Unpalatable food for thought : Let marketing research guide effective public obesity interventions

Holden, Stephen S., Zlatevska, Natalina, Parkinson, Joy, Cadario, Romain, Dubelaar, Chris, Lei, Jing, Moore, Elizabeth, Sayarh, Nada, Van Kerckhove, Anneleen and Werle, Carolina

Zlatevska Natalina (Orcid ID: 0000-0002-0936-1529)

Unpalatable food for thought: let marketing research guide effective public obesity interventions

Stephen S. Holden, Macquarie School of Business, Macquarie University, Sydney, Australia

Natalina Zlatevska*, Faculty of Business, University of Technology Sydney, Sydney, Australia

Joy Parkinson, Griffith Business School, Griffith University, Brisbane, Australia

Romain Cadario, Rotterdam School of Management, Erasmus University, Rotterdam, The Netherlands

Chris Dubelaar, Department of Marketing, Deakin University, Melbourne, Australia

Jing Lei, Department of Management & Marketing, University of Melbourne, Melbourne, Australia

Elizabeth Moore, Mendoza College of Business, University of Notre Dame, South Bend, IN, USA

Nada Sayarh, Geneva School of Economics & Management, University of Geneva, Geneva, Switzerland

Anneleen Van Kerckhove, Faculty of Economics & Business Administration, Ghent University, Ghent, Belgium

Carolina Werle, Grenoble Ecole de Management, Grenoble, France & IREGE, Université Savoie Mont Blanc, Chambery, France

Keywords: obesity, public health interventions, marketing research, meta-analysis

Running title: Unpalatable food for thought

* Corresponding author: natalina.zlatevska@uts.edu.au, +61 2 9514 3265

No potential conflicts of interest

Abstract

The prevalence of obesity is growing unabatedly despite the considerable efforts directed at the problem. While abundant research has contributed to our understanding of the multi-factorial causes of obesity, there is less attention to research that is relevant for guiding social marketers, public health professionals and policy makers in delivering public health interventions for countering and/or preventing the problem of obesity.

This review offers six points for identifying and developing research relevant for guiding community-wide obesity interventions based on the idea that an applied marketing research perspective offers a better model for identifying effective interventions than more theoretical

academic research. Specifically, the research guiding public health & social marketing interventions needs to (1) provide information on ultimate outcomes (weight, health, unintended consequences) more than intermediate outcomes (beliefs, attitudes, behavior), (2) report on observations collected over the longer term, (3) use natural settings (even at a cost of internal validity), (4) endeavor to overcome observer-effects, (5) report effect sizes (rather than statistical significance), (6) use moderator analyses to capture variation in how a population responds to interventions.

1. Introduction: We do not know which obesity interventions will work

Obesity has been identified as a major public-health problem affecting health, longevity, quality of life of people and also having major economic consequences[1–3] with the medical burden of obesity estimated to represent 10% of all medical spending in the US.[4] Governments, health agencies and social marketers around the world are responding to this problem with an investment of approximately $5 billion per annum worldwide on interventions designed to reduce obesity.[5] Interventions in this context are broad social-based programs aimed at bringing about a desired change in terms of reducing and/or preventing further growth of obesity. Such programs may be delivered broadly in the form of education, marketing, law or some combination[6] aimed at consumers directly or “upstream” to retailers, manufacturers, etc.

Despite these substantial efforts, the prevalence of obesity continues to grow.[7,8] The continuing growth in the rate of public obesity suggests that current efforts are inadequate. Why is this so? One view is that the war on obesity requires more financial resources which is roughly equivalent to a brute marketing view that a bigger budget will solve the problem. A second view is that the available tools are not those best-suited to the task: better tools are needed rather than more money. A third, pessimistic view is that the war on public obesity may be unwinnable if it cannot be reversed or prevented with public health efforts. Which view is true appears to require a better understanding of the effectiveness of broad-scale, public, whole-of-market interventions than we currently have available.

Research to date has revealed that the problem of obesity is exceedingly complex, and correspondingly, that finding a solution calls for changed thinking.[9–11] A seductively simple physiological and thermodynamic explanation for obesity points to the role of an energy surplus arising when the energy consumed by an individual exceeds the energy expended over time. This simple model is however incomplete as obesity is affected by a wide range of genetic, environmental, and psychosocial factors[12–15] and arises as an outcome of an exceedingly complex system as illustrated in the Foresight systems map.[16] Energy balance is at the heart of the problem, but it is influenced by a complex network of interdependent factors - biological, developmental, psychological, social, economic and media. Much obesity research is legitimately focused on understanding this complex web of influences.

Despite extensive research aimed at understanding the general problem of obesity, research providing clear guidance for effective public health interventions remains elusive. Knowing more about the effectiveness of broad-scale, public, whole-of-market interventions for addressing public obesity is needed to improve public health efforts by answering the following questions. Are there public interventions that can reverse or forestall prevalence of public obesity? Which interventions are most effective in preventing obesity? How effective are these interventions relative to one another? What are the potential negative, unintended consequences of these public interventions? Despite the abundance of obesity research, there appears to be no clear consensus on what interventions work best or even how well they work relative to one another. The US Institute of Medicine observes that there is "a striking contrast between the high prevalence and consequent importance of addressing obesity and the paucity of the knowledge base with which to inform prevention efforts".[4] On what basis then, are views on how to prevent or reverse obesity being developed? Casazza et al. suggest that such views are based presumptions and myths rather than scientific evidence.[17] Others offer similar views that the evidence we have about what causes obesity and interventions might reduce obesity is "both equivocal and largely circumstantial".[18]

Collectively, we are seeking to answer the question of what public/market-based interventions can be implemented to tackle public obesity. The research available is suggestive but not conclusive. We argue that efforts to shift public opinion, behavior and health within this complex network of influences requires a research re-direction, an orienting towards the practice of applied research in marketing and business. In this kind of research, the focus is on identifying successful interventions and measuring their effectiveness with less attention and interest being directed to developing a comprehensive theory focused on understanding and explaining the phenomenon of obesity.

2. Relevant Research

Public health and policy makers are actively searching for guidance from relevant research. However, research where "the effects of a policy remain speculative"[20] is not particularly helpful. We propose that public health and policy makers will be better served by more applied marketing research which: (1) focuses on the ultimate outcomes that interventions seek to change; (2) focuses on research designs that test interventions in real-world settings over time; (3) focuses on analyses of effect sizes (rather than statistical significance). We discuss these recommendations in the following three sections.

Outcome measures

Much of the extant research in the area of obesity focuses on how variables affect cognitions (including beliefs, attitudes) and behaviors (such as food purchase, consumption, or exercise undertaken). Much less research attends to whether the ultimate desired effects in terms of obesity and health are achieved. While changes in intermediate outcomes – such as a change in beliefs or a change in behavior – may be promising, whether they mediate obesity-related outcomes (such as BMI or weight change) is at best untested or speculative as recognized by multiple researchers.[19–23]

Chain of effects

Whether consciously or unconsciously, most assume that public interventions aimed at reducing obesity operate through a chain of effects consisting of a series of intermediate changes in cognition and behavior leading ultimately to changes in weight and health (see Figure 1).

- - - - - - - - - - -

Insert Fig 1 about here

- - - - - - - - - - -

We identify three broad groups of variables which operate within the chain of effects. The first are public health interventions themselves which can include public information or educational campaigns, manipulations of price (e.g., sugar-sweetened beverage taxes, subsidized exercise programs), or of products (portion size, labeling), or of product availability, or choice-architecture (nudges). The second set of variables relate to cognitions (perceptions, motivations, attitudes, intentions), and behaviors (such as selecting, purchasing, serving and/or consuming less (unhealthy) food, and/or choosing and engaging in more activities). The third and final set represent the ultimate outcomes: long-term or sustained energy balance, and more simply obesity, weight and health.

