Association suggestive of beneficial effect	86
Effects of altitude of residence on obesity	Quasi-packet-randomized experiment	Association suggestive of beneficial effect	87
Effects of environmental factors influencing birthweight on adult BMI	Co-twin control	No support	88

Abbreviations: BMI, body mass index; EATS, extended association test.

**Role of observational evidence: of OATs and extended association tests**

Observational evidence generally has a vital role in assessing the likely value of proposed policies. Observational studies are useful in generating hypotheses that can inform the conduct of more rigorous studies (that is, randomized trials) to begin to establish causality. With regard to policies developed to address the obesity epidemic, observational studies have been used to investigate associations between the initiations of policies and relevant outcomes. That said, not all observational evidence is of equal value. Here we distinguish between two broad classes of observational evidence, which we will call OATs and extended association tests (EATs).

*Ordinary association tests.* We define OATs to be observational studies on samples of individuals in which the sole or primary means of controlling for potential confounding factors is inclusion of measures of some potential confounding factors as covariates in statistical models (or stratifying by measures of such factors). OATs are heavily relied upon in thinking about plausible effects of policies, but have also been heavily criticized in general<sup>53,54</sup> and in the obesity and nutrition domains in particular<sup>55–57</sup> for multiple reasons. We refer the reader to those references for details.

*Extended association tests.* Most dialog and research in obesity does not consider the evidence continuum between OATs, which do not offer strong assessments of causal effects, and RCTs, which do offer strong inferences, but cannot be done in all circumstances. In contrast to this polarized view, there are techniques that we refer to as extended association tests that lie intermediary between OATs and RCTs, including but not limited to quasi-experimental studies and natural experiments. Such designs are increasingly used, especially in the disciplines of economics and genetics, but are rarely used in obesity research. However, the ability to draw causal inferences in obesity research could be strengthened by increased judicious use of such approaches. In-depth understanding and appropriate use of the full continuum of these methods requires input from disciplines including statistics,

economics, psychology, epidemiology, mathematics, philosophy and in some cases behavioral or statistical genetics. The application of these techniques, however, does not involve routine well-known ‘cookbook’ approaches but requires understanding of underlying principles so the investigator can tailor approaches to specific and varying situations.

Some of the key methods in use for situations where standard RCTs may not be available include natural experiments, quasi-experiments and experiments in which true randomization is used but subjects are not randomized directly to levels of the independent variable, as described with examples in Table 1.

Quasi-experiments are a useful type of observational study that can be used to investigate the impact of environmental changes on obesity, that is, changes that the investigator did not manipulate. In this case, investigators merely measure outcomes before and after the implementation of a new policy, regulation or other factor that has changed. Within the context of efforts at obesity modification, quasi-experiments have been used to assess the effectiveness of new policies (for example, inclusion of calorie information on menus, implementation of environmental elements thought to promote physical activity (that is, parks, bike lanes, walking trails), use of school-based obesity screening and body mass index report cards).<sup>58–61</sup>

A prime example comes from the US Food and Drug Administration, which implemented regulations requiring franchise restaurant chains with 20 or more locations to provide calorie information on their menus and menu boards. In a quasi-experiment conducted in New York City, receipts were collected from patrons of fast food restaurants before and after menu labeling was implemented. The investigators found that adding calorie information to the menus did not appear to influence the food choices of parents or adolescents.<sup>58</sup> Quasi-experiments such as this are a cost-effective way to evaluate the effects of obesity policies, as well as provide information that might inform modifications to existing policies.

The existence of extended association tests seems to be less well known to many investigators in public health, medicine, psychology and related fields. We believe that many questions

about behavioral, psychological and economic influences on obesity-related variables and many applied questions about the effects of extant or proposed interventions can be addressed more informatively and more rigorously if more investigators availed themselves of these evolving methods related to causal inference from a basis of a sound understanding of fundamental principles.

