INSEAD
The Business School for the World®

Working Paper Series

2017/77/MKT

Which Healthy Eating Nudges Work Best? A Meta-Analysis of Field Experiments

Romain Cadario
IESEG, r.cadario@ieseg.fr

Pierre Chandon
INSEAD, pierre.chandon@insead.edu

December 18, 2017

We examine the effectiveness in field settings of seven healthy eating nudges, classified according to whether they influence 1) cognition, via "descriptive nutritional labeling," "evaluative nutritional labeling," or "salience enhancements"; 2) affect, via "hedonic or sensory cues" or "healthy eating prods"; or 3) behavior, via "convenience enhancements" or "plate and portion size changes." Compared with existing univariate meta-analyses, our multivariate three-level meta-analysis of 277 effect sizes controlling for eating behavior, population, and study characteristics yields smaller effect sizes overall. These effect sizes increase as the focus of the intervention shifts from cognition (d=.08, equivalent to -45 kcal/day) to affect (d=.22, -121 kcal) to behavior (d=.35, -186 kcal). Interventions are more effective at reducing unhealthy eating than at increasing healthy eating or reducing total eating. Effect sizes are larger in the US than in other countries, in restaurants or cafeterias than in grocery stores, and in studies including a control group. Effect sizes are similar for food selection vs. consumption, for children vs. adults, and are independent of study duration. Compared to the typical study, one testing the best nudge scenario should expect a fourfold increase in effectiveness, with half due to switching from cognitive to behavioral nudges.

Keywords: Meta-Analysis; Health; Food; Field Experiment; Nudge; Choice Architecture

Electronic copy available at: http://ssrn.com/abstract=3090829

Healthy eating nudges: A « living » meta-analysis

We have created an online tool to allow researchers to correct and update our data, turning it into a "living" meta-analysis that can be continuously updated. Do not hesitate to share this URL to your colleagues! We will periodically update the meta-analysis to include the most recent results. The updated results will be posted online, available to all.

Helpful comments on various aspects of this research were provided by Janet Schwartz, Alexander Chernev, Maria Langlois, Yann Cornil, Gareth Hollands, and those who participated when the authors presented this research at the TCR conference, the BSPA conference, the SJDM conference, the University of Toronto and the University of British Columbia.

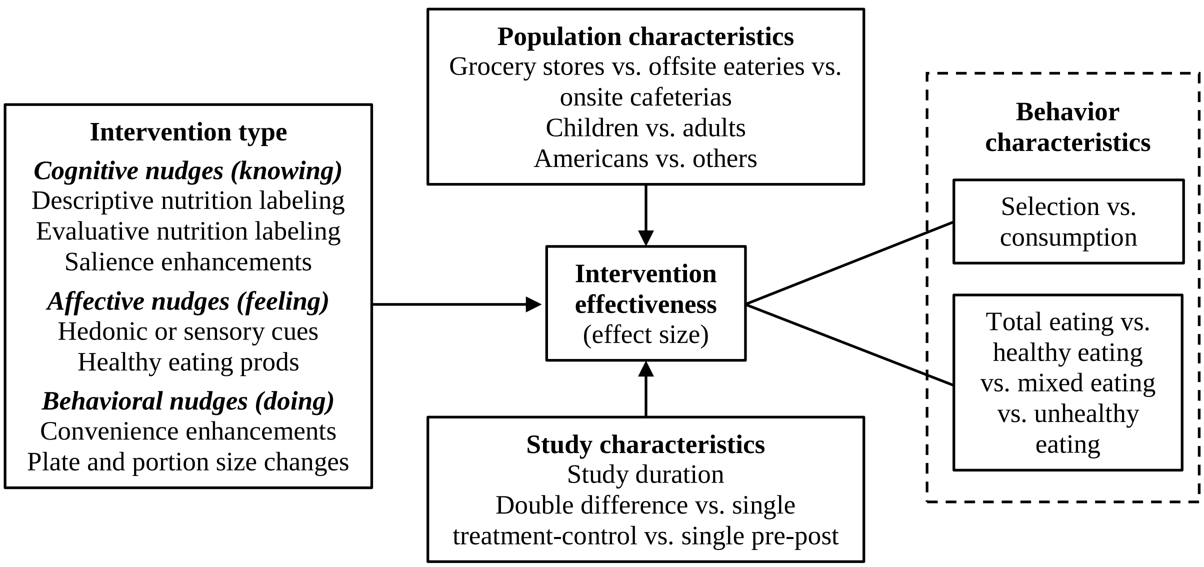
1. Introduction

Unhealthy eating is a key risk factor in non-communicable diseases such as cardiovascular disorders and diabetes, which account for 63% of all deaths worldwide and will cost an estimated US$30 trillion in the next 20 years (Bloom et al. 2012). Traditional approaches to promote healthier eating include economic incentives, such as soda taxes (for a recent review, see Afshin et al. 2017) and nutrition education (for a recent review, see Murimi et al. 2017).

More recently, interest has grown in nudge interventions as a spur to healthier eating. Disappointingly, existing meta-analyses have only found average effect sizes ranging from weak or null (e. g., Cecchini and Warin 2016; Littlewood et al. 2016; Long et al. 2015) to moderate (e.g., Arno and Thomas 2016; Hollands et al. 2015). However, these were based on a small number of studies (e.g., 19 for Long et al. 2015), specific foods (e.g., vegetables for Broers et al. 2017), specific settings (e.g., catering outlets for Nikolaou et al. 2014), or included online or laboratory studies (e.g., Sinclair et al. 2014) where effect sizes tend to be different than in field studies (Holden et al. 2016; Long et al. 2015). The field still lacks a comprehensive meta-analysis that attests to the effectiveness of a broader range of healthy eating nudges in field settings. More important, it lacks a conceptual framework within which to categorize interventions and predict their effectiveness for different consumption behaviors (e.g., healthy eating, unhealthy eating, or total eating) and different consumption settings (e.g., restaurants, cafeterias, or grocery stores), as well as for different populations and study characteristics.

Our study examines the effectiveness of interventions aimed at promoting healthy eating without resorting to either economic incentives or nutrition or health education. These simple, inexpensive, freedom-preserving modifications to the choice environment, defined as “nudges” by Thaler and Sunstein (2009, p.6), refer to “any aspect of the choice architecture that alters people’s behavior in a predictable way (1) without forbidding any options or (2) significantly changing their economic incentives. Putting fruit at eye level counts as a nudge; banning junk food does not." According to this definition, which has been adopted by other influential review papers (e.g., Hollands et al. 2013; Skov et al. 2013), healthy eating nudges encompass a wide variety of interventions, including nutrition labeling in supermarkets, healthy eating cues in cafeterias, and plate and portion size changes in restaurants, but excluding financial incentives like price changes or sales promotions.

Figure 1: Conceptual framework



To achieve this goal, we identify seven types of healthy eating nudges classified in three categories: cognitive, affective, and behavioral. As shown in Figure 1, our framework also accounts for the type of eating behavior (food selection or actual consumption) and distinguishes between healthy eating, unhealthy eating, and total energy intake. It also considers population characteristics such as age (children vs. adults), consumption setting (onsite cafeterias vs. offsite restaurants, cafes vs. grocery stores), and location of the study (US vs. other countries), as well as characteristics such as the duration of the study and its design. We test this framework with a three-level meta-analysis of 277 effect sizes from 81 articles and 87 field studies published until the end of 2016.

Table 1: Comparing Meta-Analyses of Healthy Eating Nudges

	Scale and scope			Method			Categorization of predictors			
Reference	Effect sizes (K)	Articles (N)	Setting	Hypo-theses	Accounting for repeated observations	Model	Intervention type	Consump-tion vs. selection	Healthy vs. un-healthy	Other control variables
This meta-analysis	277	81	Field only	Yes	Yes (3 levels)	Multivariate (16 *df*)	3 pure and 2 mixed types, (7 subtypes)	Yes	Yes	5 study & population characteristics
Arno and Thomas (2016)	42	36	Field & lab	No	No	Intercept only	1	No	No	None
Broers et al. (2017)	14	12	Field & lab	No	No	Intercept only	1	No	No	None
Cecchini and Warin (2016)	31	9	Field & lab	No	No	Univariate	1 (only labeling)	No	Yes	None
Holden et al. (2016)	56	20	Field & lab	No	No	Univariate	1 (only size studies)	Yes	No	Manipulation type, field vs. lab
Hollands et al. (2015)	135	69	Field & lab	No	No	Univariate	1 (only size studies)	Yes	Yes	Manipulation type, age, design
Littlewood et al. (2016)	20	14	Field & lab	No	No	Univariate	1 (only labeling)	Yes	No	None
Long et al. (2015)	23	19	Field & lab	No	No	Univariate	1 (only labeling)	No	No	Design
Nikolaou et al. (2014)	10	6	Field only	No	No	Univariate	1 (only labeling)	No	No	None
Robinson et al. (2014)	15	7	Field & lab	No	No	Intercept only	1 (only size studies)	No	No	None
Sinclair et al. (2014)	42	17	Field & lab	No	No	Univariate	2 (desc. vs. eval. label.)	Yes	No	None
Zlatevska et al. (2014)	104	30	Field & lab	No	No	Univariate	1 (only size studies)	No	Yes	Field vs. lab, age, sex, BMI

As shown in Table 1, our work contributes to the many useful existing meta-analyses in terms of (1) scale and scope, (2) method, and (3) categorization of predictors. In terms of scale and scope, we examine more than twice as many effect sizes as the largest existing meta-analysis. This is achieved despite focusing only on field experiments involving actual food choices (vs. perception, evaluation, or choice intentions) and conducted in field settings (onsite cafeterias, offsite eateries, or grocery stores) rather than in a laboratory or online. This allows us to offer guidance to restaurants, supermarket chains, and foodservice companies who want to help their customers eat more healthily but do not know which intervention will work best in their particular context; and to provide guidance for policy makers who need to forecast the effects that these nudges would have in real-world settings.

Methodologically, our meta-analysis differs from earlier ones on three levels. First, we formulate hypotheses about which healthy eating nudges work best and about the effects of eating behavior and of population and study factors. Second, to reduce the risk of confounds from univariate analyses, we employ a multivariate model incorporating all predictors simultaneously. Third, we include a three-level analysis to take into account the hierarchical structure of our data.

Finally, Table 1 shows that we use a more granular predictor structure compared to existing meta-analyses, which either estimated the effect size of a single type of healthy eating nudge or compared the effect of one single difference (say, descriptive vs. evaluative labeling) and which rarely incorporated behavior, population, and study characteristics.