The challenge for public health is that an intervention's effectiveness must be reflected at the level of health, but much of the available research focuses on intermediate effects far short of the ultimate outcome, and what promise they may offer remains largely unknown. An enormous quantity of existing research focuses on the degree to which cognitions (e.g., beliefs, attitudes) explain behaviors as reflected in the dominant theoretical models such as the Health Belief model, Theory of Planned Behavior, etc.[24,25] Some criticize these models arguing that the cognitions are “difficult to measure, they do not predict behavioral outcomes very well, there is little evidence that they cause behavior, and they are hard to change directly”.[26] This assessment is supported by evidence that behavior change research provides relatively “modest” effects with an R^2 of .3 or less.[24,27] It is troubling that the effect of interventions on cognitions and/or behaviors receive much attention while the effect of behaviors on the ultimate outcomes are much less researched, and much less well-established. In general, the evidence that public health interventions reduce population-wide obesity show limited effectiveness.[28,29]

There are multiple factors that may contribute to significant mitigation of an intervention’s effect as communicated through the chain of effects for obesity. Compensatory behaviors are one well-recognized problem whereby a change in energy-intake or energy-output is compensated in some way. A person might reduce the quantity of food that they eat, but increase the energy-density of the food eaten. Substituting 100g of steak for 100g of cabbage would see a ten-fold increase in calories: 25kcals to 270kcals. Similarly, a person who exercises longer may diminish the benefit by exercising less intensely or eating a treat afterwards. Evidence of a wide range of compensatory processes abound.[30–34]

While most researchers acknowledge the potential for compensation, the degree of compensation is contentious. Some see compensation as incomplete or partial.[35–38] Others claim that compensation completely offsets any reduction in calories consumed[21] which would imply an intervention has no effect. Part of the problem may be that compensation is often observed over time, and so a failure to observe compensation in a short time period may overlook compensation over a longer period. For instance, at the more macro-level, compensation appears to be evident in the way in which an individual's body weight tends to be remarkably stable over time.[13,21,39–41] The operation of compensation is important as it forces a winding back of any assumption that behavioral changes in the short term lead necessarily to the desired longer-term changes in obesity. Compensation undermines the utility of short-term reporting of behavioral outcomes or even short-term changes in energy intake or energy expenditure.[39]

A related issue is that outcome measures often lack validity. For instance, while studies show that food prices through taxes and subsidies do affect purchases,[34,42–44] a problem remains about whether purchases translate to meaningful changes in consumption. In general, "sales and purchase data [are] used as a proxy for consumption",[42] but the validity remains dubious.

To be clear, many if not most obesity researchers wish to genuinely contribute useful knowledge that might help reduce obesity. Obesity-related effects or outcomes are however, difficult to measure, and most studies such as those focused on cognitive and behavioral outcomes are forced to rely on speculation about how their research will contribute to reducing obesity. Baranowski et al. ask in their eponymous paper : "Are current health behavioral change models helpful in guiding prevention of weight gain efforts?"[24] They find that very few of the studies reported in their review measure weight or obesity, and so their question remains to a large degree, unanswered.

The chain of effects model (Figure 1) highlights that the quality of the results depends on the dependent variables used and their proximity to the ultimate outcomes. Outcomes further from the ultimate outcomes become less valid suggesting a clear hierarchy of preferred effects: energy intake should be preferenced over food quantity consumed; consumption over purchase; and behaviors over cognitions. However, the ideal is to focus on ultimate outcomes rather than rely on measures of "presumed mediating variables".[20]

Ultimate outcomes

Which of the ultimate outcomes will be most relevant and useful for public health practitioners? And how might they be measured? The three main contenders are health, obesity and energy balance (see Figure 1).

While public health is often framed as tackling obesity, the ultimate outcome (as highlighted in the chain of effects) is health. In this regard, the single-minded pursuit of reducing obesity may be leading us in the wrong direction.[45] For instance, better health would be a positive outcome even if obesity remained unchanged. In this regard, evidence showing that exercise programs or following the Mediterranean diet can improve health regardless of whether weight is changed[46,47] are therefore encouraging and valuable. However, health outcomes are diverse: physical, metabolic, psychological, absence of disease, etc. So focusing on health *per se* may pose definitional and measurement problems.

Obesity is commonly operationalized in terms of BMI (Body Mass Index = kgs_{weight} / $meters_{height}^2$) which was developed originally by Quetelet in 1832, and was known as the Quetelet Index.[48] The term 'body mass index' was coined later by Keys and colleagues.[49] BMI is a useful measure in epidemiological studies and is argued to be "at least as good as any other relative weight index as an indicator of relative obesity" at a population level.[49] However, it is recognized as having a number of limitations. BMI is not necessarily correlated with adiposity, and it is known to overestimate obesity among taller people and muscular body types, and underestimates obesity among shorter people and children.[50–53] In addition, the common use of bounded ranges (e.g., "normal" = 18.5-24.9 kg/m^2, "overweight" = 25-29.9, "obese" = 30+)[54] are arguably misleading given that repeated observations show the lowest mortality rate is associated with BMIs at the upper end of "normal", specifically 22.5-25 kg/m^2.[55]

More simply, interventions are typically aimed at weight-change, and therefore, the ultimate measure could be weight. Weight is of course the numerator of the BMI measure and the clearest end-point of most obesity interventions; height (the denominator in BMI) is constant.

The measurement of energy balance is appealing in a conceptual sense, but offers an unlikely prospect for a number of reasons. While it is downstream from behavior and closer to the desired ultimate outcomes, it remains removed from the ultimate outcomes – obesity/weight and health. Questionnaire-based measures for energy-in and energy-out are notoriously unreliable.[56–58] "Gold standard measures" of energy-balance are available through underwater weighing or doubly-labeled water, but their feasibility in community settings is limited by expense and effort required.[59] These measurement difficulties become particularly burdensome for assessments over the longer term, leaving weight and health changes as the best measures of the ultimate effectiveness of public interventions.

Unintended outcomes

One of the challenges facing any public policy initiative is that aside from any intended positive outcome, there is the potential for negative, unintended outcomes.[29] There is however, also a potential for any given intervention to have unintended, positive outcomes such as flow-on effects in terms of changes in activities, life-styles, etc. called "co-benefits" by Emmons et al.[59]

One undesirable and potentially unintended effect would be if consumers viewed interventions as irrelevant or overly simplistic.[60] Of greater concern than wasted effort is that interventions can create backfire effects such as increased consumption of unhealthy foods[61] and weight-gain.[62,63] Disturbingly, even the act of simulating being overweight by wearing a fat suit can increase the intake of energy dense foods.[64,65]

There is also a potential for interventions to contribute to stereotyping, stigmatization and follow-on consequences such as negative body image, excessive weight preoccupation, body dissatisfaction, and eating disorders.[29,45] The potential for interventions to harm individuals by ignoring or overriding individual concerns may become grounds for opposing the intervention. For instance, public health efforts may encourage moralizing judgments and incite stereotyping, leading to stigmatization of overweight or obese individuals.[66,67] The weight loss industry, the media and the medicalization of fatness may further contribute to shaming of the obese. Evidence shows that stigmatization can lead to negative feelings, feelings of exclusion, reduced self-esteem, well-being, and depression.[63,68–70]

In sum, the first recommendation for obesity researchers seeking to provide useful information to guide public interventions is to give attention to ultimate outcomes, and particularly weight and health changes. Cognitions and behavior are certainly valuable at a diagnostic and theoretical level, but practice needs a focus on the ultimate endpoints. In addition, unintended outcomes, both positive and especially negative, need to be considered. Ineffective interventions are wasteful, harmful interventions are problematic.

Methods & design

The research methods adopted need to be focused on assessing the effectiveness of interventions in the real world. The objective of useful research is to show whether or not an intervention is effective as distinct from its potential efficacy in a carefully controlled laboratory setting.[71,72] Controlled laboratory settings are indisputably valuable for exploring variables that *can* (under ideal conditions) have an effect, but research on public interventions is more focused on the likely effectiveness of some variable or program in real world conditions. Our suggestion reinforces the call of others for more “practice-based evidence”,[73] for research evidence showing what might work in practice.[72] Research focusing on what is effective in practice needs real-world settings, longer-term observations and reduced observer effects.