### RESEARCH THAT FORMALLY SYNTHESIZES MULTIPLE SOURCES OF INFORMATION TO ESTIMATE THE CAUSAL EFFECTS OF POLICIES

Apart from the need to embrace and use the range of potential design strategies available, it is also essential to 'step back' and synthesize the multiple and varied sources of information to evaluate what they can tell us about the causal effects of policies.

#### Role of systematic reviews and meta-analysis

As a result of the growing rates of obesity around the world, the volume of evidence from obesity research has burgeoned. However, owing to variations in the quality and type of study design, implementation and the outcomes measured, determining effects from various studies can be challenging. Debates on obesity policies are often fueled by the contradictory findings of empirical studies, such as those regarding the influence of sugar-sweetened beverage consumption on childhood obesity.<sup>40</sup> As such, high-quality systematic reviews and meta-analyses can be useful when attempting to evaluate the state of the evidence related to a particular intervention or policy with objective approaches to identifying and integrating evidence.<sup>62</sup> That said, as Ingram Olkin once wrote, 'Doing a meta-analysis is easy. Doing one well is hard',<sup>63</sup> and we have found that errors in obesity-related meta-analyses abound.<sup>64</sup> Hence, although meta-analyses are vital, our field needs to improve their execution, and meta-analyses should be as critically reviewed as are any other studies.

#### Role of modeling

One drawback of RCTs, noted above, is that they often are not large enough to capture the entire spectrum of effects (both desired and undesired) that a policy may have.<sup>65</sup> Mathematical and computational models of health policies are tools that can be used to predict the outcomes of an obesity policy and to identify implementation barriers before the policy is adopted.<sup>66</sup> Moreover, the modeling of obesity policy enables policymakers to estimate the costs of implementing policies and to determine the resource allocation required to implement a given policy.<sup>67–69</sup> For example, a dynamic weight loss model was used to estimate the effects of a tax on sugar-sweetened beverages on the prevalence of obesity in New York City.<sup>70</sup> The model suggested that there would be decreases in obesity prevalence over a 10-year period.<sup>70</sup> The model also estimated the magnitude of the projected reductions in obesity prevalence, allowing readers to better judge the potential public health impact of such a policy.<sup>70</sup>

Models are also valuable for monitoring the effects of policies over time. Evidence has shown that the effects of health policies can increase or diminish with the passage of time.<sup>69</sup> Therefore, new data concerning the effects of a policy should be continually generated to estimate its effects in order to allow policymakers to revise or even discontinue implementation of the policy if it is shown to be ineffective.<sup>65</sup>

Despite the benefits of using models in the development and refinement of health policies, some challenges and limitations must be recognized. For example, health policy modelers are not often integrated into the health policymaking process. Therefore, models are seen as 'one-offs' rather than as tools that should be used during the lifecycle of the policy to ensure that it retains its value. Perhaps most importantly, models offer projections of

effects, not demonstrations of effects. Such projections can be heavily dependent on the input parameters (that is, assumptions) of the model, and some published modeling activities<sup>71</sup> are so heavy on assumptions of efficacy of the policies considered that the modeling can be seen as an instance of *petitio principii*.<sup>72</sup>

### STANDARDS FOR EVIDENCE AND RELATED FACTORS INFLUENCING POLICY DECISIONS

According to Donaldson *et al.*,<sup>73</sup> most obesity-prevention bills enacted between 2010 and 2013 were based on initiating strategies (for example, 'initiated farmer's markets, increased access to walking trails, local menu labeling') that had little to no evidence of benefit. But is this wrong? A vital consideration, often not made explicit *a priori*, concerns the standards for evidence that will be used to both generate a policy decision and to evaluate its effect once implemented. In general, the standards of evidence for a scientific conclusion are thought to be far more rigorous, because they are based on long-established methods that are considered to be objective, repeatable and relatively immune to biases of the individuals conducting the study. In contrast, the evidence (if any) needed to reach a policy decision (which is distinct from reaching a scientific conclusion) depends on many factors and is not constant across circumstances. Opinions can also vary. For example, the Society for Prevention Research states, 'To be ready for broad dissemination, a program must not only be of proven effectiveness, but it must also meet other criteria...' (emphasis added).<sup>74</sup> This stands in marked contrast to the statement of District Judge Richard J Holwell quoted above that 'even if there are no data demonstrating conclusively that Regulation 81.50 will be effective, conclusive proof is not required...' and in the context of the legal proceedings, his interpretation of law is what determined the evidence standard. There are yet other standards in different