2. Conceptual framework

2.1. Classifying interventions

The many existing classifications of healthy eating interventions tend to be ad hoc, focusing on mnemonic acronyms rather than theoretical grounding. They have also shown a tendency towards category inflation over the years. For example, Chance et al. (2014) suggested four P's (possibilities, process, persuasion, and person) where Kraak et al. (2017) recommended eight P's (place, profile, portion, pricing, promotion, healthy default picks, prompting or priming, and proximity). Hollands et al. (2013) initially distinguished three nudge categories, but even the simplified version of their more recent TIPPME typology contains 18 categories (Hollands et al. 2017). Unlike the other classifications, Wansink's (2015) CAN model is based on three hypothesized mechanisms of action (convenience, attractiveness, or norms). Unfortunately, many interventions operate via multiple mechanisms, making the CAN model less suitable for a classification than for a conceptual framework. Moreover, none of the existing classifications makes predictions about which type of intervention is most effective.

We draw on the classic tripartite classification of mental activities into cognition, affect, and behavior (or conation), which dates back to eighteenth-century German philosophy (Hilgard 1980). The trilogy of mind has long been adopted in psychology and marketing, both to understand consumer behavior and to predict the effectiveness of marketing actions (Barry and Howard 1990; Breckler 1984; Hanssens et al. 2014; Oliver 1999; Srinivasan et al. 2010). We distinguish between 1) cognitive interventions that seek to influence what consumers know; 2) affective interventions that seek to influence how consumers feel, without necessarily changing what they know; and 3) behavioral interventions that seek to influence what consumers do, without necessarily changing what they know or how they feel.

Within each type of intervention, we further distinguish two or three subtypes that share similar characteristics and that have been tested by enough studies to enable a meaningful meta-analysis at their level. This subcategorization is based on existing classifications, such as the distinction between descriptive and evaluative nutritional labeling (Fernandes et al. 2016; Sinclair et al. 2014). Table 2 provides definitions and illustrations of the seven types of intervention and lists the studies that tested them. In the analysis section, we explain how we account for the fact that some studies implemented multiple types of interventions at once.

Cognitive interventions. As described in Table 2, we first grouped the three types of cognitive interventions that aim to nudge consumers by informing them about the healthiness of their food options. The first type, "descriptive nutritional labeling," provides calorie count or information about other nutrients, be it on menus or menu boards in restaurants, or on labels on the food packaging or near the foods in self-service cafeteria and grocery stores. The second type, "evaluative nutritional labeling," typically (but not always) provides nutrition information but also helps consumers interpret it through color coding (e.g., red, yellow, green as nutritive value increases) or by adding special symbols or marks (e.g., heart-healthy logos or smileys on menus). Although the third type, "salience enhancement," does not directly provide health or nutrition information, it is a cognitive intervention because it informs consumers about the availability of healthy options by increasing their visibility on grocery or cafeteria shelves (e.g., by placing healthy options at eye level and unhealthy options on the bottom shelves) or on restaurant menus (e.g., by placing healthy options on the first page and burying unhealthy ones in the middle).

Affective interventions. The first type of affective interventions, dubbed "hedonic or sensory cues," seeks to increase the hedonic or sensory appeal of healthy options by using vivid sensory descriptions (e.g., "amazing broccoli") or attractive displays, photos, or containers (e.g., "pyramids of fruits"). To date, no field experiment has sought to reduce hedonic or sensory expectations for unhealthy options by using disparaging descriptions or unattractive photos. These interventions are affective because, rather than focusing on informing consumers about the nutritional quality of food options or their likely health impact, they focus on the more affective hedonic or sensory consequences of eating the food.

Table 2: Categorization of nudge interventions

Intervention	Target: Healthy eating (k = 170)	Target: Unhealthy eating (k = 71)
	Knowing: Cognitive interventions	
Descriptive nutritional labeling (k = 31)	Calorie or nutrition labeling (Auchincloss et al. 2013; Bollinger et al. 2011; Brissette et al. 2013; Chu et al. 2009; Downs et al. 2013; Dubbert et al. 1984; Dumanovsky et al. 2011; Elbel et al. 2011; Elbel et al. 2009; Elbel et al. 2013; Ellison et al. 2013; Finkelstein et al. 2011; Krieger et al. 2013; Pulos and Leng 2010; Roberto et al. 2010; Tandon et al. 2011; Vanderlee and Hammond 2014; Webb et al. 2011)	
Evaluative nutritional labeling (k = 34)	Green stickers, smileys, "heart healthy" logos (Cawley et al. 2015; Ensaff et al. 2015; Gaigi et al. 2015; Hoefkens et al. 2011; Kiesel and Villas-Boas 2013; Levin 1996; Levy et al. 2012; Ogawa et al. 2011; Olstad et al. 2015; Reicks et al. 2012; Thorndike et al. 2014; Thorndike et al. 2012)	Red stickers next to unhealthier options (Crockett et al. 2014; Hoefkens et al. 2011; Levy et al. 2012; Olstad et al. 2015; Shah et al. 2014; Thorndike et al. 2014; Thorndike et al. 2012)
Salience enhancements (k = 25)	Healthier options more visible: e.g., eye-level shelf position, transparent containers, placed near cash register (Bartholomew and Jowers 2006; Cohen et al. 2015; Ensaff et al. 2015; Foster et al. 2014; Gamburzew et al. 2016; Geaney et al. 2016; Hanks et al. 2013; Kroese et al. 2016; Levy et al. 2012; Meyers and Stunkard 1980; Perry et al. 2004; Policastro et al. 2017)	Unhealthier options less visible (not eye-level shelf positions, middle of the menu), previous unhealthier consumption more visible: e.g., leftover chicken wings un-bussed (Bartholomew and Jowers 2006; Baskin et al. 2016; Dayan and Bar-Hillel 2011; Meyers and Stunkard 1980; Wansink and Payne 2007)
	Feeling: Affective interventions	
Hedonic or sensory cues (k = 12)	Vivid hedonic descriptions (e.g., "amazing broccoli") or attractive displays, photos, or containers (Cohen et al. 2015; Ensaff et al. 2015; Hanks et al. 2013; Morizet et al. 2012; Olstad et al. 2014; Perry et al. 2004; Wansink et al. 2012; Wilson et al. 2016b)	
Healthy eating prods (k = 35)	Written or oral injunction to choose healthier options: e.g., "Make a fresh choice" or "Revitalize yourself" (Buscher et al. 2001; Cohen et al. 2015; Ensaff et al. 2015; Hanks et al. 2013; Hubbard et al. 2015; Perry et al. 2004; Schwartz 2007; van Kleef et al. 2015)	Written or oral injunctions to change unhealthy choices: e.g., "Your meal doesn't look balanced" or "Would you like to take a half portion?" (Freedman 2011; Miller et al. 2016; Mollen et al. 2013; Schwartz et al. 2012)
	Doing: Behavioral interventions	
Convenience enhancements (k = 56)	Healthier options are easier to select or consume: e.g., default choice, convenient utensils, "grab and go" line, placed earlier in cafeteria line, pre-sliced, pre-portioned, or pre-served (Adams et al. 2005; Buscher et al. 2001; Cohen et al. 2015; de Wijk et al. 2016; Elsbernd et al. 2016; Goto et al. 2013; Hanks et al. 2012; Lachat et al. 2009; Olstad et al. 2014; Redden et al. 2015; Rozin et al. 2011; Steenhuis et al. 2004; Tal and Wansink 2015; Wansink et al. 2016; Wansink and Hanks 2013; Wansink et al. 2013; Wilson et al. 2016b)	Unhealthier options are less convenient to select or consume: e.g., placed later in cafeteria line when tray is fuller, less accessible or harder to reach, less convenient serving utensils (Hanks et al. 2012; Mishra et al. 2012; Rozin et al. 2011; Wansink and Hanks 2013)
Plate and portion size change (k = 17)	Larger plates for healthier options (DiSantis et al. 2013)	Smaller plates or portions for unhealthy options (Diliberti et al. 2004; DiSantis et al. 2013; Freedman and Brochado 2010; van Ittersum and Wansink 2013; Wansink and Kim 2005; Wansink and van Ittersum 2013; Wansink et al. 2006; Wansink et al. 2014)

The second type of affective interventions, dubbed “healthy eating prods,” provides direct injunctions, orally or in writing, to eat better. This can be done either by prodding people to choose a healthy option (e.g., “Make a fresh choice” or “Revitalize yourself by snacking on a fresh basket of crisp red peppers, juicy tomatoes, and crunchy carrots. Easy to eat on the run!”) or to change their unhealthy choices (e.g., “Your meal doesn’t look like a balanced meal” or “Would you like to take half a portion of your side dish?”). These interventions are affective because, rather than changing beliefs, they seek to directly change people’s eating goals from taste to health through injunctions. By asking lunch ladies or cashiers to comment on people’s food choices, healthy eating prods also rely on the power of interpersonal communication and norms.

Behavioral interventions. The third group consists of two types of interventions that aim to impact people’s eating behaviors without necessarily influencing what they know or how they feel, and therefore often without people being aware of their presence. “Convenience enhancements” make it easier for people to select healthy options (e.g., by making them the default option or placing them in faster “grab & go” cafeteria lines) or to consume them (e.g., by pre-slicing fruits or pre-serving vegetables), or make it more cumbersome to select or consume unhealthy options (e.g., by placing them later in the cafeteria line when trays are already full or by providing less convenient serving utensils). The second type, which we call “plate and portion size changes,” modifies the size of the plate, bowl, or glass, or the size of pre-plated portions, either increasing the healthy options they contain or, most commonly, reducing unhealthy options.

Hypotheses. The tripartite categorization of healthy eating nudges allows us to make predictions about their effectiveness. First, in the domain of food, cognitive factors tend to be less predictive of choice than affective factors, which have been shown to strongly influence even restrained eaters who eat according to cognitive rules (Macht 2008). For example, when asked about what drives their food choices, American and European consumers place affective factors like taste in first position, well ahead of cognitive factors like nutrition or weight control (Glanz et al. 1998; Januszewska et al. 2011). Even interventions that successfully change beliefs about the health consequences of behaviors often fail to lead to meaningful behavioral changes (Carpenter 2010; Sniehotta et al. 2014).

Second, because eating is largely habitual and prone to self-regulation failures, affective factors tend to be less predictive of food choices than behavioral factors (Herman and Polivy 2008; Ouellette and Wood 1998). For example, directly changing the eating environment (e.g., avoiding exposure to tempting food) is a more successful self-control strategy than cognitive strategies (e.g., not looking at the food, thinking about its nutrition content) or affective strategies such as relying on willpower (Duckworth et al. 2016; Wansink and Chandon 2014). Similar conclusions were reached in a large study of the drivers of the sales elasticity for 74 mostly food brands (Srinivasan et al. 2010). This study found that changes in distribution (a behavioral intervention) have a larger impact than changes in advertising (a cognitive or affective intervention) and that affective changes in liking are more predictive of brand choice than cognitive changes in awareness. We therefore expect that the effectiveness of healthy eating interventions increases as their focus switches from cognition to affect and to behavior.