Real-world settings

A number of researchers have called for more effort to show the effectiveness of public interventions in reducing obesity in real-world settings.[4,59,74,75] "Most of the evidence [guiding public health] is not very practice-based" and "too much… comes from artificially controlled research that does not fit the realities of practice".[73] For instance, in many food studies, participants are given food for free, and sometimes limited from accessing other food.[76] Self-reported measures of consumption and nutritional intake such as the food frequency questionnaire (FFQ) are unreliable.[56–58] Artificiality of measures reaches extremes in measures of intended consumption, and especially in measures where participants serve themselves fake food from a buffet[77] or draw the amount of food they would serve themselves on a piece of paper.[78]

Artificial manipulations and measures are specific examples of the more general problem of the artificiality of experimental methods. In the field of health and medicine, randomized control trials (RCTs) are held to represent "the gold standard".[79,80] However, some doubt is raised as to whether this is true in the testing of public interventions which ask somewhat different questions. The best or proper approach is one that is best suited to answering the research question being asked.[72,81]

A great deal of the debate hinges around a key distinction between whether a treatment *can* have an effect, and what type of effect it is likely to have in practice. Or stated otherwise, the efficacy of a treatment versus the effectiveness of a social intervention. For instance, a great deal of RCTs show that diets can lead to weight loss. However, in an RCT, the diets will be trialed on volunteers, very likely overweight or those considered at high risk, very likely motivated to lose weight, with non-compliant participants possibly being excluded. In a public context, the public health and policy-makers have much less control. The participants in a real world intervention will often comprise both at-risk and not at-risk targets, persons who have not volunteered, who may have low motivation or even be unwilling to participate, and whose non-compliance will be reflected in the final results. In that context, the observed effectiveness of the intervention is likely to be considerably diminished relative to the potential efficacy of the diet itself. An experiment shows what can happen—in a controlled situation. It shows much less about what will happen in the uncontrolled, real world.

Experimental designs, and particularly those with random assignment to multiple highly controlled conditions (i.e., RCTs) are ideally suited to examining the efficacy of a treatment (the effects of a drug or a diet), and to establish the explanatory potential of a cause. However, the conditions in the real world are uncontrolled, and ethics and practicality limit the degree to which people in the real world can be relied on to stick to an intervention which endeavors to change their behavior (eating, activity, or both). Observational research methods can address such concerns, but these designs are limited, most notably due to threats to internal validity through confounding and selection bias.[59,82]

Observational designs are arguably "the gold standard for assessing the effects of real-world interventions" even while they face significant limitations.[59] For instance, the linking of obesity to higher rates of mortality from all causes comes from observational research.[7,55] While efforts can be made to address concerns with observational research such as using prospective designs and exclusion of periods of time “to limit reverse causality”,[55] it remains true that observational designs are imperfect. And so are RCTs.

RCTs are indispensable to the progress of science, but their limitations, and whether they are fit to purpose is important to consider.[81,83] “Threats to validity [of RCTs are] well known, but they often seem to be forgotten, or to be treated as minor irritants to be handled with some reassuring words or a robustness study, rather than as fundamental limitations on what can be learned from a particular dataset”.[84]

The respective limitations of experimental and observational research methods are overcome in recognizing that these designs have complementary values and roles.[79] The specific contributions of each are recognized in the call for more openness to observational and emerging hybrid research designs for investigating public prevention efforts in obesity[4,59,74] and public health more broadly.[73,80] Other hybrid research designs that might be considered include pragmatic clinical trials,[85] quasi-experimental designs, natural experiments and field experiments.[4,20,86–88] In quasi-experimental designs for instance, researchers may sacrifice full random assignment, but can nevertheless examine the success of interventions *in situ*. There is a need to add more studies to the relatively limited set of innovative, pragmatic trials conducted in public settings.[33,89–92] The results may need to be treated with caution, but this is true for most research being applied to guide public interventions, and notably to much nutrition research which is accused of generating implausible results.[56]

Longer duration

Too many studies purporting to address weight- and obesity-change are based on short term, even one-off eating occasions.[20,93,94] Weight- and obesity-change are long-term phenomena and so, researchers need to “provide the necessary link between individual short-term food choices and long-term weight gain [or change]”.[23] Many researchers are aware of the problem and frequently allude to “a need for high-quality studies … particularly in the longer term”.[95] An article offering early results from a sugar beverage tax in Philadelphia noted that “future studies are needed to evaluate longer-term impact of the tax on sugared beverage consumption and substitutions”.[96] Observations over extended periods of time are particularly important given that longer-term studies of diets show that weight loss tends to plateau after about six months, and in many instances, reverses.[41,97] To know whether an intervention has the desired effect, long-term observations are essential. Obesity like diabetes needs to be managed over the lifetime, and can never rely on one-off interventions. The problem of course, is that studying long-term interventions is time-consuming and costly.[93]

Overcoming observer effects

One potential bias that severely limits the value of experimental designs is that the participants know that they are being observed. As stated by poet and author Robert Penn Warren, “If you look at a thing, the very fact of your looking changes it”.[98] In social research sciences, this is the problem of observer effects,[99] but also includes demand effects.[100–102] While demand effects refers to guessing the researcher’s hypotheses—which is a strong possibility in dieting studies—a more pervasive and pernicious effect is when people change their behavior simply because they are being observed so as to manage how they are perceived by others.[99] This observer effect – sometimes characterized as social facilitation – affects both food consumption and exercise which are often public activities. The power of observers to improve sports and other motor performance through social facilitation has been acknowledged for nearly 100 years.[103] More recent meta-analytic reviews confirm that the mere presence of an observer has small but robust effects on motor performance.[104,105] And importantly, exercise science researchers have acknowledged that social facilitation has the potential to undermine internal validity in experimental studies.[106]

Meanwhile, food and nutrition researchers also acknowledge the extent to which an observer can influence consumption behavior. While research shows that eating with others tends to increase consumption (social facilitation), eating with a passive, non-eating observer is a powerful moderator of consumption.[107] The effect of the passive observer is considered to "trump" the social facilitation effect.[107] This result holds significant implications for research into food consumption where the researcher or assistants could be characterized as passive, non-eating observers with the potential to reduce or mitigate consumption and potentially undermine the validity of a study. A number of recent meta-analyses and studies reveal that when participants are aware they are being observed, they reduce their food consumption even if in the control group.[108–110]

Observer effects are problematic. Even with successful blinding to the treatment, informed consent means that participants are necessarily aware they are participating in a study, and this may lead to a change in behavior which may undermine the study's objectives. Robinson et al. note that observer effects have received little consideration in food consumption research,[109] and have been overlooked in exercise science.[106]

Observer effects undermine the validity of tests of obesity interventions, and it is essential that they are addressed.[111] The first hurdle is acknowledging that observer effects are operating. More practical steps are to attempt to blind participants to both the manipulations and to the measures (through covert recruitment, use of cover stories), and to use *post hoc* questioning to assess participants' awareness of the experimental purpose.[100,102,111,112] Longer term studies measuring weight- or health-related outcomes in response to food manipulations may also mitigate observer effects.

Trials and triangulation

Obesity is a systemic problem, a function of multiple causes.[13,16,59] As outlined at the outset, we are not clear whether the apparent failure of existing interventions represents a lack of sufficient resources, a failure to identify the most effective interventions, or whether the problem can actually be reversed or prevented by existing interventions.

Resolving the problem requires multiple, coordinated efforts over an extended period of time.[5,11,20,56,113,114] There are no silver bullets, there are no short cuts. We add our call to the emerging recognition of a need for evidence of public interventions that work in the real world settings over extended durations of time, away from research setting where motivation and/or observer effects may threaten internal validity. Primarily, there is a need to recognize the limitations of both RCTs and observational designs, and the complementary nature of research designs: "there is no perfect gold standard" in research.[115]

Analyses

The key guide for any analysis is to get an answer to the research question at hand which in this case is whether an intervention is effective or not. Unfortunately, the current predominant research culture encourages analysts to focus on statistical significance rather than effect sizes. This creates and encourages a bias whereby researchers strive to find (and publish) studies which meet the statistical significance threshold, typically $p<.05$.[116,117] The focus on statistical significance is deeply embedded within the world of research, and drives both publication, prestige, and promotion. It is, however, not well-fitted to the requirements of public health and policy practitioners looking for effective interventions to reduce public obesity. The key data needed to guide public interventions are effect-sizes. In particular, the key question we want to know for a public intervention is by how much will weight, BMI and/or health be changed?

Effect sizes

The long-standing focus on statistical significance is being challenged and incisively criticized by many.[118,119] The problem is that statistical significance is widely misunderstood, misused and leads to a misdirection of research efforts.[120] As Cohen remarks, the emperor's new clothes do not exist, and "this naked emperor has been shamelessly running around a long time".[118] In the context of obesity research, it is the effect that matters, and even small effects can be powerfully helpful in real world, public settings.[83]

What we want to know is the probability of whether the observed effect is true. What statistical significance tells us is the probability of an observed datum (D) *given that there is no effect* (H_o), i.e., $p(D/H_o)$. In short, statistical significance testing is related to the probability of a false alarm (Type I error), it is not related to the probability of a research hypothesis, an intervention for example, being true. The confusion is remarkably simply shown, but frequently overlooked: $p(D/H_o) \neq p(H_o/D)$. Statistical significance focuses on the first expression, but many often mistakenly conclude the second expression despite the inequality of the two[*]. A bigger issue (leading to an even greater misunderstanding) is that researchers are really interested in the probability of H_1 given the data, i.e., $p(H_1/D)$. Given the data, how likely is it that the theory (H_1) is true?