**Box 1** Contrasting ideas on the amount and rigor of evidence regarding policies targeted at sugar consumption (emphases added below)

'It's a classic example. It's industry-funded authors saying that the dietary guidelines recommendations about sugar aren't based on science. I'm laughing because what kind of evidence do you need? Sugar is calories and no nutrients and everybody would be healthier eating less of it.'—Marion Nestle.<sup>75</sup>

'I would be very surprised if any one policy had any effect on obesity. And in some ways I think it's a trap to expect it to do that because so many things are driving the problem. So many things are affecting the food supply, levels of physical activity, and [...] structural things like poverty, education, and access to health care. To expect any one policy to turn that around I think is wishful thinking. But it is certainly important that it be studied as much as it can so that you know at the end of the day, so that you know [...] where you get the biggest impact per dollar of policy change'—Kelly Brownell.<sup>76</sup>

'In public health, when someone is going to act, particularly for something that is a public policy my goodness, we have to have some pretty strong sense that it's going to be effective. Public policy cannot be enacted simply based on a good idea. There has to be reason that it's going to make a difference and a difference relative to public interest. '—Nancy E Kass.<sup>77</sup>

'But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like.'—Sir Austin Bradford Hill.<sup>78</sup>

**Box 2** Clarification within two domains: evidence for scientific conclusions and for advocacy of policy decisions

Evidence for scientific conclusions	Greater candor in scientific presentations <sup>89–92</sup> Acceptance by empirically minded scientists that action can sometimes legitimately precede without strong evidence Articulating distinctions between our values and our assessments of empirical evidence <sup>91</sup>
Evidence for advocacy of policy decisions	Eschewing fallacious rhetorical arguments Acceptance by advocates that advocacy neither requires nor justifies making evidence seem stronger than it is

contexts and so no universal rule about how much evidence is or is not needed for policymaking can be given. This stands in contrast to occasional statements from academics that seem to state from no formal basis of authority that a particular amount of evidence is or is not needed to enact a policy.

The four quotations listed (see Box 1) are from discussions and presentations involving policies directed at curbing sugar intake in the public. They reflect the varying perspectives of differing standards of evidence among researchers. The first two<sup>75,76</sup> put rigor of evidence aside and instead emphasize that the decision to develop policy is the priority based on a decision that seems to have already been committed to based on some combination of suggestive evidence or intuition. In contrast, the third and fourth statements progress from needing 'a strong sense that it will be effective'<sup>77</sup> to confidently requiring 'strong evidence' before any public policy decision.<sup>78</sup> Thus, disagreement on the amount and rigor of evidence needed to enact a policy exists even among researchers discussing a particular target (sugar) of public policy. They illustrate the subjectivity of the standards of evidence for decision making.

**IN SUMMATION**

In closing, our field will benefit from a greater emphasis on probative research. Probative research would meaningfully move us forward in our ability to state that a given treatment or prevention strategy does or does not have a particular effect.<sup>79</sup> This is in contrast to studies that merely continue to draw attention to the plausibility of some treatment having some effect but do not increase our knowledge that such an effect actually exists.<sup>79</sup> Finally, the quest for rigorous evidence and scrupulous truthfulness in reporting is fully compatible with the quest for beneficence and passionate pursuit of action for the betterment of others. Recognizing these comparabilities (see Box 2) may pave the way for public health dialog in obesity that is both more honest and more collegial.