2.2. The role of eating behavior type

As shown in Figure 1, we differentiate between different types of eating behaviors. First, some studies measure actual food consumption whereas others only capture food selection (e.g., the purchase of food in a grocery store, cafeteria, or restaurant) without knowing whether the food was entirely consumed. One may expect larger effect sizes for selection than for consumption if some consumers, after being nudged to try a healthier food, are disappointed by its taste and only consume part of it. On the other hand, people usually have a stronger preference for *what* to eat relative to *how much* to eat (Wansink and Chandon 2014). Interventions aiming to influence consumption once the food has already been selected, like plate and portion size changes, may be more effective than those that attempt to influence what food is selected. This would imply larger effect sizes when measuring actual consumption than selection. For these reasons, we expect similar effect sizes for food selection and consumption. This hypothesis is consistent with existing meta-analyses of specific types of healthy eating nudges, which found no differences between studies measuring selection and those measuring actual consumption (Holden et al. 2016; Hollands et al. 2015; Littlewood et al. 2016; Sinclair et al. 2014).

The second aspect of eating behavior that we examine is whether studies measure total eating (e.g., the total number of calories of the food selected or consumed) or focus on the selection or consumption of healthy or unhealthy foods. We expect smaller effect sizes when the dependent variable is total amount of food ordered or consumed for two reasons. The first is that people must eat—it is difficult, psychologically and physiologically, to sustain an imbalance between energy intake and energy expenditure. In contrast, people have more flexibility in choosing how to allocate their total calorie intake between healthy and unhealthy foods. Furthermore, healthy foods have calories too, so replacing unhealthy food with healthier options—although clearly a form of healthier eating—does not necessarily mean a reduction in the total quantity of food ordered or consumed. Total eating therefore underestimates certain forms of healthier eating that are better captured by measuring changes in healthy and unhealthy eating separately. This hypothesis is consistent with an existing meta-analysis of interpretive nutrition labels, which found that they are more effective in helping consumers in choosing healthier products than in changing total intake (Cecchini and Warin 2016).

Finally, we hypothesize that interventions aimed at reducing unhealthy eating have a stronger effect size than those aimed at promoting healthy eating. This prediction is based on the fact that more than two thirds of Americans are overweight or obese, and about half of the latter are actively trying to lose weight in any given year (Snook et al. 2017). Dieters should therefore be particularly receptive to interventions that help reduce calorie intake, which is most effectively accomplished by reducing the consumption of unhealthy foods rather than by increasing healthy food consumption. Indeed, dieters and overweight consumers may be wary of increasing their intake of foods presented as "healthy," which often actually have a high energy density (Chernev 2011; Wansink and Chandon 2006). More generally, people often exhibit dynamically inconsistent preferences, choosing unhealthy food in the short term and regretting it later (Prelec and Loewenstein 1998; Wertenbroch 1998). Interventions that help people resist the temptation of unhealthy food should therefore be particularly attractive to the large number of people who have a long-term healthy eating goal and are aware that they need help resisting unhealthy foods despite their best intentions. Our hypothesis is consistent with the results of two existing meta-analyses, which found that the effectiveness of plate and portion size changes is higher for unhealthy foods compared to healthy foods (Hollands et al. 2015; Zlatevska et al. 2014).

2.3. The role of population characteristics

We consider the influence of three population characteristics. We distinguish between studies conducted in onsite eating settings (e.g., university or worksite cafeterias), offsite eating eateries (e.g., restaurants, cinemas, cafes), and grocery stores. We expect weaker effects in grocery stores compared to the other two settings. This is because it should be easier to respond to healthy eating nudges when choosing for oneself, from among a limited number of options, and for a single immediate consumption in a cafeteria or a restaurant, than when choosing for the entire family, from among a huge variety of tempting options, and for multiple consumption occasions in a grocery store. This hypothesis is consistent with research showing that uncertainty about future preferences (when buying for the entire family, for example) increases the variety of choices (Walsh 1995), thereby mitigating the effects of nudges. It is also consistent with the systematic review conducted by Seymour et al. (2004), which concluded that healthy eating nudges have weaker effects in grocery stores than in restaurants or in university or worksite cafeterias.

Second, prior research has established that adults are more interested in nutrition than children (Croll et al. 2001). Adults are also more sensitive than children to portion size changes (Hollands et al. 2015; Zlatevska et al. 2014). We thus expect them to be more responsive to all types of interventions than children. Finally, we expect to find higher effect sizes in studies conducted in the US than in other countries. The higher proportion of overweight people in the US, the larger size of portions there (Rozin et al. 2003), Americans' higher interest in and knowledge of the health consequences of eating (Rozin et al. 1999), and their greater reliance on external than internal cues when making food decisions (Wansink et al. 2007) should all make interventions more effective in the US than in the other Western countries in our sample.

2.4. The role of study characteristics

We expect two study characteristics to influence the effectiveness of healthy eating nudges. First, food interventions vary in duration from single-exposure studies of one consumption occasion to longitudinal interventions lasting several months. In field experiments, treatment effects usually decay over time as people revert to habitual behavior (Brandon et al. 2017). We therefore expect to find larger effect sizes for shorter studies than for longer ones.

Finally, we distinguish between studies using a pre-post design without control, those using a single-difference treatment-control design, and those using a double-difference design. Although designs with stronger levels of control should have a lower statistical bias in the estimation of the effect size, the type of design itself should not influence the size of the effect. Therefore, we cannot formulate hypotheses about the effect of design on effect sizes, and include this factor simply as a control variable.

3. Data collection

3.1. Inclusion criteria

Appendix A provides detailed information on the search strategy, including the SPICE (Setting, Population, Intervention, Comparison, Valuation) framework (Booth 2006) for the selection of keywords and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Moher et al. 2009) flow diagram showing the number of articles included and excluded. Briefly, we searched for relevant articles published in scholarly journals until December 2016 through keyword searches on Science Direct, PubMed, and Google Scholar. We also examined all the references from 11 meta-analyses (Arno and Thomas 2016; Broers et al. 2017; Cecchini and Warin 2016; Holden et al. 2016; Hollands et al. 2015; Littlewood et al. 2016; Long et al. 2015; Nikolaou et al. 2014; Robinson et al. 2014; Sinclair et al. 2014; Zlatevska et al. 2014) and 7 systematic reviews (Bucher et al. 2016; Hollands et al. 2013; Nornberg et al. 2016; Roy et al. 2015; Skov et al. 2013; Thapaa and Lyford 2014; Wilson et al. 2016a). Both authors developed the protocol detailing the search and inclusion criteria, coding categories for predictors, and computation rules. The first author was trained to code all the studies and was responsible for extracting data. The second author checked the results, and disagreements were solved through discussion.

To be included in the meta-analysis, the study had to test a nudge intervention consistent with our definition (e.g., not a price change nor a nutrition education campaign). Since our focus was on pure nudges, studies (or conditions in studies) combining nudges with changes in economic incentives or education efforts were not included. The intervention had to be tested in a field experiment in which participants were not aware of the intervention, and thus not in a laboratory or online setting. This is important because previous reviews found marked differences between studies conducted in the field and those conducted in a laboratory or online (Long et al. 2015), and between studies conducted with aware or unaware participants (Holden et al. 2016). Finally, the dependent variable of the study had to provide an objective measure of food selection or consumption (either in weight or energy). We rejected studies relying on consumption intentions.

Overall, the meta-analysis includes 277 effect sizes derived from 87 studies published in 81 articles. The number of observations per study ranged from 36 to 100 million, with a median of 1,168. Even after excluding the two outlier studies, one with 100 million transactions (Bollinger et al. 2011) and one with 29 million transactions (Nikolova and Inman 2015), the meta-analysis represents more than 4.1 million observations.

3.2. Effect sizes calculations

We calculated the effect sizes of studies with a binary outcome, such as the number of participants who chose a healthy option, by computing the log odds ratio or by obtaining it directly from the paper in the few cases when it was available. We computed the odds ratio as the odds of a healthy selection in the treatment group divided by the odds of a healthy selection in the control group. We then computed its standard error. We calculated the effect sizes of studies with a continuous dependent variable, such as unhealthy food intake, by computing the standardized mean difference, except in the few instances when the standardized mean difference, also known as Cohen's (1988) *d*, was already reported in the paper. We computed the *d* value as the mean difference in consumption between the treatment and control condition, divided by the pooled standard deviation. Given that we had two different effect-size metrics, we converted the log odds ratio into *d* using the formula proposed by Borenstein et al. (2009). After the conversion, the 135 effects sizes originally calculated as log odds ratio were not statistically different from the 142 effect sizes computed as *d* (p = .30). Hence, we report Cohen's *d* in the paper because it is the most common measure and because it allows direct comparisons with other meta-analyses.

The most common unreported data was the sample size per experimental condition (e.g., intervention vs. control). When only the total sample was reported, we divided the total number of observations by the number of conditions. When only the number of observations in the control group or in the intervention group was reported, we used the same number for the other group. Whenever several assumptions were possible, we conservatively chose the assumption that yielded the smaller effect size or the largest standard error.

When results were reported separately for each food in the same study (e.g., Ensaff et al. 2015), we calculated separate effect sizes per food and accounted for their dependence in the statistical analysis. We also computed separate effect sizes for the few studies (e.g., Schwartz 2007) that measured both food selection (e.g., putting a food item on a cafeteria tray) and consumption (e.g., how much of it was consumed). When a study had a two-phase intervention (e.g., one intervention during the first phase and then another intervention during a later phase, Thorndike et al. 2012), we computed separate effect sizes for each phase and compared both phases to the baseline period. When a study tested multiple interventions separately (e.g., Mollen et al. 2013), we computed separate effect sizes for each intervention. Because only two studies reported results separately for men and women (Baskin et al. 2016; Wansink and Payne 2007), we could not examine the role of gender in the meta-analysis and calculated the average effect size for these two studies.

3.3. Coding

Each intervention was categorized into one of the seven types discussed earlier. Field experiments that implemented multiple interventions at once were treated separately. There are not enough studies testing each possible combination of cognitive, affective, and behavioral interventions (for example, only one study mixed a cognitive and a behavioral intervention), and so we had to rely on an ad hoc coding of “mixed interventions” based on their frequency in our sample.