In addition to the problematic focus on statistical significance, there is strong tendency for researchers to focus on a specific alpha level, typically $p<.05$. This however encourages a dichotomous view about support for a research hypothesis rather than reflecting a judgment based on an arbitrary cut on a continuous distribution marking the probability of a false positive or Type I error.[118,121,122]

[*] The error is one that is often made by students (and others) examining contingency tables. It is like concluding that the probability that someone is divorced given they have ever been married is equal to the probability of ever have been married given they have been divorced (where the former might be about .5, the latter is 1). The confusion matrix which classifies the kinds of errors we make in judgments (Type 1 = false alarms, Type II = misses) is a specific type of contingency table.

Ultimately, statistical significance which is important in the social academic domain is of limited value in the public policy domain. For instance, a good deal of excellent research has focused on whether a low-carbohydrate or low-fat diet is best.[39,123,124] For all the debate however, it seems apparent that *both* diets work to reduce weight; how well they each work is much less clear.[123,124] Much of the debate is fairly arcane and well-removed from what might work in the real-world where dietary regimes are much more difficult to control. In fairness, the studies have tended to focus on which is most efficacious as this is important at an explanatory level. But they provide little insight on which might be most effective in a public setting.

A related issue which highlights again the distinction between efficacy of a treatment and effectiveness of a treatment is the question of whether experiments should be analyzed after excluding participants who have not faithfully followed the research protocol, or including all, even those that have not complied and correctly followed the research protocol. The answer to this depends on what question the researcher is asking.[125] If we are investigating the potential effectiveness of an intervention in a public health setting, then the inclusion of participants who did not adhere to the research regime in a so-called "intention-to treat" analysis is preferred. Returning to calorie-restrained diets, if individuals find some diets easier to follow, then that will be a better diet to follow.[97,124,125] Both diets may be efficacious among compliant participants, but effectiveness includes consideration of efficacity and compliance. A problem with the intention-to-treat analysis is that efficacy and adherence (or compliance) are confounded,[125] but this typically represents the more useful analysis for real-world effectiveness where treatment *is* confounded with compliance.

While a number of researchers have recognized the need for more attention to effect sizes and confidence intervals,[118,126] the need is even more relevant for public health practice. However, public health researchers will likely find more utility from some effect size metrics than others. Standardized mean difference measures (e.g., Cohen's *d*, Hedge's *g*) are widely reported, but their value for identifying effective obesity interventions is less clear. As Cohen observes, to understand a standardized effect-size metric, it is necessary to try and understand the scale,[127] but this is not easy when measures such as Cohen's *d* are open-ended. Cohen's response has been to classify *d*=.2 as a small effect, *d*=.5 a medium effect and *d*=.8 a large effect. The remaining opacity of the Cohen's d measure is penetrated to some degree when expressed in terms of an equivalent Pearson's correlation coefficient: r=.1 is a small effect, r=.3 is a medium effect, and r=.5 is a strong effect. Hence, a medium effect represented by a correlation of .3 accounts for less than 10% of the variance in the dependent variable. Cohen explains that the medium effect size was set to capture an effect that was "likely to be visible to the naked eye of a careful observer".[127]

Studies that wish to report on the effectiveness of an intervention on some meaningful outcome (rather than scales without any intrinsic meaning) will find raw means[128] and/or ratios may prove to be more useful. That is, a direct measure of how much weight, obesity or health have changed as a result of an intervention will be more helpful than a standardized effect-size measure. In terms of ratios, the effect size might be reported as a percentage change in the dependent variable, or even better, a percentage change in the dependent variable as a ratio to the percentage change in the independent variable (i.e., an elasticity). Similarly, researchers may choose to employ a relative risk or risk ratio (RR) to report the probability of an event occurring in the intervention group versus the control group. Or again, in the context of interventions aimed at reducing obesity, another possibility is to report the effect-sizes in terms of raw means, namely the change in BMI or weight change (lbs or kgs).[129] As weight change is slow, takes place over time, and is subject to dynamic physiological processes,[39] a better effect size measure would be the rate of change in weight over time (e.g., kgs / month). Effect sizes reported in this way are practical, likely to be widely and easily understood, and useful for providing a standard by which different interventions might be measured.

Perhaps the key concern that limits people from reporting effect sizes is the recognition, perhaps unconscious, that many interventions will tend to have relatively small effects. Ioannidis suggests that most treatment effects in medical and social research tend to be small (d<=.5). [83] Notwithstanding, health policy will be better guided by a realistic view of effect-sizes rather than implausible results,[56] and besides, small effects can have major benefits at the population level. Moreover, single interventions with small effects may be combined with multiple other interventions with many thinking this is what will be needed to solve obesity overall.[5,11,20,56,113,114]

Replication & meta-analyses

While science values new, and novel contributions, public health needs a different kind of evidence that justifies the allocation of resources. Science faces something of a problem due to its reliance on testing of statistical significance in many underpowered studies leading to a situation where much, even most published, statistically significant results are likely false.[116,117] The growing acknowledgement of problems of replication in medicine, business, and psychology[116,130,131] confirm the problem and highlight the limitations of single studies.

Effectiveness of public health interventions is not determined from single studies, but from multiple studies which allow for the average effectiveness of the intervention to be determined. When there is an abundance of separate research projects, effect sizes and confidence intervals can be determined from published (and unpublished) research papers through meta-analysis.

In addition to simple estimation of the effectiveness of an intervention, meta-analysis can also be used to test whether effect sizes vary as a function of various individual and environmental factors, typically labeled sub-group or moderator analyses. Meta-analyses reveal wide variation in effects which is a useful reminder of the unreliability of single studies. The opportunity to explore the conditions under which an intervention works well versus conditions where it does not work are important at a public health level. Similarly such analyses are useful if they reveal that something about unintended outcomes.

Analyses of multiple studies conducted in multiple settings observing different levels of control variables offer an opportunity to treat the collected studies as offering a quasi-experimental design with imperfect sampling of the population of replications.[132] With a small number of studies in a meta-analysis, each sub-group (moderator) analysis will be conducted independently. If more studies are available, any potential confounding of moderators maybe avoided through meta-regression.[133] However, the search for moderators in this fashion needs some caveats. The first issue is that meta-analysts can only consider as moderators those variables for which the original research provides results for different levels of the potential moderator. For instance, if researchers report an effect size for participants, but do not report a result for men and women separately, then that study cannot be used in examining the influence of gender on the meta-analytic effect size. The second issue, related to the first, is that the results of any moderator analysis in a meta-analytic setting is *post hoc*. It is an observational analysis based on a constrained and incomplete sampling of the domain of studies that might have addressed this issue—both those that have been conducted, and those that have not. In this regard then, moderator or more correctly, sub-group analyses cannot be relied on as reflecting certainty. but do offer promising directions for future research to explore and resolve.

Moderator analyses conducted within the meta-analytic framework have the potential to provide both useful information for guiding health policy, and useful hypotheses for future research efforts. In particular, the search for moderators serves as a useful reminder that one size of obesity intervention does not fit all. While current efforts tend to focus on “mass-market” obesity solutions, it is well understood that not all people are likely to respond equally to the same intervention. Accordingly, researchers should give attention to identifying variables that moderate responses to obesity interventions, so that public policy can be fine-tuned.

Conclusions

Obesity continues to grow, and we remain unclear as to the relative effectiveness of potential public interventions. Which interventions are most effective? Do they reverse obesity, or prevent future obesity? Are they capable of solving the problem of obesity?

Public-health practice needs what Green identified as “practice-based evidence” rather than the more artificially-controlled, clinical, trial-based evidence that typically supports health science.[73] Public policy is likely to be better served by research that prioritizes effectiveness over explanatory power and efficacy. Practice-based evidence requires thinking more like a commercial marketer, using research that shows what works with less concern for developing an intricate understanding of why.

The first step is to focus on relevant outcome measures. The corpus of obesity research is populated with many studies that focus on intermediate outcomes (cognitions, behaviors), but relatively few that focus on ultimate outcomes. While research on intermediate outcomes is very helpful for understanding and explaining theory (and even diagnosing how public interventions work with some but not others), it is less helpful for identifying and guiding effective public interventions. In particular, the promise of an intervention based on an intermediate outcome is speculative and likely limited given the powerful and acknowledged effects of physiological compensation.[39]

The second aspect is to recognize that despite their limitations, observational and pragmatic, real-world experiments conducted over time maybe more suited to exploring effectiveness of public interventions than RCTs. Designs that are short-term, limited to artificial laboratory settings and that do not attempt to control for observer effects will provide limited insights about what will work in the real world.