**CONFLICT OF INTEREST**

DBA has received financial support from numerous nonprofit and from profit organizations including government, litigators, food and beverage, pharmaceutical and publishing companies. The remaining authors declare no conflict of interest.

**ACKNOWLEDGEMENTS**

This work is supported in part by the NIH (T32HL105349) and the Nutrition Obesity Research Center (P30DK0563360). The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization. We gratefully acknowledge Gregory Pavela, Alexandra S Allison, Ted Kyle, Jennifer Holmes, Olivia Affuso, Paula Chandler-Laney and Daniel L. Smith Jr for input on this manuscript.

**REFERENCES**

1 Dawson W. Taxation by weight. Vol. 1. Printed and Published by the Office of the British Medical Association: London, UK. Available from: <https://books.google.com/books?id=xY5MAQAAMAAJ&pg=PA1094&lpg=PA1094&dq=>

22taxation+by+weight%22+BMJ&source=bl&ots=Wx-ddlm8\_n&sig=RyJYezZqM\_5t7vV1LVGmcX3BLBM&hl=en&sa=X&ved=0ahUKEWjR0cqO6qnLahXBPIYKHR14AbkQ6AEIHDA#v=onepage&q=%22taxation%20by%20weight%22%20BMJ&f=false (accessed on 11 March 2016).

2 Holwell RJ. Case 1:08-cv-01000-RJH. [Memorandum Opinion and Order]. Doc 42. 2008: 1–27. Available at: <http://lawyersusaonline.com/wp-files/pdfs/CALORIE.pdf> (accessed on March 2016).

3 Cawley J. Does anything work to reduce obesity? (Yes, modestly). *J Health Polit Policy Law* 2016; **41**: 463–472.

4 Essington M, Hertelendy AJ. Legislating weight loss: are antiobesity public health policies making an impact? *J Health Polit Policy Law* 2016; **41**: 453–461.

5 Dhurandhar EJ, Kaiser KA, Dawson JA, Alcorn AS, Keating KD, Allison DB. Predicting adult weight change in the real world: a systematic review and meta-analysis accounting for compensatory changes in energy intake or expenditure. *Int J Obes* 2015; **39**: 1181–1187.

6 Wang YC, Hsiao A, Orleans CT, Gortmaker SL. The caloric calculator: average caloric impact of childhood obesity interventions. *Am J Prev Med* 2013; **45**: e3–e13.

7 Thaler RH, Sunstein CR. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press: New Haven, CT, USA, 2008.

8 Rozin R, Scott S, Dingley M, Urbanek JK, Jiang H, Kaltenbach M. Nudge to nobesity I: minor changes in the accessibility decrease food intake. *Judgm Decis Mak* 2011; **6**: 323–332.

9 Hanks A, Wansink B, Just D, Schulze W. From Coke to Coors: a field study of a fat tax and its unintended consequences. *J Nutr Educ Behav* 2013; **45**: S40.

10 Matjasko JL, Cawley JH, Baker-Goering MM, Yokum DV. Applying behavioral economics to public health policy: illustrative examples and promising directions. *Am J Prev Med* 2016; **50**: S13–S19.

11 Howlett M. Matching Policy Tools & Their Targets: Beyond Nudges and Utility Maximization in Policy Design. Presented at European Consortium for Political Research; 24–28 April 2016, Pisa, Italy.

12 Marlow ML. Market test or government guess? Are government efforts to 'nudge' us to lose weight really based on science?. *Regulation* 2014; **37**: 48–54.

13 Kaiser KA, Brown AW, Bohan Brown MM, Shikany JM, Mattes RD, Allison DB. Increased fruit and vegetable intake has no discernible effect on weight loss: a systematic review and meta-analysis. *Am J Clin Nutr* 2014; **100**: 567–576.

14 Ledoux TA, Hingle MD, Baranowski T. Relationship of fruit and vegetable intake with adiposity: a systematic review. *Obes Rev* 2011; **12**: e143–e150.