The first type of mixed intervention consists of studies mixing cognitive interventions with affective and/or behavioral interventions (e.g., descriptive nutrition labeling and hedonic or sensory cues). We named them “mixed: cognitive present.” The second type of mixed interventions consists of studies combining affective and behavioral interventions. We named them “mixed: cognitive absent.” We therefore have five categories for behavioral interventions: three for pure cognitive, affective, and behavioral interventions and two for mixed interventions (mixed: cognitive present and mixed: cognitive absent).

When the dependent variable of the study was the total amount of food selected or consumed, it was categorized as “total eating.” When it was the selection or consumption of fruits, vegetables, and water, or foods color-coded green in the study, we categorized it as “healthy eating.” We categorized the selection or consumption of calorie-dense and nutrient-poor foods such as desserts or sodas and those color-coded red in studies as “unhealthy eating.” We created a fourth category (“mixed eating”) for foods that could not be categorized as healthy or unhealthy or which were color-coded yellow by the researcher (rather than green or red). Because our goal is to examine healthy eating, we reverse coded the effect sizes for unhealthy eating and for total eating.

We coded population characteristics according to where the study was conducted (school or workplace onsite cafeterias; offsite restaurants, cinemas, or cafes; or grocery stores). We coded whether the participants were children or adults. Finally, we distinguished between studies conducted in the United States and those conducted in other countries (Belgium, Canada, France, Ireland, Israel, Japan, Netherlands, and the United Kingdom). The number of studies in these other countries was too low to enable a more refined level of analysis.

Regarding study characteristics, we measured the duration of the treatment as the number of weeks of the intervention period. For example, if a study with a pre-post design measured food choices in the 4 weeks prior to the intervention and in the 2 weeks after the intervention was implemented, study duration is coded as 2 weeks. Because all the interventions that we studied were applied continuously, duration captures—in the extreme case—the difference between the effects of a single exposure to the nudge on a single food choice and the effects of repeated exposures to the nudge over multiple food choices over time.

We distinguished between "double-difference" designs (which assigned participants to two independent control and treatment conditions, with observations before and after the intervention), "single-difference treatment-control" designs (which assigned respondents to two independent control and treatment conditions), and "single-difference pre-post" designs (which used a pre-post study design without a control group, comparing observations before and after the intervention). Note that all are quasi-experiments because the randomization was not done at the participant level but at the level of the store, restaurant, cafeteria, or at best, cafeteria line.

4. Analyses and results

As indicated in Table 1, the 11 existing meta-analyses used a standard two-level meta-analytical model (Borenstein et al. 2009). In contrast, we used a three-level model (Cheung 2014), which accounts for the fact that some observations come from the same field experiment (e.g., studies testing two types of interventions or measuring their impact on healthy and unhealthy foods separately). We estimated a mixed-effects three-level meta-analytic model with the "metafor" R package provided in Viechtbauer (2010), via maximum likelihood.

4.1. Average meta-analytical effect: Intercept-only model

Let y_{ij} be the i^{th} effect size in the j^{th} study. The equations from the three levels are:

$$y_{ij} = \lambda_{ij} + e_{ij} \tag{1}$$

$$\lambda_{ij} = \kappa_j + u_{(2)ij} \tag{2}$$

$$\kappa_j = d_0 + u_{(3)j} \tag{3}$$

where λ_{ij} is the true effect size and $\mathrm{Var}(e_{ij}) = v_{ij}$ is the known sampling variance in the i^{th} effect size in the j^{th} study. κ_j is the average effect size in the j^{th} study and $\mathrm{Var}(u_{(2)ij}) = \tau^2_{(2)}$ captures the heterogeneity in effect sizes between different eating behaviors (e.g., selection or consumption, healthy or unhealthy food) within the same study, when more than one outcome was measured. d_0 is the meta-analytic effect size estimated across all studies, and $\mathrm{Var}(\mathrm{u}_{(3)\mathrm{j}}) = \tau^2_{(3)}$ captures the heterogeneity between studies after controlling for the presence of multiple observations at level 2. The three equations can be combined as follows:

$$y_{ij} = d_0 + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{4}$$

We assessed the magnitude of effect size heterogeneity through the I^2 index (Higgins and Thompson 2002). We also report the decomposition of heterogeneity within-studies $I^2_{(2)}$ and between-studies $I^2_{(3)}$ as derived in Cheung (2014). Heterogeneity is considered to be low if the I^2 index is below 25%, medium if it is between 25% and 75%, and high if it is above 75% (Higgins and Thompson 2002).

The standard two-level model yields a statistically significant average effect size (d = .22, z = 13.05, p < .001) with a very large amount of heterogeneity (I^2 = 99.9%). The proposed three-level model fits the data significantly better than the two-level model ($\chi^2(1)$ = 81, p < .001) and yields a slightly larger estimate of the average effect size (d = .27, z = 9.31, p < .001). This effect size is considered small as per Cohen's (1988) definition. The three-level random-effects model shows that the total heterogeneity is lower within studies ($I^2_{(2)}$ = 32.4%) than between studies ($I^2_{(3)}$ = 67.5%). Additional analyses reported in detail in Appendix B (p-curve, trim and fill, sensitivity analyses) suggest minimal publication bias (Rothstein et al. 2006).

4.2. Influence of predictors: univariate vs. multivariate model selection

As shown in Table 1, the 11 existing meta-analyses on healthy eating nudges use univariate meta-analyses (i.e., they separately test the impact of each predictor/outcome). Univariate analyses exclude control variables and increase the possibility that significant differences are due to potential confounds. Multivariate models help to provide estimates with better statistical properties, as well as reduce the risk of bias such that a significant result in univariate analyses may not hold using the multivariate model (Jackson et al. 2011). We performed both univariate and multivariate analyses and confirm that the latter lead to a higher model fit as well as more conservative average effect sizes (Figure 2, Appendix C).

Univariate models. We estimated one univariate meta-regression for each predictor *x*. These univariate analyses provide benchmark values which can be compared to the estimates obtained in the full multivariate model. When the predictor is categorical (for intervention type, for example), the univariate model in Equation 5 estimates S coefficients β_s corresponding to each level of the categorical predictor, without any covariate. The third and fourth column of Figure 2 show, respectively, the mean and standard errors of the β_s coefficients, which capture the effect size for each level of the categorical variables as estimated in a univariate regression.

$$y_{ij} = \sum_{1}^{S} \beta_s x_{ij} + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{5}$$

For study duration, which is a continuous variable measured in weeks, the univariate model estimated one intercept and one parameter, as shown in Equation 6. To provide a point estimate for short and long study durations, Figure 2 shows the model's intercept estimated at the first quartile (1 week) and third quartile (15 weeks) of the distribution of duration.

$$y_{ij} = d_0 + \beta_{Duration} Duration_{ij} + u_{(2)ij} + u_{(3)j} + e_{ij} \tag{6}$$

Univariate analyses suggested that the effectiveness of healthy eating nudges varies by intervention type (R^2 = 29%, $\chi^2(4) = 35$, $p < .001$). Figure 2 shows, for example, that the estimated effect sizes in the univariate analysis of intervention type vary between $d = .12$ for cognitive interventions and $d = .48$ for behavioral interventions, and are all statistically different from zero. Univariate analyses found no difference between selection and consumption (R^2 = 4%, $\chi^2(1) = 2.8$, $p = .09$) but significant effect depending on the behavior measured (total eating, healthy eating, mixed eating, or unhealthy eating: R^2 = 17%, $\chi^2(3) = 26$, $p < .001$). They found no differences between adults and children (R^2 = 2%, $\chi^2(1) = 1.1$, $p = .29$); between grocery stores, offsite eateries, or onsite cafeterias (R^2 = 7%, $\chi^2(2) = 5.5$, $p = .07$), or between studies conducted in the US and outside the US (R^2 = 2%, $\chi^2(1) = 2.9$, $p = .09$). Finally, they found a significant effect of duration (R^2 = 13%, $\chi^2(1) = 10.9$, $p = .02$) and study design (R^2 = 9%, $\chi^2(2) = 7.0$, $p = .03$).

Multivariate model. We estimated a full model with all the predictors entered simultaneously as shown in equation 7, where *s* corresponds to the categories for each predictor *k*. The multivariate model explained 46% of the variance, a significant improvement over the intercept-only model ($\chi^2(15) = 70$, $p < .001$). It is also a significant improvement over the best univariate model, the one with intervention type ($\chi^2(7) = 35$, $p < .001$). This suggests that our overall conceptual framework captured a substantial variation in the effect sizes, much more than any separate univariate model.

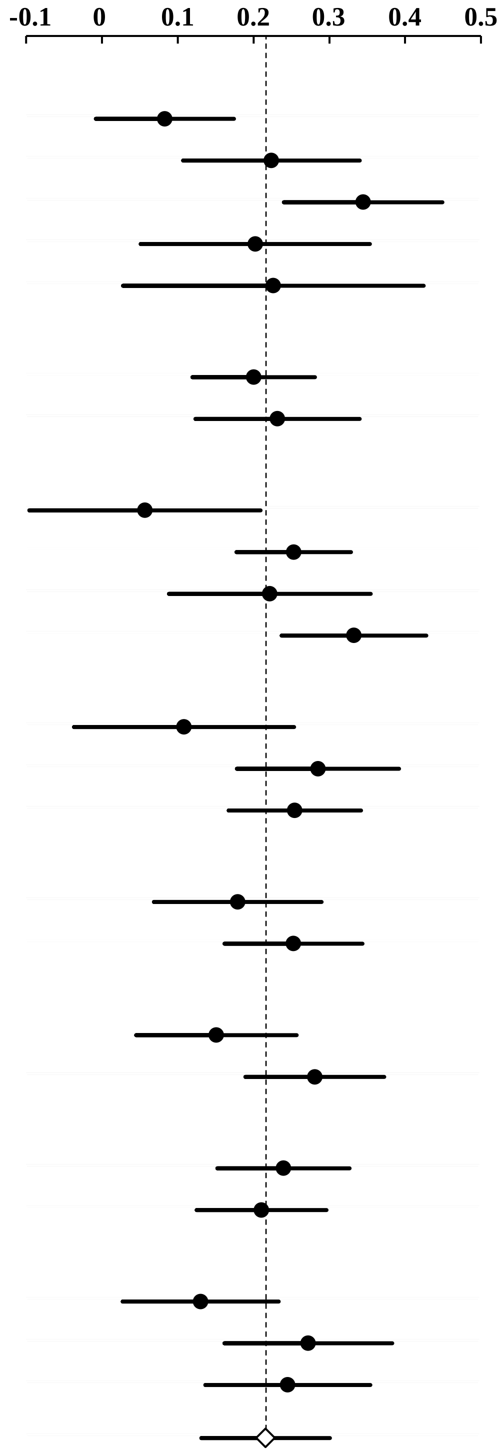
$$y_{ij} = d_0 + \sum_{s}^{S-1} \sum_{k}^{K} \beta_{ks} x_{kij} + u_{(2)ij} + u_{(3)j} + e_{ij} \qquad (7)$$

The fifth and sixth columns in Figure 2 show the multivariate effect sizes estimated for each level of a given predictor, when all the other predictors are at their mean value. Overall, the multivariate model yielded effect sizes that are 16.4% smaller than those of the univariate models.