Finally, reports of effect sizes are more helpful than reports of statistical significance for judging the effectiveness of an intervention in changing weight or BMI or health over time. In addition, meta-analyses, ideally reporting raw effect sizes (e.g., weight change) are needed to show the average effect of an intervention, and moderator analyses are needed to show conditions under which this effectiveness may vary.

The intention of this paper is call attention to the kind of research that is needed to identify and assess effective public obesity interventions. We acknowledge that there *is* considerable research that has been conducted that meets the terms proposed in this paper. The challenge is to accept that these types of research deserve more weight when designing public interventions than would be accorded to them in the academic world. The second, corollary to the first, is that researchers should be encouraged rather than discouraged to conduct the required designs. In effect, we are arguing that the best type of research for identifying effective public interventions is *not* the same as the best type of research for explaining the phenomenon of obesity.

In addition to challenging entrenched beliefs (e.g., RCTs are the gold standard, statistical significance is the key analysis), our recommendations raise a number of other legitimate concerns such expense (e.g., long duration, natural settings), and ethical concerns (e.g., blinding participants to the intervention). These issues deserve attention, but ongoing pursuit of approaches that do not raise these concerns risks policy decisions being made largely in the dark, without a clear understanding of which interventions are effective and which are not.

Perhaps the greatest constraint to adopting the proposed agenda is a sense that many of the proposed interventions for reducing or arresting the growing prevalence of obesity may prove to be less effective than many might hope. Hope however ought not to trump evidence, and this paper is a call to explicitly label the effectiveness of various public interventions to better guide public health and social marketing agencies tackling obesity. Perhaps effects will be small, but it is better to be realistic, and to be encouraged to not "confuse small change with no change".[134] Even small effects can contribute to the desired results in much the same way as a surfeit or deficit of just 50-100kcals per day in an individual's energy balance will, if sustained over time, create weight change.[135] Researchers widely recognize that obesity will not be solved with a single intervention,[5,11,20,56,113,114] nor will it be solved in a short time. Public obesity will be reduced by multiple, small interventions, gently moving weight and health in the right direction.

Heading in the right direction requires the right research, research that reveals what is effective and what is not.

References

1. Gortmaker SL, Swinburn BA, Levy D, et al. Changing the future of obesity: Science, policy, and action. *Lancet*. 2011;378(9793):838-847. doi:10.1016/S0140-6736(11)60815-5
2. Finkelstein EA, Kruger E. Meta- and cost-effectiveness analysis of commercial weight loss strategies. *Obesity*. 2014;22(9):1942-1951. doi:10.1002/oby.20824
3. Finkelstein EA, Ruhm CJ, Kosa KM. Economic Causes and Consequences of Obesity. *Annu Rev Public Health*. 2005;26(1):239-257. doi:10.1146/annurev.publhealth.26.021304.144628
4. Institute of Medicine (IOM). *Bridging the Evidence Gap in Obesity Prevention: A Framework to Inform Decision Making*. Washington DC; 2010. https://www.nap.edu/read/12847.
5. Dobbs R, Sawers C, Thompson F, et al. Overcoming obesity : An initial economic analysis. *McKinsey Glob Inst*. 2014;(November):120.
6. Rothschild ML. Carrots, sticks, and promises: A conceptual framework for the management of public health and social issue behaviors. *J Mark*. 1999;63(4):24-37. doi:10.1080/15245004.2000.9961146
7. Flegal KM, Kruszon-Moran D, Carroll MD, Fryar CD, Ogden CL. Trends in obesity among adults in the united states, 2005 to 2014. *Jama*. 2016;315(21):2284-2291. doi:10.1001/jama.2016.6458

8. World Health Organization (WHO). Obesity and overweight. Fact sheet. http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight. Published 2017.

9. Sanabria E. CIRCULATING IGNORANCE: Complexity and Agnogenesis in the Obesity "Epidemic." *Cult Anthropol*. 2016;31(1):131-158. doi:10.14506/ca31.1.07

10. Frood S, Johnston LM, Matteson CL, Finegood DT. Obesity, Complexity, and the Role of the Health System. *Curr Obes Rep*. 2013;2(4):320-326. doi:10.1007/s13679-013-0072-9

11. Parkinson J, Dubelaar C, Carins J, Holden S, Newton F, Pescud M. Approaching the wicked problem of obesity: an introduction to the food system compass. *J Soc Mark*. 2017;7(4):387-404. doi:10.1108/JSOCM-03-2017-0021

12. Hall KD, Heymsfield SB, Kemnitz JW, Klein S, Schoeller DA, Speakman JR. Energy balance and its components: Implications for body weight regulation. *Am J Clin Nutr*. 2012;95(4):989-994. doi:10.3945/ajcn.112.036350

13. Spiegelman BM, Flier JS. Obesity and the regulation of energy balance. *Cell*. 2001;104(4):531-543. doi:10.1016/S0092-8674(01)00240-9

14. Jackson SE, Llewellyn CH, Smith L. The obesity epidemic – Nature via nurture : A narrative review of high-income countries. *Sage Open Med*. 2020;8:1-11. doi:10.1177/2050312120918265

15. Rodgers A, Woodward A, Swinburn B, Dietz WH. Prevalence trends tell us what did not precipitate the US obesity epidemic. *Lancet Public Heal*. 2018;3(4):e162-e163.

doi:10.1016/S2468-2667(18)30021-5

16. Vandenbroeck IP, Goossens J, Clemens M. Foresight Tackling Obesities: Future Choices—Building the Obesity System Map. Government Office for Science, UK Government's Foresight Programme.

17. Casazza K, Brown A, Astrup A, et al. Weighing the Evidence of Common Beliefs in Obesity Research. *Crit Rev Food Sci Nutr*. 2015;55(14):2014-2053. doi:10.1080/10408398.2014.922044

18. Keith SW, Redden DT, Katzmarzyk PT, et al. Putative contributors to the secular increase in obesity: Exploring the roads less traveled. *Int J Obes*. 2006;30(11):1585-1594. doi:10.1038/sj.ijo.0803326

19. Herman CP, Polivy J, Vartanian LR, Pliner P. Are large portions responsible for the obesity epidemic? *Physiol Behav*. 2016;156:177-181. doi:10.1016/j.physbeh.2016.01.024

20. Richardson MB, Williams MS, Fontaine KR, Allison DB. The development of scientific evidence for health policies for obesity: why and how? *Int J Obes*. 2017;41(6):840-848. doi:10.1038/ijo.2017.71

21. Benton D, Young HA. Reducing Calorie Intake May Not Help You Lose Body Weight. *Perspect Psychol Sci*. 2017;12(5):703-714. doi:10.1177/1745691617690878

22. Ello-Martin JA, Ledikwe JH, Rolls BJ. The influence of food portion size and energy density on energy intake: implications for weight management. *Am J Clin Nutr*. 2005;82(1 Supplement):241S.

23. Chandon P, Wansink B. Is Food Marketing Making us Fat? A Multi-Disciplinary Review. *Found Trends Mark*. 2011;5(3):113-196. doi:10.1561/1700000016

24. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. Are Current Health Behavioral Change Models Helpful in Guiding Prevention of Weight Gain Efforts? *Obes Res*. 2003;11(S10):23S-43S. doi:10.1038/oby.2003.222

25. Glanz K, Bader MDM, Iyer S. Retail grocery store marketing strategies and obesity: An integrative review. *Am J Prev Med*. 2012;42(5):503-512. doi:10.1016/j.amepre.2012.01.013

26. Jeffery RW. How can Health Behavior Theory be made more useful for intervention research? *Int J Behav Nutr Phys Act*. 2004;1(1):10. doi:10.1186/1479-5868-1-10

27. Glasman LR, Albarracin D. Forming Attitudes That Predict Future Behavior: A Meta-Analysis of the Attitude–Behavior Relation. *Psychol Bull*. 2006;132(5):778-822. doi:10.1111/obr.12065.Variation

28. Walls HL, Peeters A, Proietto J, McNeil JJ. Public health campaigns and obesity - A critique. *BMC Public Health*. 2011;11:1-7. doi:10.1186/1471-2458-11-136

29. Salas XR. The ineffectiveness and unintended consequences of the public health war on obesity. *Can J Public Heal*. 2015;106(2):e79-e81. doi:10.17269/CJPH.106.4757

30. Chandon P, Wansink B. The Biasing Health Halos of Fast-Food Restaurant Health Claims: Lower Calorie Estimates and Higher Side-Dish Consumption Intentions. *J Consum Res*. 2007;34(October):301-314. doi:10.1086/519499

31. Werle COC, Wansink B, Payne CR. Is it fun or exercise? The framing of physical activity biases subsequent snacking. *Mark Lett*. 2015;26(4):691-702. doi:10.1007/s11002-014-9301-6

32. Fenzl N, Bartsch K, Koenigstorfer J. Labeling exercise fat-burning increases post-exercise food consumption in self-imposed exercisers. *Appetite*. 2014;81:1-7.