15 Mook DG. In defense of external invalidity. *Am Psychol* 1983; **38**: 379–387.

16 Reynolds S, Streiner D. Why we do not abstract analogue studies of treatment outcome and scale development. *Evid Based Ment Health* 1998; **1**: 101–102.

17 Epstein LH, Dearing KK, Roba LG, Finkelstein E. The influence of taxes and subsidies on energy purchased in an experimental purchasing study. *Psychol Sci* 2010; **21**: 406–414.

18 Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci* 2011; **13**: 217–224.

19 Allison DB. Evidence, discourse and values in obesity-oriented policy: menu labeling as a conversation starter. *Int J Obes* 2011; **35**: 464–471.

20 Brown AW, Allison DB. Unintended consequences of obesity-targeted health policy. *Virtual Mentor* 2013; **15**: 339–346.

21 Ramos Salas X. The ineffectiveness and unintended consequences of the public health war on obesity. *Can J Public Health* 2015; **106**: e79–e81.

22 Gendall P, Hoek J, Taylor R, Mann J, Krebs J, Parry-Strong A. Should support for obesity interventions or perceptions of their perceived effectiveness shape policy? *Aust N Z J Public Health* 2015; **39**: 172–176.

23 Suh Y, Puhl R, Liu S, Milici FF. Support for laws to prohibit weight discrimination in the United States: public attitudes from 2011 to 2013. *Obesity (Silver Spring)* 2014; **22**: 1872–1879.

24 Pollard CM, Daly A, Moore M, Binns CW. Public say food regulatory policies to improve health in Western Australia are important: population survey results. *Aust N Z J Public Health* 2013; **37**: 475–482.