Figure 2: Effect sizes in the univariate and full multivariate models

		Univariate		Multi-variate	
	k	***d***	***se***	***d***	***se***
Intervention type					
Cognitive	113	.12*	.04	.08	.05
Affective	47	.33*	.06	.22*	.06
Behavioral	73	.48*	.05	.35*	.06
Mixed: cognitive present	37	.20*	.08	.20*	.08
Mixed: cognitive absent	7	.30*	.10	.23*	.10
Eating behavior type					
Selection	231	.25*	.03	.20*	.04
Consumption	46	.34*	.05	.23*	.06
Eating behavior measure					
Total eating	22	.04	.06	.06	.08
Healthy eating	170	.27*	.03	.26*	.04
Mixed eating	14	.24*	.06	.22*	.07
Unhealthy eating	71	.39*	.04	.33*	.05
Population setting					
Grocery stores	38	.12	.08	.11	.08
Offsite eateries	67	.24*	.05	.29*	.05
Onsite cafeterias	172	.32*	.04	.25*	.05
Population age					
Children	96	.31*	.05	.18*	.06
Adults	181	.25*	.04	.25*	.05
Population country					
Other countries	66	.18*	.06	.15*	.05
US	211	.30*	.03	.28*	.05
Study duration					
Short (1 week)	277	.32*	.05	.24*	.04
Long (15 weeks)	277	.24*	.05	.21*	.04
Study design					
Single pre-post	151	.24*	.04	.13*	.05
Single treatment-control	71	.38*	.05	.28*	.06
Double-difference	55	.19*	.06	.25*	.06
Average effect	**277**	**.27***	**.03**	**.22***	**.04**

Forrest plot (mean and 95% confidence interval from the multivariate model)



* $p < .05$.

Table 3: Parameter estimates of the multivariate model

	β	***se***	***Z***
Intercept	.22***	.04	5.00
Intervention type			
Cognitive	(ref)		
Affective	.14*	.06	2.26
Behavioral	.26***	.06	4.44
Mixed: cognitive present	.12	.08	1.42
Mixed: cognitive absent	.14	.10	1.39
Eating behavior type			
Selection	(ref)		
Consumption	.03	.05	.70
Eating behavior measure			
Total eating	-.20**	.08	-2.58
Healthy eating	(ref)		
Mixed eating	-.03	.07	-.54
Unhealthy eating	.08*	.04	2.18
Population setting			
Grocery stores	(ref)		
Offsite eateries	.18*	.08	2.23
Onsite cafeterias	.15*	.07	2.10
Population age			
Children	(ref)		
Adults	.07	.06	1.30
Population country			
Other countries	(ref)		
US	.13*	.05	2.49
Study duration			
Intervention length (week)	-.002	.001	-1.59
Study design			
Single-difference pre-post	(ref)		
Single-difference treatment-control	.14**	.05	2.78
Double-difference	.12	.06	1.95
K (observations)	277		
N (studies)	87		
R^2	.46		
LR test vs. intercept-only model	70***		

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.
Note: Each coefficient is interpreted as the difference with the reference category, denoted as “(ref).”

After controlling for all covariates, the average effect size computed across all 277 observations shrinks slightly, from d =.27 to d = .22, but remains significantly different from zero (z = 5.00, p < .001). Other effect sizes show stronger reductions. Importantly, the effect size for cognitive intervention shrinks from .12 to .08 and is no longer statistically significant. The reduction is particularly strong for the largest effect sizes, like behavioral interventions (which shrinks from .48 to .35). In the next section, we examine whether these smaller differences are still statistically significant.

4.3. Hypothesis testing

To test our hypotheses, we estimated the full multivariate model (equation 7). We used ANOVA coding (e.g., consumption = ½, selection = -½) so that the coefficients of the categorical predictors represent a contrast with the reference category (see Table 3).

Effect sizes vary significantly between the three types of interventions. As hypothesized, cognitive interventions are significantly less effective than affective (β = -.14, z = -2.26, p = .02) or behavioral interventions (β = -.26, z = -4.44, p < .001). As expected, affective interventions are less effective than behavioral interventions (β = -.12, z = -1.97, p = .049). Note that we chose to report two-tailed p-values throughout the paper to avoid confusion and to be conservative, but that a one-tailed test (yielding p = .025) would also be appropriate given that our hypothesis is about the ordering of cognitive, affective, and behavioral interventions. Finally, mixed interventions are not more effective than pure cognitive interventions, whether they include a cognitive intervention or not (respectively, β = .12, z = 1.42, p = .15 and β = .14, z = 1.40, p = .16).

As expected, effect sizes are similar for food selection and actual consumption (β = .03, z = .70, p = .48). However, effect sizes are significantly lower for total eating compared with healthy eating (β = -.20, z = -2.58, p = .01) or unhealthy eating (β = -.28, z = -3.58, p < .001). As hypothesized as well, effect sizes are significantly higher for unhealthy eating than for healthy eating (β = .08, z = 2.18, p = .03).

As expected, and in contrast to what the univariate analyses suggested, effect sizes are significantly lower for grocery stores compared to offsite eateries (β = -.18, z = -2.23, p = .03) or onsite eateries (β = -.15, z = -2.10, p = .04). There are no differences between onsite and offsite eateries (β = -.03, z = -.54, p = .60). As expected, and contrary to the univariate results, effect sizes are significantly higher in the US than in other countries (β = .13, z = 2.50, p = .01). Contrary to our hypothesis, there is no difference between children and adults (β =.07, z = 1.30, p = .19). Similarly, effect sizes are unrelated to study duration (β = -.002, z = -1.60, p = .11), contrary to our hypothesis and to the univariate results. Last, effect sizes are significantly lower in pre-post studies than in double-difference studies (β = -.14, z = -2.78, p < .01), and (marginally) lower than in single-difference studies (β = -.12, z = -1.95, p = .051). There is no difference between studies using a single- and double-difference design (β = -.03, z = -.42, p = .67).

5. Discussion

It is easy to understand the growing enthusiasm for healthy eating nudges in academic and policy circles. They promise to improve people's diet at a fraction of the cost of economic incentives or education programs and without imposing new taxes or constraints on businesses or consumers. But do they really deliver on this promise? The encouraging results of existing reviews and meta-analyses are derived from analyses of only a subset of interventions, and often include results from studies conducted in favorable laboratory or online settings. More important, existing meta-analyses relied on univariate comparisons between two or three groups of studies and failed to control for important differences in eating behaviors, population, and studies in their sample or for the fact that some studies yielded multiple effect sizes.

5.1. Do healthy nudges work, and to what extent?

Our analysis of 277 effect sizes derived from 81 articles and 87 field experiments shows that the average effect size of healthy eating nudges is d = .217, with a 95% confidence interval of [.132; .302]). Although this number is lower than what would have been obtained without controlling for the characteristics of the eating behaviors, population, and studies, it is still considered "small" (Cohen 1988).

In order to provide a more intuitive grasp of what this means, we computed the daily energy equivalent that one would expect from such an effect size using the method described in Hollands et al. (2015). Since d is the standardized mean difference, a d of .217 means that, on average, healthy eating nudges increase healthy eating by .217 standard deviations. Assuming that the standard deviation in daily energy intake is 537 kcal for an adult (Hollands et al. 2015), the average effect size of .217 translates into a .217*537 = 117 kcal change in daily energy intake (-6.8% of the 1,727 kcal average energy intake). Given that a teaspoon of sugar contains 16 kcal, this is equivalent to about 7 fewer teaspoons of sugar per day (see Table 4).

Table 4: Expected daily energy equivalents by intervention type

	Effect sizes	**Daily equivalents** [a]		
	Cohen's d (SMD)	Energy intake change (kcal)	Energy intake change (%)	Teaspoons sugar [b]
Cognitive interventions	.084	-45	-2.6%	-2.8
Affective interventions	.225	-121	-7.0%	-7.5
Behavioral interventions	.346	-186	-10.8%	-11.6
Overall meta-analytical effect	.217	-117	-6.8%	-7.3

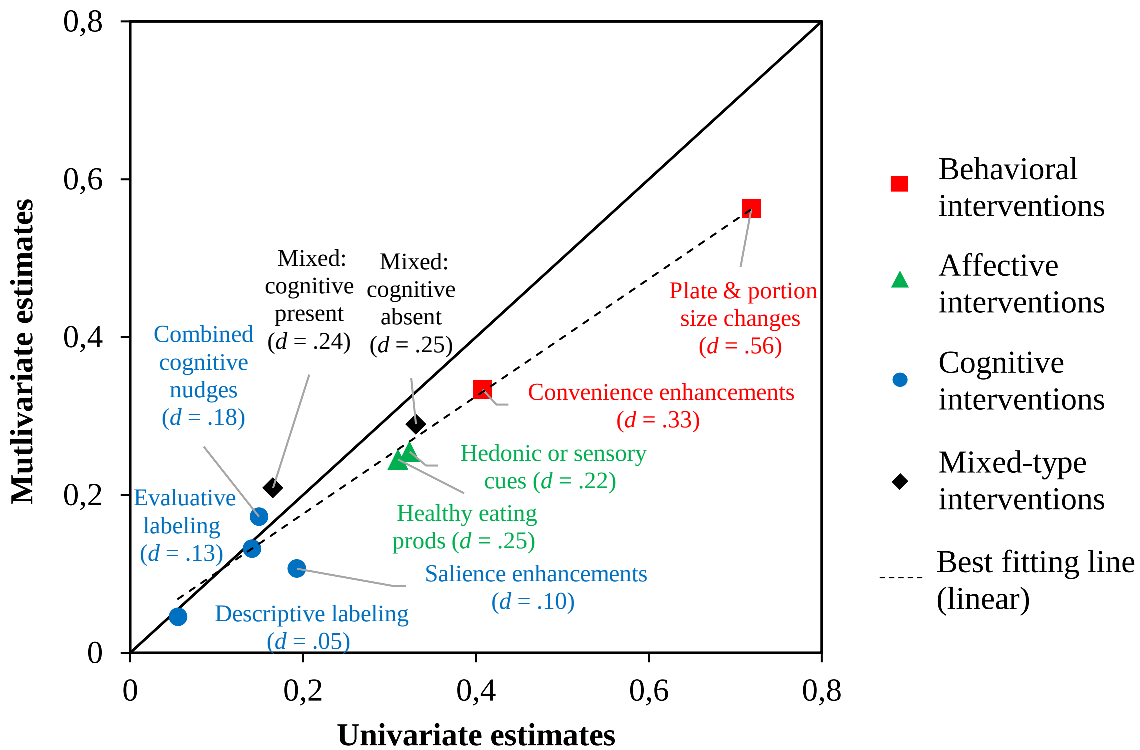
Notes: [a] The daily equivalents are computed using the mean and standard deviation in daily energy intake of 1,727 ± 537 kcal reported in Hollands et al. (2015). [b] One teaspoon of sugar contains 16 kcal.