33. Hollands GJ, Cartwright E, Pilling M, et al. Impact of reducing portion sizes in worksite cafeterias: A stepped wedge randomised controlled pilot trial. *Int J Behav Nutr Phys Act*. 2018;15(1):1-14. doi:10.1186/s12966-018-0705-1

34. Cabrera Escobar MA, Veerman JL, Tollman SM, Bertram MY, Hofman KJ. Evidence that a tax on sugar sweetened beverages reduces the obesity rate : a meta- analysis Evidence that a tax on sugar sweetened beverages reduces the obesity rate : a. *BMC Public Health*. 2013;13:1072-1081. doi:10.1186/1471-2458-13-1072

35. Poppitt SDD, Prentice AMM. Energy density and its role in the control of food intake: evidence from metabolic and community studies. *Appetite*. 1996;26(2):153-174. doi:S0195-6663(96)90013-6 [pii]\n10.1006/appe.1996.0013

36. Looney SM, Raynor HA. Impact of Portion Size and Energy Density on Snack Intake in Preschool-Aged Children. *J Am Diet Assoc*. 2011;111(3):414-418. doi:10.1016/j.jada.2010.11.016

37. Rolls BJ, Roe LS, Beach AM, Kris-Etherton PM. Provision of Foods Differing in Energy Density Affects Long-Term Weight Loss. *Obesity*. 2005;13(6):1052-1060.

doi:10.1038/oby.2005.123

38. Rolls BJ. What is the role of portion control in weight management? *Int J Obes*. 2014;38(10):1-882. doi:10.1038/ijo.2014.82

39. Hall KD, Guo J. Obesity energetics: Body weight regulation and the effects of diet composition. *Gastroenterology*. 2017;152(7):1718-1727. doi:10.1016/S2214-109X(16)30265-0.Cost-effectiveness

40. Fildes A, Charlton J, Rudisill C, Littlejohns P, Prevost AT, Gulliford MC. Probability of an obese person attaining normal body weight: cohort study using electronic health records. *Am J Public Health*. 2015;359(54-59). doi:10.2105/AJPH.2015.302773

41. Fothergill E, Guo J, Howard L, et al. Persistent metabolic adaptation 6 years after “The Biggest Loser” competition. *Obesity*. 2016;24(8):1612-1619.

42. Afshin A, Peñalvo JL, Gobbo L Del, et al. The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLoS One*. 2017;12(3). doi:10.1371/journal.pone.0172277

43. Andreyeva T, Long MW, Brownell KD. The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food. *Am J Public Health*. 2010;100(2):216-222. doi:10.2105/AJPH.2008.151415

44. Colchero MA, Popkin BM, Rivera JA, et al. Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study. *BMJ*. 2016;352:h6704. doi:10.1136/bmj.h6704

45. Sturgiss E, Jay M, Campbell-scherer D, Weel C Van. Challenging assumptions in obesity research. 2017;5303(November):1-4. doi:10.1136/bmj.j5303

46. Shaw KA, Gennat HC, O'Rourke P, Del Mar C. Exercise for overweight or obesity (Review). *Cochrane Database Syst Rev*. 2006;4(Art No. CD003817):1-85. doi:10.1002/14651858.CD003817.pub3

47. Sofi F, Macchi C, Abbate R, Gensini GF, Casini A. Mediterranean diet and health status: An updated meta-analysis and a proposal for a literature-based adherence score. *Public Health Nutr*. 2013;17(12):2769-2782. doi:10.1017/S1368980013003169

48. Eknoyan G. Adolphe Quetelet (1796-1874) - The average man and indices of obesity. *Nephrol Dial Transplant*. 2008;23(1):47-51. doi:10.1093/ndt/gfm517

49. Keys A, Fidanza F, Karvonen MJ, Kimura N, Taylor HL. Indices of relative weight and obesity. *J Chronic Dis*. 1972;25(6-7):329-343. doi:10.1016/0021-9681(72)90027-6

50. Blackburn H, Jacobs D. Commentary: Origins and evolution of Body Mass Index (BMI): Continuing saga. *Int J Epidemiol*. 2014;43(3):665-669. doi:10.1093/ije/dyu061

51. Etchison WC, Bloodgood EA, Minton CP, et al. Body mass index and percentage of body fat as indicators for obesity in an adolescent athletic population. *Sports Health*. 2011;3(3):249-252. doi:10.1177/1941738111404655

52. Taylor RS. To the editor: Use of body mass index for monitoring growth and obesity. *Paediatr Child Health*. 2010;15(5):258. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2912631&tool=pmcentrez&re

ndertype=abstract.

53. Marchand V. Author's response: Use of body mass index for monitoring growth and obesity. *Paediatr Child Health*. 2010;15(5):258.

54. National Institutes of Health. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health. *Obes Res*. 1998;6:51S-209S.

55. Whitlock G, Lewington S, Sherliker P, et al. Body-mass index and cause-specific mortality in 900 000 adults: Collaborative analyses of 57 prospective studies. *Lancet*. 2009;373(9669):1083-1096. doi:10.1016/S0140-6736(09)60318-4

56. Ioannidis JPA. Implausible results in human nutrition research. *BMJ*. 2013;347:f6698-f6699. doi:10.1136/bmj.f6698

57. Kristal AR, Peters U, Potter JD. Is It Time to Abandon the Food Frequency Questionnaire? *Cancer Epidemiol Biomarkers Prevlogy, Biomarkers, Prev*. 2005;14(December):2826-2829. doi:10.1158/1055-9965.EPI-editorial

58. Schaefer EJ, Augustin JL, Schaefer MM, et al. Lack of efficacy of a food-frequency questionnaire in assessing dietary macronutrient intakes in subjects consuming diets of known composition. *Am J Clin Nutr*. 2000;71(3):746-751. doi:10.1093/ajcn/71.3.746

59. Emmons KM, Doubeni CA, Fernandez ME, et al. National Institutes of Health Pathways to Prevention Workshop : Methods for Evaluating Natural Experiments in Obesity. *Ann Intern Med*. 2018;168(11):809-814. doi:10.7326/M18-0501

60. Lewis S, Thomas SL, Hyde J, Castle D, Blood RW, Komesaroff K. "I don't eat a hamburger and large chips every day!" A qualitative study of the impact of public health messages about obesity on obese adults. *BMC Public Health*. 2010;10(309).

61. Pham N, Mandel N, Morales AC. Messages from the Food Police: How Food-Related Warnings Backfire among Dieters. *J Assoc Consum Res*. 2016;1(1):175-190. doi:10.1086/684394

62. Robinson E, Hunger JM, Daly M. Perceived weight status and risk of weight gain across life in US and UK adults. *Int J Obes*. 2015;39(12):1721-1726. doi:10.1038/ijo.2015.143

63. Major B, Hunger JM, Bunyan DP, Miller CT. The ironic effects of weight stigma. *J Exp Soc Psychol*. 2014. doi:10.1016/j.jesp.2013.11.009

64. Incollingo Rodriguez AC, Heldreth CM, Tomiyama AJ. Putting on weight stigma: A randomized study of the effects of wearing a fat suit on eating, well-being, and cortisol. *Obesity*. 2016;24(9):1892-1898. doi:10.1002/oby.21575

65. Robinson E, Proctor M, Oldham M, Masic U. The effect of heightened awareness of observation on consumption of a multi-item laboratory test meal in females. *Physiol Behav*. 2016;163:129-135. doi:10.1016/j.physbeh.2016.04.044

66. Askegaard S, Cheung T, Mathras D, Franziska Junghans A, Miller EG. Moralities in food and health research. *J Mark Manag*. 2014. doi:10.1080/0267257X.2014.959034

67. Cordell G, Ronai CR. Identity management among overweight women: narrative resistance to stigma. In: Sobal J, Maurer D, eds. *Interpreting Weight: The Social*

Management of Fatness and Thinness. New Brunswick, NJ: Transaction Publishers; 1999:24-49.