- 25 Zarefsky D. 'Public Sentiment Is Everything': Lincoln's View of Political Persuasion. *J Abraham Lincoln Assoc* 1994; **15**: 23–40.
- 26 Robbins RNJ, Lundell H, Meyerson J. Views of city, county, and state policy makers about childhood obesity in New York State, 2010–2011. *Prev Chronic Dis* 2013; **10**: E195.
- 27 Kyle T. UK Proposal for Explicit Weight Discrimination in Healthcare. Available at: <http://conscienhealth.org/2016/09/uk-proposal-explicit-weight-discrimination-healthcare/> (accessed 5 September 2016).
- 28 Abrams L. A Case for Shaming Obese People, Tastefully. *The Atlantic*. Published January 23, 2013. Available at: <http://www.theatlantic.com/health/archive/2013/01/a-case-for-shaming-obese-people-tastefully/267446/> (accessed 8 September 2016).
- 29 Obesity Action Coalition. Obesity Action Coalition Calls on the State of Mississippi House of Representatives to Withdraw House Bill 282 Discriminating Against Those Affected by Obesity. Published 1 February 2008. Available at: <http://www.prnewswire.com/news-releases/obesity-action-coalition-calls-on-the-state-of-mississippi-house-of-representatives-to-withdraw-house-bill-282-discriminating-against-those-affected-by-obesity-56763767.html> (accessed 8 September 2016).
- 30 Gelman A. Tracking public opinion with biased polls. *The Washington Post*. Published 8 April 2014. Available at: <https://www.washingtonpost.com/news/monkey-cage/wp/2014/04/09/tracking-public-opinion-with-biased-polls/> (accessed 8 September 2016).
- 31 Newport F. The Scientific Approach to Question Wording Differences. Gallup. Published 5 September 2003. Available at: <http://www.gallup.com/poll/9193/scientific-approach-question-wording-differences.aspx> (accessed 8 September 2016).
- 32 Adams J. American Imprint Collection (Library of Congress). A Defence of the Constitutions of Government of the United States of America. London printed, Boston re-printed and sold by Edmund Freeman 1788.
- 33 Mill JS. On liberty. In: Eliot CW (ed). *The Internet Wiretap*, vol 25. Harvard Classics, P. F. Collier & Son: New York, NY, USA, 1909.
- 34 Kylie M. Science Says Fat Shaming Backfires — So Can We Finally Stop It? Self Published on 15 September 2014. Available at: <http://www.yourtango.com/2014225975/love-your-body-science-says-fat-shaming-backfires-so-can-we-finally-stop-it> (accessed 8 September 2016).
- 35 Jackson SE, Steptoe A, Beeken RJ, Croker H, Wardle J. Perceived weight discrimination in England: a population-based study of adults aged 50 years. *Int J Obes* 2015; **39**: 858–864.
- 36 Swinburn BA, Caterson I, Seidell JC, James WP. Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutr* 2004; **7**: 123–146.
- 37 Byar DP, Simon RM, Friedewald WT, Schlesselman JJ, DeMets DL, Ellenberg JH *et al*. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976; **295**: 74–80.
- 38 Mozaffarian D, Ludwig DS. Why is the federal government afraid of fat? *New York Times*. Published 9 July 2015. Available at: [http://www.nytimes.com/2015/07/09/opinion/why-is-the-federal-government-afraid-of-fat.html?\\_r=0](http://www.nytimes.com/2015/07/09/opinion/why-is-the-federal-government-afraid-of-fat.html?_r=0) (accessed 8 September 2016).
- 39 Mozaffarian D, Ludwig DS. The 2015 US Dietary Guidelines: lifting the ban on total dietary fat. *JAMA* 2015; **313**: 2421–2422.
- 40 Keller A, Bucher Della Torre S. Sugar-sweetened beverages and obesity among children and adolescents: a review of systematic literature reviews. *Child Obes* 2015; **11**: 338–346.
- 41 Peikes D, Geonntotti K, Wang W. *Using Pragmatic Clinical Trials to Test the Effectiveness of Patient-Centered Medical Home Model in Real-World Setting*. Agency for Healthcare Research and Quality: Rockville, MD, USA, 2013.
- 42 Ludwig J, Sanbonmatsu L, Gennetian L, Adam E, Duncan GJ, Katz LF *et al*. Neighborhoods, obesity, and diabetes—a randomized social experiment. *N Engl J Med* 2011; **365**: 1509–1519.
- 43 Van Eemeren FH, Grootendorst R. Fallacies in pragma-dialectical perspective. *Argumentation* 1987; **1**: 283–301.
- 44 Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003; **327**: 1459–1461.
- 45 Cummings L. Argument from analogy. *Reasoning and Public Health: New Ways of Coping with Uncertainty*. Springer International Publishing: Cham, Switzerland, 2015, pp 93–120.
- 46 Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965; **58**: 295–300.
- 47 Grutters JP, van Asselt MB, Chalkidou K, Joore MA. Healthy decisions: towards uncertainty tolerance in healthcare policy. *Pharmacoeconomics* 2015; **33**: 1–4.
- 48 Majumdar SR, Soumerai SB. The unhealthy state of health policy research. *Health Aff (Millwood)* 2009; **28**: w900–w908.
- 49 Rubin DB. Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics* 1991; **47**: 1213–1234.
- 50 Ouadah-Boussouf N, Babin PJ. Pharmacological evaluation of the mechanisms involved in increased adiposity in zebrafish triggered by the environmental contaminant tributyltin. *Toxicol Appl Pharmacol* 2016; **294**: 32–42.
- 51 Grun F, Blumberg B. Environmental obesogens: organotins and endocrine disruption via nuclear receptor signaling. *Endocrinology* 2006; **147**: S50–S55.