5.2. Which type of healthy eating nudge works best?

Table 4 provides the daily equivalents, in calories and teaspoons of sugar, of the average effect sizes of cognitive, affective, and behavioral interventions. It shows that effect sizes increase by 168% between cognitive and affective interventions (reducing daily energy intake from 45 kcal to 121 kcal). Even more remarkable, moving from a cognitive to a behavioral intervention is estimated to increases effect sizes by a factor of 4 (reducing daily energy intake from 45 to 186 kcal per day).

There are also important differences between each type of cognitive, affective, and behavioral nudges. As detailed in Appendix C, we estimated another regression which, instead of estimating five effect sizes (for the three pure types and the two mixed types), estimated a separate effect size for each of the seven subcategories, for the two mixed types of intervention, and for a tenth subcategory consisting of studies combining multiple cognitive interventions (e.g., evaluative nutrition labeling and salience enhancements). There were no studies combining the two types of affective interventions or the two types of behavioral interventions. Figure 3 plots the univariate (x-axis) and multivariate (y-axis) estimates of these ten effect sizes.

Figure 3: Effect sizes of intervention types estimated by univariate and multivariate models



Note: Effect sizes in parentheses are those estimated by the multivariate model.

As Figure 3 shows, effect sizes are smaller (by 12% on average) in the multivariate analyses and the magnitude of the reduction increases with the size of the effect. For example, the univariate average effect size for plate and portion size shrinks by 22% (from d = .72 to d = .56). Note that this univariate estimate (d = .72) is very similar to the value (d = .76) reported by Holden et al. (2016) for studies with unaware participants (e.g., excluding laboratory or online studies), suggesting that the reduction found by our multivariate model would apply also to the sample of studies examined in prior meta-analyses. In addition to overestimating effect sizes, the univariate analysis incorrectly ranks some of the interventions, suggesting, for example, that salience enhancements are more effective than evaluative labeling, when they are not.

5.3. Which other factors influence the effectiveness of healthy eating nudges?

By explaining 46% of the variance among effect sizes, our study shows that some of the characteristics of the eating behavior, population, and study significantly impact the effectiveness of healthy eating nudges. First, we find that interventions more easily reduce unhealthy eating than improve healthy eating or decrease total eating. In other words, it is easier to make people eat less chocolate cake than to make them eat more vegetables, and the most difficult is to make them generally eat less. Indeed, the effect size of healthy eating nudges on total eating (d = .06, z = .73, p = .46) is not statistically different from zero. This finding is consistent with what we know about the difficulty—perhaps even pointlessness—of hypocaloric diets.

Our result of a 33% stronger effect size for reducing unhealthy eating than for increasing healthy eating is consistent with prior self-control research (Prelec and Loewenstein 1998; Wertenbroch 1998). Dynamically inconsistent preferences and self-control lapses can explain why people would particularly welcome interventions that reduce unhealthy eating and help them stick to their long-term goals and avoid regret (Schwartz et al. 2014). However, future research is necessary to test the robustness of this finding, which may be explained, at least in part, by the fact that the most effective interventions (plate and portion size changes) have been disproportionally implemented to reduce unhealthy eating rather than to promote healthy eating.

On the other hand, we replicate prior findings of similar effect sizes for food selection rather than actual consumption. This is an important result because it suggests that researchers or practitioners may not need to measure actual consumption to test the impact of their interventions, which is usually considerably more onerous to measure than just the number of consumers picking healthier options.

Importantly, and contrary to what we expected and to what had been reported in the literature, effect sizes are unaffected by the duration of the study. Although it is a reassuring result for anyone concerned about the long-term effectiveness of healthy eating nudges, we should note that the trend is nevertheless toward smaller effects for longer studies. Figure 2 shows that our model predicts that increasing the duration of the study from 1 week to 15 weeks would reduce effect size by 12% (from $d = .24$ to $d = .21$). It remains to be seen by how much effect size would change if a nudge were conducted over a longer period.

We also find that effect sizes increase with the level of control in the design of the study. On average, effect sizes are 98% larger in studies with a control group compared with those with a simple pre-post design without a control group. This suggests that researchers should use stronger controls as much as possible. It also provides a way to correct the effect sizes found in pre-post studies and to forecast what they might have been in a more controlled setting.

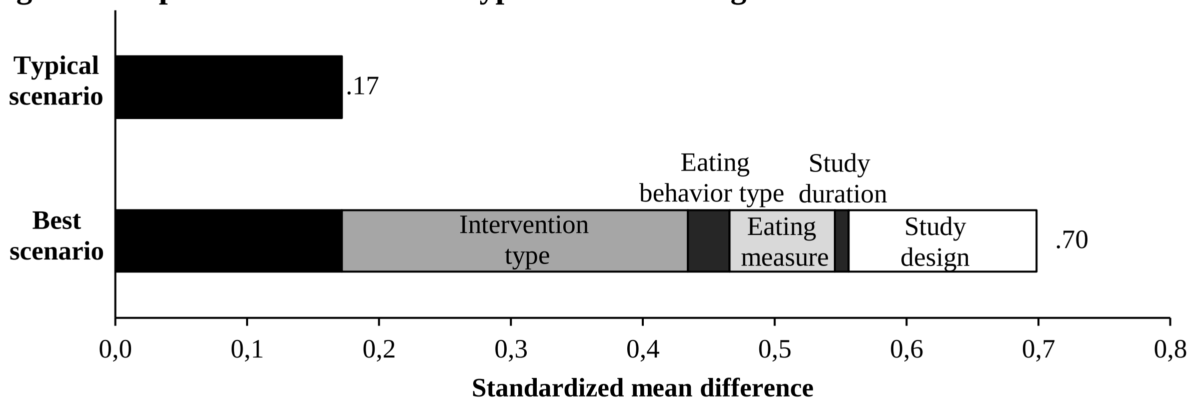
Turning to population characteristics, effect sizes are 60% smaller on average among grocery shoppers than among cafeteria or restaurant eaters. This is consistent with our hypothesis and with the literature, although more research is needed to determine if it is because of the differences between choosing for immediate or future consumption, because of different levels of competition, or because different goals are salient when grocery shopping vs. eating. Also consistent with our hypothesis, effect sizes are 85% larger in studies conducted in the US than in other countries. This may be because Americans focus less on the experience and more on the health effects of eating (Rozin et al. 1999) or because they rely more strongly on external eating cues than internal ones (Wansink et al. 2007). It could also be caused by the higher proportion of overweight people in the US and the larger size of portions (Rozin et al. 2003). On the other hand, the difference in effect sizes between adults and children is not statistically significant. Still, compared to the univariate analyses which suggest larger effects for children than for adults, our results are in the direction (smaller effects for children) that we hypothesized and that is consistent with the literature. To determine conclusively whether children and adults respond differently to healthy eating nudges, more research is needed, especially on cognitive interventions which have, so far, been tested primarily with adults.

Table 5: Expected effectiveness increase between typical and best nudge study

Predictor	Typical scenario	Best scenario	Increase (*d*)	Increase (contribution)
Intervention type	Cognitive	Behavioral	.26	50%
Eating behavior type	Selection	Consumption	.03	6%
Eating behavior measure	Healthy	Unhealthy	.08	15%
Study Duration	6 weeks	1 week	.01	2%
Study Design	Pre-post	Single-difference	.14	27%
Population: Country	US	US		
Population: Location	Onsite cafeterias	Onsite cafeterias		
Population: Age	Adults	Adults		
Effect Size (*d*)	**.17**	**.70**	**.53**	**100%**

Our analysis allows us to predict what effect size one could expect when conducting a field experiment with any combination of predictors, including the most typical and the most effective combination. Table 5 summarizes the typical and best scenarios, while Figure 4 shows the contribution of the different predictors. Table 5 shows that researchers choosing the most typical level of each predictor (studying the effects of a cognitive intervention on the healthy food selection of US adult cafeteria eaters for a pre-post 6-week study) could expect an effect size of only d = .17, 95% CI [.07, .28]). In contrast, researchers choosing the best combination of predictors (studying the effects of a behavioral intervention on the unhealthy food consumption of adult cafeteria eaters for a single-difference 1-week study) could expect an effect size four and a half times larger (d = .70, 95% CI [.56, .84]). Computing the daily energy equivalents, we get a reduction by 94 kcal for the typical nudge study, and a reduction by 374 kcal for the best one.

Figure 4: Expected effectiveness in typical vs. best nudge scenarios



5.4. Directions for future research

Our findings offer insights into where more research is needed and where it is not. Table 2 shows the number of observations by intervention type and target eating behavior (healthy or unhealthy). This table makes it immediately apparent that no field experiment has tested the effectiveness of displaying unattractive product descriptions or photos of unhealthy foods, like the dissuasive photos used on cigarette packs in some countries (Kees et al. 2006). Although degrading other brands may be difficult because of trademark laws, it has shown promise in laboratory studies (Hollands et al. 2011), and retailers or restaurants could test this strategy with their own unhealthy products. It would also seem important to run more studies increasing portion, plate, or glass size for healthy foods and beverages, rather than for unhealthy ones. Further studies are needed to examine the effectiveness of hedonic sensory cues, for which we only have 12 effect sizes. Precedence should be given to testing interventions in grocery stores and outside the US, and for unhealthy foods. These issues should have priority over other well-researched topics, such as studying the effects of cognitive interventions on healthy eating in cafeterias using a pre-post design.

Beyond filling out the underpopulated cells of the framework, three research areas appear particularly fruitful. The first is to study interaction effects. Because of lack of data in our sample, it is not possible to estimate interactions effects between each intervention type and the other predictors. In Appendix D, we report the results of a simplified model using linear coding for intervention type and eating behavior (from total eating to healthy eating). This preliminary analysis suggests that shifting from cognitive to behavioral interventions is particularly impactful on consumption (vs. selection), for adults (vs. children), and in the US (vs. in other countries). These results qualify the lack of main effect for eating behavior and for participant age reported in the main results. However, additional field experiments orthogonally manipulating intervention type and population or study characteristics would be necessary to confirm these results.

