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

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/obr.13141

This article is protected by copyright. All rights reserved.
Carolina Werle, Grenoble Ecole de Management, Grenoble, France & IREGE, Université Savoie Mont Blanc, Chambéry, 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 multifactorial 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.\textsuperscript{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.\textsuperscript{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\textsuperscript{12–15} and arises as an outcome of an exceedingly complex system as illustrated in the Foresight systems map.\textsuperscript{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”. 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. Others offer similar views that the evidence we have about what causes obesity and interventions might reduce obesity is “both equivocal and largely circumstantial”.

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” 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.\textsuperscript{19-23}

\textbf{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).

\begin{figure*}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Figure 1: Chain of effects.}
\end{figure*}

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. 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”. This assessment is supported by evidence that behavior change research provides relatively “modest” effects with an $R^2$ of .3 or less. 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.

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.
While most researchers acknowledge the potential for compensation, the degree of compensation is contentious. Some see compensation as incomplete or partial.\textsuperscript{35–38} Others claim that compensation completely offsets any reduction in calories consumed\textsuperscript{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.\textsuperscript{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.\textsuperscript{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,\textsuperscript{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”,\textsuperscript{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? 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”.

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. 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 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 = \( \frac{\text{kgs}_{\text{weight}}}{\text{meters}_{\text{height}}^2} \)) 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 foods61 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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”,\textsuperscript{73} for research evidence showing what might work in practice.\textsuperscript{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.\textsuperscript{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”.\textsuperscript{73} For instance, in many food studies, participants are given food for free, and sometimes limited from accessing other food.\textsuperscript{76} Self-reported measures of consumption and nutritional intake such as the food frequency questionnaire (FFQ) are unreliable.\textsuperscript{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\textsuperscript{77} or draw the amount of food they would serve themselves on a piece of paper.\textsuperscript{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”.\textsuperscript{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.\textsuperscript{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”; 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. “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”.

The respective limitations of experimental and observational research methods are overcome in recognizing that these designs have complementary values and roles. 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 and public health more broadly. Other hybrid research designs that might be considered include pragmatic clinical trials, quasi-experimental designs, natural experiments and field experiments. 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. 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.
Longer duration

Too many studies purporting to address weight- and obesity-change are based on short term, even one-off eating occasions.\textsuperscript{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]”.\textsuperscript{23} Many researchers are aware of the problem and frequently allude to “a need for high-quality studies … particularly in the longer term”.\textsuperscript{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”.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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”.\textsuperscript{98} In social research sciences, this is the problem of observer effects,\textsuperscript{99} but also includes demand effects.\textsuperscript{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, and have been overlooked in exercise science.

Observer effects undermine the validity of tests of obesity interventions, and it is essential that they are addressed. 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. 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. 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. 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.\textsuperscript{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 $p<.05$.\textsuperscript{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.\textsuperscript{118,119} The problem is that statistical significance is widely misunderstood, misused and leads to a misdirection of research efforts.\textsuperscript{120} As Cohen remarks, the emperor’s new clothes do not exist, and “this naked emperor has been shamelessly running around a long
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. 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 (H₀), i.e., p(D/H₀). 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/H₀) ≠ p(H₀/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 H₁ given the data, i.e., p(H₁/D). Given the data, how likely is it that the theory (H₁) 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 I = 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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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,\textsuperscript{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, 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, 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”.

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 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). As weight change is slow, takes
place over time, and is subject to dynamic physiological processes, 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 \((d<=.5).\) Notwithstanding, health policy will be better guided by a realistic view of effect-sizes rather than
implausible results, 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.
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.\textsuperscript{116,117} The growing acknowledgement of problems of replication in medicine, business, and psychology\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 \textit{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. 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


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


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


This article is protected by copyright. All rights reserved.


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


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


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


This article is protected by copyright. All rights reserved.


Figure 1: Chain of effects from intervention to health
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Holden, SS; Zlatevska, N; Parkinson, J; Cadario, R; Dubelaar, C; Lei, J; Moore, E; Sayarh, N; Van Kerckhove, A; Werle, C

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

Date:
2021-02-01

Citation:

Persistent Link:
http://hdl.handle.net/11343/276254