- 52 Reference Manual on Scientific Evidence, 3rd edn. The National Academies Press: Washington, DC, USA, 2011.
- 53 Albert RK. 'Lies, damned lies...' and observational studies in comparative effectiveness research. *Am J Respir Crit Care Med* 2013; **187**: 1173–1177.
- 54 Taubes G. Epidemiology faces its limits. *Science* 1995; **269**: 164–169.
- 55 Schoenfeld JD, Ioannidis JP. Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr* 2013; **97**: 127–134.
- 56 Bohan Brown MM, Brown AW, Allison DB. Nutritional epidemiology in practice: learning from data or promulgating beliefs? *Am J Clin Nutr* 2013; **97**: 5–6.
- 57 Young SS, Karr A. Deming, data and observational studies: a process out of control and needing fixing. *Significance* 2011; **8**: 116–120.
- 58 Elbel B, Gyamfi J, Kersh R. Child and adolescent fast-food choice and the influence of calorie labeling: a natural experiment. *Int J Obes* 2011; **35**: 493–500.
- 59 Naylor PJ, Olstad DL, Therrien S. An intervention to enhance the food environment in public recreation and sport settings: a natural experiment in British Columbia, Canada. *Child Obes* 2015; **11**: 364–374.
- 60 Benjamin Neelon SE, Namenek Brouwer RJ, Østbye T, Evenson KR, Neelon B, Martinie A *et al*. A community-based intervention increases physical activity and reduces obesity in school-age children in North Carolina. *Child Obes* 2015; **11**: 297–303.
- 61 Madsen KA. School-based body mass index screening and parent notification: a cross-site natural experiment. *Arch Pediatr Adolesc Med* 2011; **165**: 987–992.
- 62 Crombie I, Davies H. *What is Meta-Analysis?*. Hayword Medical Communications: London, UK, 2009.
- 63 Shuster JJ, Gieser PW. Meta-analysis and prospective meta-analysis in childhood leukemia clinical research. *Ann Oncol* 1996; **7**: 1009–1014.
- 64 George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J *et al*. Common scientific and statistical errors in obesity research. *Obesity (Silver Spring)* 2016; **24**: 781–790.
- 65 Metcalf CJ, Edmunds WJ, Lessler J. Six challenges in modelling for public health policy. *Epidemics* 2015; **10**: 93–96.
- 66 Hawkes C, Smith TG, Jewell J, Wardle J, Hammond RA, Friel S *et al*. Smart food policies for obesity prevention. *Lancet* 2015; **385**: 2410–2421.
- 67 Backholer K, Beauchamp A, Ball K, Turrell G, Martin J, Woods J *et al*. A framework for evaluating the impact of obesity prevention strategies on socioeconomic inequalities in weight. *Am J Public Health* 2014; **104**: e43–e50.
- 68 Dietz WH, Bland MG, Gortmaker SL, Molloy M, Schmid TL. Policy tools for the childhood obesity epidemic. *J Law Med Ethics* 2002; **30**(suppl): 83–87.
- 69 Levy DT, Mabry PL, Wang YC, Gortmaker S, Huang TT, Marsh T *et al*. Simulation models of obesity: a review of the literature and implications for research and policy. *Obes Rev* 2011; **12**: 378–394.
- 70 Ruff RR, Zhen C. Estimating the effects of a calorie-based sugar-sweetened beverage tax on weight and obesity in New York City adults using dynamic loss models. *Ann Epidemiol* 2015; **25**: 350–357.
- 71 Gortmaker SL, Wang YC, Long MW, Giles CM, Ward ZJ, Barrett JL *et al*. Three interventions that reduce childhood obesity are projected to save more than they cost to implement. *Health Aff (Millwood)* 2015; **34**: 1932–1939.
- 72 Cummings L. Informal fallacies as cognitive heuristics in public health reasoning. *Informal Log* 2014; **34**: 1–37.
- 73 Donaldson EA, Cohen JE, Villanti AC, Kanarek NF, Barry CL, Rutkow L. Patterns and predictors of state adult obesity prevention legislation enactment in US states: 2010–2013. *Prev Med* 2015; **74**: 117–122.
- 74 Society for Prevention Research. Standards of Evidence. Criteria for Efficacy, Effectiveness, and Dissemination. Available at: <http://www.preventionresearch.org/StandardsOfEvidencebook.pdf> (accessed 8 2016).
- 75 National Public Radio (Producer). How Much Is Too Much? New Study Casts Doubts On Sugar Guidelines, 19 December 2016. Retrieved from: <http://www.npr.org/templates/transcript/transcript.php?storyId=505867535>.
- 76 National Academy of Medicine (Panel). 2016 Annual Meeting Session 3 Question and Answer: Evidence of Preventing Obesity and Diabetes at the Population Level and Promising New Directions, 24 October 2016. Retrieved from: [https://www.youtube.com/watch?v=I8RyxGfQyWg&list=PLqRL5HQ\\_hA8faH-DrOXZV7sA\\_HuGMdktom&index=26](https://www.youtube.com/watch?v=I8RyxGfQyWg&list=PLqRL5HQ_hA8faH-DrOXZV7sA_HuGMdktom&index=26).
- 77 Nancy Kass (Presenter). Obesity Prevention policies in the United States: which approaches make the most ethical sense? UT Southwestern Health Sciences Digital Library and Learning Center, 8 November 2016. Retrieved from: <https://repositories.tdl.org/utswmed-ir/handle/2152.5/3892>.
- 78 Bradford Hill, Austin (Presenter). President's Address Section of Occupational Medicine: The Environment and Disease: Association or Causation? 295–300, 14 January 1965. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-0010.pdf>.