68. Tomiyama AJ. Weight stigma is stressful. A review of evidence for the Cyclic Obesity/Weight-Based Stigma model. *Appetite*. 2014;82:8-15.

69. Vartanian LR, Porter AM. Weight stigma and eating behavior: a review of the literature. *Appetite*. 2016;102:3-14.

70. Major B, Eliezer D, Rieck H. The Psychological Weight of Weight Stigma. *Soc Psychol Personal Sci*. 2012;3(6):651-658.

71. Gartlehner G, Hansen R, Nissman D, Lohr K, Carey TS. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. *Agency Healthc Res Qual*. 2006;(12):1-28.

72. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. *BMJ*. 1999;319(7211):652-653. doi:10.1136/bmj.319.7211.652

73. Green LW. Public health asks of systems science: To advance our evidence-based practice, can you help us get more practice-based evidence? *Am J Public Health*. 2006;96(3):406-409. doi:10.2105/AJPH.2005.066035

74. Bennett WL, Lawrence Cheskin MJ, Wilson RF, et al. Methods for Evaluating Natural Experiments in Obesity: Systematic Evidence Review. *Ann Intern Med*. 2018;(15):1-10. doi:10.7326/M18-0309

75. Sallis JF, Story M, Lou D. Study designs and analytic strategies for environmental and policy research on obesity, physical activity, and diet: recommendations from a meeting of experts. *Am J Prev Med*. 2009;36(2 Suppl):s72-s77. doi:10.1016/j.pmrj.2014.02.014.Lumbar

76. Johnstone AM, Horgan GW, Murison SD, Bremner DM, Lobley GE. Effects of a high-protein ketogenic diet on hunger, appetite, and weight loss in obese men feeding ad libitum. *Am J Clin Nutr*. 2008;87(1):44-55.

77. Libotte E, Siegrist M, Bucher T. The influence of plate size on meal composition. Literature review and experiment. *Appetite*. 2014;82:91-96.

78. Sharp D, Sobal J. Using plate mapping to examine sensitivity to plate size in food portions and meal composition among college students. *Appetite*. 2012;59(3):639-645.

79. Barton S. Which clinical studies provide the best evidence? *Br Med J*. 2000;321:255-256.

80. Kroke A, Boeing H, Rossnagel K, Willich SN. History of the concept of “levels of evidence” and their current status in relation to primary prevention through lifestyle interventions. *Public Health Nutr*. 2004;7(2):279-284. doi:10.1079/PHN2003535

81. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018;210(October 2017):2-21. doi:10.1016/j.socscimed.2017.12.005

82. Cook TD, Campbell DT. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton-Mifflin; 1979.

83. Ioannidis JPA. Randomized controlled trials: Often flawed, mostly useless, clearly

indispensable: A commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210(April):53-56. doi:10.1016/j.socscimed.2018.04.029

84. Gelman A. Benefits and limitations of randomized controlled trials: A commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210(April):48-49. doi:10.1016/j.socscimed.2018.04.034

85. Ioannidis JPA. Why Most Clinical Research Is Not Useful. *PLoS Med*. 2016;13(6):1-10. doi:10.1371/journal.pmed.1002049

86. Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *J Policy Anal Manag*. 2008;27(4):724-750.

87. Shadish WR, Cook TD. The Renaissance of Field Experimentation in Evaluating Interventions. *Annu Rev Psychol*. 2008;60(1):607-629. doi:10.1146/annurev.psych.60.110707.163544

88. Bennett WL, Lawrence Cheskin MJ, Wilson RF, et al. Methods for Evaluating Natural Experiments in Obesity: Systematic Evidence Review. 2017;(15). doi:10.7326/M18-0309

89. Wansink B, Payne CR. Eating Behavior and Obesity at Chinese Buffets. *Obesity*. 2008;16(8):1957-1960. doi:10.1038/oby.2008.286

90. Scheibehenne B, Todd PM, Wansink B. Dining in the dark. The importance of visual cues for food consumption and satiety. *Appetite*. 2010;55(3):710-713. doi:10.1016/j.appet.2010.08.002

91. Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: A systematic review of naturally occurring experiments. *Obes Rev*. 2015;16(5):362-375. doi:10.1111/obr.12269

92. Hollands GJ, Shemilt I, Marteau TM, et al. Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. *BMC Public Health*. 2013;13. doi:10.1186/1471-2458-13-1218

93. Haws KL, Liu PJ, Redden JP, Silver HJ. Exploring the Relationship Between Varieties of Variety and Weight Loss: When More Variety Can Help People Lose Weight. *J Mark Res*. 2017;54(4):619-635. doi:10.1509/jmr.15.0206

94. Wansink B. Change their choice! Changing behavior using the CAN approach and activism research. *Psychol Mark*. 2015;32(5):486-500.

95. Bucher T, Collins C, Rollo ME, et al. Nudging consumers towards healthier choices: a systematic review of positional influences on food choice. *Br J Nutr*. 2016;115(12):2252-2263. doi:10.1017/S0007114516001653

96. Zhong Y, Auchincloss AH, Lee BK, Kanter GP. The Short-Term Impacts of the Philadelphia Beverage Tax on Beverage Consumption. *Am J Prev Med*. 2018;55(1):26-34. doi:10.1016/j.amepre.2018.02.017

97. Freedhoff Y, Hall KD. Weight loss diet studies: we need help not hype. *Lancet*. 2016;388(10047):849-851. doi:10.1016/S0140-6736(16)31338-1

98. Warren RP. *Flood: A Romance of Our Time*. NY, NY: Random House; 1964.

99. Guerin B. Mere presence effects in humans: A review. *J Exp Soc Psychol*. 1986;22(1):38-77. doi:10.1016/0022-1031(86)90040-5

100. Orne MT. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *Am Psychol*. 1962;17(11):776–783. doi:10.1037/h0043424

101. Sawyer AG. Demand artifacts in laboratory experiments in consumer research. *J Consum Res*. 1975;1(4):20-30.

102. Sharpe D, Whelton WJ. Frightened by an old scarecrow: The remarkable resilience of demand characteristics. *Rev Gen Psychol*. 2016;20(4):349-368. doi:10.1037/gpr0000087

103. Allpert FH. *Social Psychology*. Cambridge, MA: Riverside Press; 1924.

104. Bond CF, Titus LJ. Social facilitation: A meta-analysis of 241 studies. *Psychol Bull*. 1983;94(2):265-292. doi:10.1037/0033-2909.94.2.265

105. Oviatt DP. SOCIAL FACILITATION AND MOTOR PERFORMANCE: A META-ANALYSIS. 2005.

106. Halperin I, Pyne DB, Martin DT. Threats to internal validity in exercise science: A review of overlooked confounding variables. *Int J Sports Physiol Perform*. 2015;10(7):823-829. doi:10.1123/IJSPP.2014-0566

107. Herman CP. The social facilitation of eating. A review. *Appetite*. 2015;86:61-73. doi:10.1016/j.appet.2014.09.016

108. Holden SS, Zlatevska N, Dubelaar C. Whether Smaller Plates Reduce Consumption

Depends on Who's Serving and Who's Looking: A Meta-Analysis. *J Assoc Consum Res*. 2016;1(1):134-146. doi:10.1086/684441

109. Robinson E, Hardman CA, Halford JCG, Jones A. Eating under observation : a systematic review and meta-analysis of the effect that heightened awareness of observation has on laboratory measured energy intake 1 , 2. *Am J Clin Nutr*. 2015;102(2):324-337. doi:10.3945/ajcn.115.111195.1

110. Zlatevska N, Dubelaar C, Holden SS. Sizing Up the Effect of Portion Size on Consumption: A Meta-Analytic Review. *J Mark*. 2014;78(3):140-154.