Second, it would be important to study the interplay between interventions and economic incentives. One would hope to find synergies between the two that allow, for example, reduction in the magnitude of economic incentives. Similarly, it would be interesting to compare behavioral and economic interventions, both in terms of their effects on healthy eating and in terms of their cost effectiveness.

Finally, the prevalence and severity of non-communicable diseases is strongly associated with socioeconomic and cultural factors such as income, education, gender, ethnicity, and culture (Bartley 2017). Surprisingly, this data was almost never available in the studies that we analyzed. Future research should therefore measure socioeconomic data, as well as biomarkers such as body mass or diabetes diagnoses, and traits such as cognitive restraint or impulsivity that strongly influence food choices and health (Ma et al. 2013; Sutin et al. 2011). Such information should be provided, at a minimum, to better characterize the respondent population; it would be even better to report results separately for different population types. This should be done systematically even in the absence of significant differences. When prior research or theory predicts an effect, finding none can be informative.

More broadly, future research should expand the dependent variables beyond purchase and consumption. To encourage the adoption of healthy eating nudges in commercial operations, it is important to measure their impact on the consumer's experience, satisfaction, and perception of value, as well as on the company's top and bottom lines. Even interventions that lead to a reduction in consumption can be good for business if they attract new consumers who value their ability to nudge them away from unhealthy choices that they will later regret. Similarly, one of the core tenets of nudges is that they improve consumer welfare as judged by consumers themselves (Sunstein 2017). Future studies should therefore measure whether making the intervention more salient, by alerting consumers to it, for example, would influence its effectiveness. It would be important to know whether people, upon learning that they have been subject to an intervention, would agree that it led them to make better decisions compared to the status quo ante, but also compared to other interventions such as taxes and other economic incentives. This is important because, while they preserve freedom of choice, the interventions analyzed here are nevertheless paternalistic. Finding that consumers welcome these interventions, and that they are compatible with commercial goals, would go a long way to encourage their adoption in multiple contexts.

5.5. Toward a "living" meta-analysis

One of the biggest challenges of studying healthy eating nudges is the exponential increase in the studies carried out. Of the 277 effect sizes we analyzed, 166 (60%) were published in the past 5 years, a 133% increase over the previous 5 years. By simple extrapolation, we can predict that 222 new effect sizes will be estimated between 2017 and 2021. This upsurge makes meta-analyses even more valuable but also means that they rapidly become out of date. Compounding the problem, research on healthy eating nudges is published in a wide variety of scientific publications in marketing, nutrition, psychology, and health sciences, which are indexed in different databases and not always available to all researchers. To mitigate these problems, encourage the diffusion of our results, and correct possible categorization errors, the spreadsheet containing the raw data will be made available online (post publication). In addition, we have created a simple survey (available at http://tinyurl.com/healthy-eating-nudge) to allow researchers to correct and update the database by entering information about their study. We hope that this "living" meta-analysis will encourage the consolidation and diffusion of knowledge and contribute to making science more open.

6. References[*]

* Adams, M.A., R.L. Pelletier, M.M. Zive, J.F. Sallis. 2005. Salad bars and fruit and vegetable consumption in elementary schools: A plate waste study. *J Am Diet Assoc*. **105**(11) 1789-1792.

Afshin, A., J.L. Peñalvo, L. Del Gobbo, J. Silva, M. Michaelson, M. O'Flaherty, S. Capewell, D. Spiegelman, G. Danaei, D. Mozaffarian. 2017. The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLoS One*. **12**(3) e0172277.

Arno, A., S. Thomas. 2016. The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BMC Public Health*. **16**(1) 676.

* Auchincloss, A.H., G.G. Mallya, B.L. Leonberg, A. Ricchezza, K. Glanz, D.F. Schwarz. 2013. Customer responses to mandatory menu labeling at full-service restaurants. *Am J Prev Med*. **45**(6) 710-719.

Barry, T.E., D.J. Howard. 1990. A review and critique of the hierarchy of effects in advertising. *Int J Advert*. **9**(2) 121-135.

* Bartholomew, J.B., E.M. Jowers. 2006. Increasing frequency of lower-fat entrees offered at school lunch: An environmental change strategy to increase healthful selections. *J Acad Nutr Diet*. **106**(2) 248-252.

Bartley, M. 2017. *Health inequality : An introduction to concepts, theories and methods*. Polity, Cambridge, UK ; Malden, MA, USA.

* Baskin, E., M. Gorlin, Z. Chance, N. Novemsky, R. Dhar, K. Huskey, M. Hatzis. 2016. Proximity of snacks to beverages increases food consumption in the workplace: A field study. *Appetite*. **103** 244-248.

Bloom, D., E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. Bloom, S. Fathima, A. Feigl, T. Gaziano, A. Hamandi, M. Mowafi. 2012. *The global economic burden of noncommunicable diseases*. Program on the Global Demography of Aging.

* Bollinger, B., P. Leslie, A. Sorensen. 2011. Calorie posting in chain restaurants. *Am Econ J Econ Policy*. **3**(1) 91-128.

Booth, A. 2006. Clear and present questions: Formulating questions for evidence based practice. *Library Hi Tech*. **24**(3) 355-368.

Borenstein, M., L.V. Hedges, J.P.T. Higgins, H.R. Rothstein. 2009. *Introduction to meta-analysis*. John Wiley & Sons, Chichester, U.K.

Brandon, A., P.J. Ferraro, J.A. List, R.D. Metcalfe, M.K. Price, F. Rundhammer. 2017. Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments. *NBER Working Paper Series*. **No. 23277**.

Breckler, S.J. 1984. Empirical validation of affect, behavior, and cognition as distinct components of attitude. *J Pers Soc Psychol*. **47**(6) 1191-1205.

* Brissette, I., A. Lowenfels, C. Noble, D. Spicer. 2013. Predictors of total calories purchased at fast-food restaurants: Restaurant characteristics, calorie awareness, and use of calorie information. *J Nutr Educ Behav*. **45**(5) 404-411.

Broers, V.J., C. De Breucker, S. Van den Broucke, O. Luminet. 2017. A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice. *Eur J Public Health*. **27**(5) 912-920.

[*] Article included in the meta-analysis.

Bucher, T., C. Collins, M.E. Rollo, T.A. McCaffrey, N. De Vlieger, D. Van der Bend, H. Truby, F.J. Perez-Cueto. 2016. Nudging consumers towards healthier choices: A systematic review of positional influences on food choice. *Br J Nutr*. **115**(12) 2252-2263.

* Buscher, L.A., K.A. Martin, S. Crocker. 2001. Point-of-purchase messages framed in terms of cost, convenience, taste, and energy improve healthful snack selection in a college foodservice setting. *J Acad Nutr Diet*. **101**(8) 909-913.

Carpenter, C.J. 2010. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Communication*. **25**(8) 661-669.

* Cawley, J., M.J. Sweeney, J. Sobal, D.R. Just, H.M. Kaiser, W.D. Schulze, E. Wethington, B. Wansink. 2015. The impact of a supermarket nutrition rating system on purchases of nutritious and less nutritious foods. *Public Health Nutr*. **18**(01) 8-14.

Cecchini, M., L. Warin. 2016. Impact of food labelling systems on food choices and eating behaviours: A systematic review and meta-analysis of randomized studies. *Obesity Reviews*. **17**(3) 201-210.

Chance, Z., M. Gorlin, R. Dhar. 2014. Why choosing healthy foods is hard, and how to help: Presenting the 4ps framework for behavior change. *Customer Needs and Solutions*. **1**(4) 253-262.

Chernev, A. 2011. The dieter's paradox. *J Consum Psychol*. **21**(2) 178-183.

Cheung, M.W.L. 2014. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods*. **19**(2) 211-229.

* Chu, Y.H., E.A. Frongillo, S.J. Jones, G.L. Kaye. 2009. Improving patrons' meal selections through the use of point-of-selection nutrition labels. *Am J Public Health*. **99**(11) 2001-2005.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, NJ.

* Cohen, J.F., S.A. Richardson, S.A. Cluggish, E. Parker, P.J. Catalano, E.B. Rimm. 2015. Effects of choice architecture and chef-enhanced meals on the selection and consumption of healthier school foods: A randomized clinical trial. *JAMA Pediatr*. **169**(5) 431-437.

* Crockett, R.A., S.A. Jebb, M. Hankins, T.M. Marteau. 2014. The impact of nutritional labels and socioeconomic status on energy intake. An experimental field study. *Appetite*. **81** 12-19.

Croll, J.K., D. Neumark-Sztainer, M. Story. 2001. Healthy eating: What does it mean to adolescents? *J Nutr Educ*. **33**(4) 193-198.

* Dayan, E., M. Bar-Hillel. 2011. Nudge to nobesity ii: Menu positions influence food orders. *Judgm Decis Mak*. **6**(4) 333.

* de Wijk, R.A., A.J. Maaskant, I.A. Polet, N.T. Holthuysen, E. van Kleef, M.H. Vingerhoeds. 2016. An in-store experiment on the effect of accessibility on sales of wholegrain and white bread in supermarkets. *PLoS One*. **11**(3) e0151915.

* Diliberti, N., P.L. Bordi, M.T. Conklin, L.S. Roe, B.J. Rolls. 2004. Increased portion size leads to increased energy intake in a restaurant meal. *Obesity*. **12**(3) 562-568.

* DiSantis, K.I., L.L. Birch, A. Davey, E.L. Serrano, J. Zhang, Y. Bruton, J.O. Fisher. 2013. Plate size and children’s appetite: Effects of larger dishware on self-served portions and intake. *Pediatrics*. **131**(5) e1451-e1458.

* Downs, J.S., J. Wisdom, B. Wansink, G. Loewenstein. 2013. Supplementing menu labeling with calorie recommendations to test for facilitation effects. *Am J Public Health*. **103**(9) 1604-1609.

* Dubbert, P.M., W.G. Johnson, D.G. Schlundt, N.W. Montague. 1984. The influence of caloric information on cafeteria food choices. *J Appl Behav Anal*. **17**(1) 85-92.

Duckworth, A.L., T.S. Gendler, J.J. Gross. 2016. Situational strategies for self-control. *Perspectives on Psychological Science*. **11**(1) 35-55.

* Dumanovsky, T., C.Y. Huang, C.A. Nonas, T.D. Matte, M.T. Bassett, L.D. Silver. 2011. Changes in energy content of lunchtime purchases from fast food restaurants after introduction of calorie labelling: Cross sectional customer surveys. *BMJ*. **343**.