- 79 Casazza K, Allison DB. Stagnation in the clinical, community and public health domain of obesity: the need for probative research. *Clin Obes* 2012; **2**: 83–85.
- 80 Metzger MW, McDade TW. Breastfeeding as obesity prevention in the United States: a sibling difference model. *Am J Hum Biol* 2010; **22**: 291–296.
- 81 Nelson MC, Gordon-Larsen P, Adair LS. Are adolescents who were breast-fed less likely to be overweight? Analyses of sibling pairs to reduce confounding. *Epidemiology* 2005; **16**: 247–253.
- 82 Fontaine KR, Robertson HT, Holst C, Desmond R, Stunkard AJ, Sørensen TI *et al*. Is socioeconomic status of the rearing environment causally related to obesity in the offspring? *PLoS ONE* 2011; **6**: e27692.
- 83 Gibson J, McKenzie D, Stillman S. What happens to diet and child health when migration splits households? Evidence from a migration lottery program. *Food Policy* 2011; **36**: 7–15.
- 84 Yakusheva O, Kapiros K, Weiss M. Peer effects and the freshman 15: evidence from a natural experiment. *Econ Hum Biol* 2011; **9**: 119–132.
- 85 Braakmann N. The causal relationship between education, health and health related behaviour: evidence from a natural experiment in England. *J Health Econ* 2011; **30**: 753–763.
- 86 Jones-Smith JC, Dow WH, Chichlowska K. Association between casino opening or expansion and risk of childhood overweight and obesity. *JAMA* 2014; **311**: 929–936.
- 87 Voss J, Allison DA, Webber BJ, Otto JL, Clark LL. Lower obesity rate during residence at high altitude among a military population with frequent migration: a quasi experimental model for investigating spatial causation. *PLoS ONE* 2014; **9**: e93493.
- 88 Allison DB, Paultre F, Heymsfield SB, Pi-Sunyer FX. Is the intra-uterine period really a critical period for the development of adiposity? *Int J Obes Relat Metab Disord* 1995; **19**: 397–402.
- 89 Mehta T, Allison DB. From measurement to analysis reporting: grand challenges in nutritional methodology. *Front Nutr* 2014; **1**: 6.
- 90 Allison DB, Bassaganya-Riera J, Burlingame B, Brown AW, le Coutre J, Dickson SL *et al*. Goals in nutrition science 2015–2020. *Front Nutr* 2015; **2**: 26.
- 91 Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: a tragedy of errors. *Nature* 2016; **530**: 27–29.
- 92 Dietz T. Bringing values and deliberation to science communication. *Proc Natl Acad Sci USA* 2013; **110**: 14081–14087.