111. Robinson E, Bevelander KE, Field M, Jones A. Methodological and reporting quality in laboratory studies of human eating behavior. *Appetite*. 2018;125(February):0-1. doi:10.1016/j.appet.2018.06.037

112. Robinson E. I'm watching you. Why we might need to lie more to participants in eating behaviour experiments. *Appetite*. 2014;83:343. doi:10.1016/j.appet.2014.06.036

113. Katz DL. School-Based Interventions for Health Promotion and Weight Control: Not Just Waiting on the World to Change. *Annu Rev Public Health*. 2009;30(1):253-272. doi:10.1146/annurev.publhealth.031308.100307

114. Roberto CA, Pomeranz JL, Fisher JO. The need for public policies to promote healthier food consumption: A comment on Wansink and Chandon (2014). *J Consum Psychol*. 2014;24(3):438-445. doi:10.1016/j.jcps.2014.03.001

115. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research.

JAMA. 2005;294(2):218-228.

116. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):0696-0701.

117. Stanley TD, Carter EC, Doucouliagos H. What Meta-Analyses Reveal about the Replicability of Psychological Research. *Psychol Bull*. 2018. doi:10.1037/bul0000169

118. Cohen J. The world is round (p<.05). *Am Psychol*. 1994;49(12):997-1003. doi:10.1037/0003-066X.49.12.997

119. Gill J. The insignificance of null hypothesis significance testing. *Polit Res Q*. 1999;52(3):647-674.

120. George BJ, Beasley TM, Brown AW, et al. Common scientific and statistical errors in obesity research. *Obesity*. 2016;24(4):781-790. doi:10.1002/oby.21449

121. McShane BB, Gal D. Blinding Us to the Obvious? The Effect of Statistical Training on the Evaluation of Evidence. *Manage Sci*. 2016;62(6):1707-1718. doi:10.1287/mnsc.2015.2212

122. McShane BB, Gal D. Blinding Us to the Obvious? The Effect of Statistical Training on the Evaluation of Evidence. *Asia-Pacific Adv Consum Res*. 2015;11:202-203. doi:43008804

123. Bueno NB, De Melo ISV, De Oliveira SL, Da Rocha Ataide T. Very-low-carbohydrate ketogenic diet v. low-fat diet for long-term weight loss: A meta-analysis of Randomised controlled trials. *Br J Nutr*. 2013;110(7):1178-1187. doi:10.1017/S0007114513000548

124. Hession M, Rolland C, Kulkarni U, Wise A, Broom J. Systematic review of randomized controlled trials of low-carbohydrate vs. low-fat/low-calorie diets in the management of obesity and its comorbidities. *Obes Rev*. 2009;10(1):36-50. doi:10.1111/j.1467-789X.2008.00518.x

125. Feinman RD. Intention-to-Treat. What is the question? *Nutr Metab*. 2009;6:1-7. doi:10.1186/1743-7075-6-1

126. Cumming G. The New Statistics: Why and How. *Psychol Sci*. 2014;25(1):7-29. doi:10.1177/0956797613504966

127. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155-159. doi:10.1037/0033-2909.112.1.155

128. Bond CF, Wiitala WL, Richard FD. Meta-Analysis of Raw Mean Differences. *Psychol Methods*. 2003;8(4):406-418. http://www.ncbi.nlm.nih.gov/pubmed/14664679.

129. Borgeraas H, Johnson LK, Skattebu J, Hertel JK, Hjelmesæth J. Effects of probiotics on body weight, body mass index, fat mass and fat percentage in subjects with overweight or obesity: a systematic review and meta-analysis of randomized controlled trials. *Obes Rev*. 2018;19(2):219-232. doi:10.1111/obr.12626

130. Lynch JG, Bradlow ET, Huber JC, Lehmann DR. Reflections on the replication corner: In praise of conceptual replications. *Int J Res Mark*. 2015;32(4):333-342. doi:10.1016/j.ijresmar.2015.09.006

131. Nosek B, Aarts A, Anderson C, Anderson J, Kappes H, Collaboration OS. Estimating the

reproducibility of psychological science. *Science (80-)*. 2015;349(6251):aac4716-1. doi:10.1126/science.aac4716

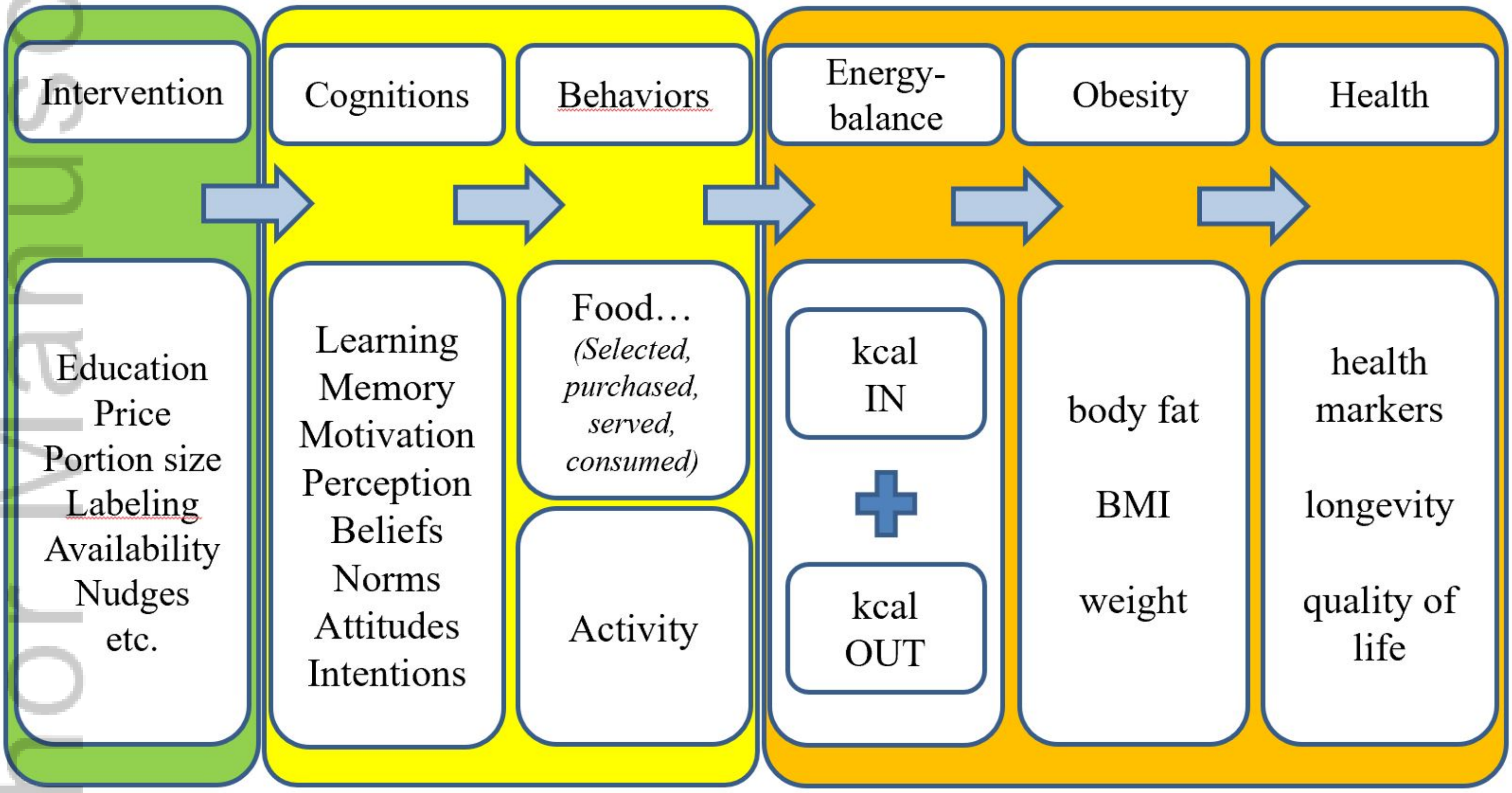
132. Assmus G, Farley JU, Lehmann DR. How Advertising Affects Sales: Meta-Analysis of Econometric Results. *J Mark Res*. 1984;21(1):65-74. doi:10.1017/CBO9781107415324.004

133. Eisend M. Meta-Analysis in Advertising Research. *J Advert*. 2017;46(1):21-35. doi:10.1080/00913367.2016.1210064

134. Rosling H, Rosling O, Rosling-Ronnland A. *Factfulness: Ten Reasons We're Wrong about the World -- and Why Things Are Better than You Think*. 1st ed. NY, NY: Flatiron Books; 2018.

135. Hill JO, Wyatt HR, Reed GW, Peters JC, Hill J. and the Environment : Where Do We Go from Here ? *Science (80-)*. 2011;299(5608):853-855. doi:10.1126/science.1079857

Figure 1: Chain of effects from intervention to health



OBR_13141_Figure 1.tif