Duval, S., R. Tweedie. 2000. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. **56**(2) 455-463.

* Elbel, B., J. Gyamfi, R. Kersh. 2011. Child and adolescent fast-food choice and the influence of calorie labeling: A natural experiment. *Int J Obes*. **35**(4) 493-500.

* Elbel, B., R. Kersh, V.L. Brescoll, L.B. Dixon. 2009. Calorie labeling and food choices: A first look at the effects on low-income people in new york city. *Health Aff*. **28**(6) w1110-1121.

* Elbel, B., T. Mijanovich, L.B. Dixon, C. Abrams, B. Weitzman, R. Kersh, A.H. Auchincloss, G. Ogedegbe. 2013. Calorie labeling, fast food purchasing and restaurant visits. *Obesity*. **21**(11) 2172-2179.

* Ellison, B., J.L. Lusk, D. Davis. 2013. Looking at the label and beyond: The effects of calorie labels, health consciousness, and demographics on caloric intake in restaurants. *Int J Behav Nutr Phys Act*. **10**(1) 21.

* Elsbernd, S., M. Reicks, T. Mann, J. Redden, E. Mykerezi, Z. Vickers. 2016. Serving vegetables first: A strategy to increase vegetable consumption in elementary school cafeterias. *Appetite*. **96** 111-115.

* Ensaff, H., M. Homer, P. Sahota, D. Braybrook, S. Coan, H. McLeod. 2015. Food choice architecture: An intervention in a secondary school and its impact on students' plant-based food choices. *Nutrients*. **7**(6) 4426-4437.

Fernandes, A.C., R.C. Oliveira, R.P.C. Proença, C.C. Curioni, V.M. Rodrigues, G.M.R. Fiates. 2016. Influence of menu labeling on food choices in real-life settings: A systematic review. *Nutr Rev*. **74**(8) 534-548.

* Finkelstein, E.A., K.L. Strombotne, N.L. Chan, J. Krieger. 2011. Mandatory menu labeling in one fast-food chain in king county, washington. *Am J Prev Med*. **40**(2) 122-127.

* Foster, G.D., A. Karpyn, A.C. Wojtanowski, E. Davis, S. Weiss, C. Brensinger, A. Tierney, W. Guo, J. Brown, C. Spross. 2014. Placement and promotion strategies to increase sales of healthier products in supermarkets in low-income, ethnically diverse neighborhoods: A randomized controlled trial. *Am J Clin Nutr*. **99**(6) 1359-1368.

* Freedman, M.R. 2011. Point-of-selection nutrition information influences choice of portion size in an all-you-can-eat university dining hall. *Journal of Foodservice Business Research*. **14**(1) 86-98.

* Freedman, M.R., C. Brochado. 2010. Reducing portion size reduces food intake and plate waste. *Obesity*. **18**(9) 1864-1866.

* Gaigi, H., S. Raffin, M. Maillot, L. Adrover, B. Ruffieux, N. Darmon. 2015. Expérimentation d'un fléchage nutritionnel dans deux supermarchés à marseille «le choix vita+». *Cahiers de Nutrition et de Diététique*. **50**(1) 16-24.

* Gamburzew, A., N. Darcel, R. Gazan, C. Dubois, M. Maillot, D. Tomé, S. Raffin, N. Darmon. 2016. In-store marketing of inexpensive foods with good nutritional quality in disadvantaged neighborhoods: Increased awareness, understanding, and purchasing. *Int J Behav Nutr Phys Act*. **13**(1) 104.

* Geaney, F., C. Kelly, J.S. Di Marrazzo, J.M. Harrington, A.P. Fitzgerald, B.A. Greiner, I.J. Perry. 2016. The effect of complex workplace dietary interventions on employees' dietary intakes, nutrition knowledge and health status: A cluster controlled trial. *Prev Med*. **89** 76-83.

Glanz, K., M. Basil, E. Maibach, J. Goldberg, D.A.N. Snyder. 1998. Why americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J Am Diet Assoc*. **98**(10) 1118-1126.
* Goto, K., A. Waite, C. Wolff, K. Chan, M. Giovanni. 2013. Do environmental interventions impact elementary school students' lunchtime milk selection? *Appl Econ Perspect Policy*. **35**(2) 360-376.
* Hanks, A.S., D.R. Just, L.E. Smith, B. Wansink. 2012. Healthy convenience: Nudging students toward healthier choices in the lunchroom. *J Public Health*. **34**(3) 370-376.
* Hanks, A.S., D.R. Just, B. Wansink. 2013. Smarter lunchrooms can address new school lunchroom guidelines and childhood obesity. *J Pediatr*. **162**(4) 867-869.
Hanssens, D.M., K.H. Pauwels, S. Srinivasan, M. Vanhuele, G. Yildirim. 2014. Consumer attitude metrics for guiding marketing mix decisions. *Marketing Sci*. **33**(4) 534-550.
Herman, C.P., J. Polivy. 2008. External cues in the control of food intake in humans: The sensory-normative distinction. *Physiol Behav*. **94**(5) 722-728.
Higgins, J., S.G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Stat Med*. **21**(11) 1539-1558.
Hilgard, E.R. 1980. The trilogy of mind: Cognition, affection, and conation. *J Hist Behav Sci*. **16**(2) 107-117.
* Hoefkens, C., C. Lachat, P. Kolsteren, J. Van Camp, W. Verbeke. 2011. Posting point-of-purchase nutrition information in university canteens does not influence meal choice and nutrient intake. *Am J Clin Nutr*. **94**(2) 562-570.
Holden, S.S., N. Zlatevska, C. Dubelaar. 2016. Whether smaller plates reduce consumption depends on who’s serving and who’s looking: A meta-analysis. *Journal of the Association for Consumer Research*. **1**(1) 134-146.
Hollands, G., G. Bignardi, M. Johnston, M. Kelly, D. Ogilvie, M. Petticrew, A. Prestwich, I. Shemilt, S. Sutton, T. Marteau. 2017. The tippme intervention typology for changing environments to change behaviour. *Nature Human Behaviour*.
Hollands, G.J., A. Prestwich, T.M. Marteau. 2011. Using aversive images to enhance healthy food choices and implicit attitudes: An experimental test of evaluative conditioning. *Health Psychol*. **30**(2) 195.
Hollands, G.J., I. Shemilt, T.M. Marteau, S.A. Jebb, M.P. Kelly, R. Nakamura, M. Suhrcke, D. Ogilvie. 2013. Altering micro-environments to change population health behaviour: Towards an evidence base for choice architecture interventions. *BMC Public Health*. **13**(1) 1218.
Hollands, G.J., I. Shemilt, T.M. Marteau, S.A. Jebb, H.B. Lewis, Y. Wei, J.P. Higgins, D. Ogilvie. 2015. Portion, package or tableware size for changing selection and consumption of food, alcohol and tobacco. *Cochrane Database Syst Rev*. **9** CD011045.
* Hubbard, K.L., L.G. Bandini, S.C. Folta, B. Wansink, M. Eliasziw, A. Must. 2015. Impact of a smarter lunchroom intervention on food selection and consumption among adolescents and young adults with intellectual and developmental disabilities in a residential school setting. *Public Health Nutr*. **18**(2) 361-371.
Jackson, D., R. Riley, I.R. White. 2011. Multivariate meta-analysis: Potential and promise. *Stat Med*. **30**(20) 2481-2498.
Januszewska, R., Z. Pieniak, W. Verbeke. 2011. Food choice questionnaire revisited in four countries. Does it still measure the same? *Appetite*. **57**(1) 94-98.
Kees, J., S. Burton, J.C. Andrews, J. Kozup. 2006. Tests of graphic visuals and cigarette package warning combinations: Implications for the framework convention on tobacco control. *J Public Policy Mark*. **25**(2) 212-223.

Kiesel, K., S.B. Villas-Boas. 2013. Can information costs affect consumer choice? Nutritional labels in a supermarket experiment. *Int J Ind Organ*. **31**(2) 153-163.
Kraak, V.I., T. Englund, S. Misyak, E.L. Serrano. 2017. A novel marketing mix and choice architecture framework to nudge restaurant customers toward healthy food environments to reduce obesity in the united states. *Obesity Reviews*. **18**(8) 852-868.
* Krieger, J.W., N.L. Chan, B.E. Saelens, M.L. Ta, D. Solet, D.W. Fleming. 2013. Menu labeling regulations and calories purchased at chain restaurants. *Am J Prev Med*. **44**(6) 595-604.
* Kroese, F.M., D.R. Marchiori, D.T. de Ridder. 2016. Nudging healthy food choices: A field experiment at the train station. *J Public Health*. **38**(2) e133-e137.
* Lachat, C.K., R. Verstraeten, B. De Meulenaer, J. Menten, L.F. Huybregts, J. Van Camp, D. Roberfroid, P.W. Kolsteren. 2009. Availability of free fruits and vegetables at canteen lunch improves lunch and daily nutritional profiles: A randomised controlled trial. *Br J Nutr*. **102**(07) 1030-1037.
* Levin, S. 1996. Pilot study of a cafeteria program relying primarily on symbols to promote healthy choices. *J Nutr Educ*. **28**(5) 282-285.
* Levy, D.E., J. Riis, L.M. Sonnenberg, S.J. Barraclough, A.N. Thorndike. 2012. Food choices of minority and low-income employees: A cafeteria intervention. *Am J Prev Med*. **43**(3) 240-248.
Littlewood, J.A., S. Lourenço, C.L. Iversen, G.L. Hansen. 2016. Menu labelling is effective in reducing energy ordered and consumed: A systematic review and meta-analysis of recent studies. *Public Health Nutr*. **19**(12) 2106-2121.
Long, M.W., D.K. Tobias, A.L. Cradock, H. Batchelder, S.L. Gortmaker. 2015. Systematic review and meta-analysis of the impact of restaurant menu calorie labeling. *Am J Public Health*. **105**(5) e11-e24.
Ma, Y., K.L. Ailawadi, D. Grewal. 2013. Soda versus cereal and sugar versus fat: Drivers of healthful food intake and the impact of diabetes diagnosis. *J Marketing*. **77**(3) 101-120.
Macht, M. 2008. How emotions affect eating: A five-way model. *Appetite*. **50**(1) 1-11.
* Meyers, A.W., A.J. Stunkard. 1980. Food accessibility and food choice: A test of schachter's externality hypothesis. *Arch Gen Psychiatry*. **37**(10) 1133-1135.
* Miller, G.F., S. Gupta, J.D. Kropp, K.A. Grogan, A. Mathews. 2016. The effects of pre-ordering and behavioral nudges on national school lunch program participants' food item selection. *J Econ Psychol*. **55** 4-16.
* Mishra, A., H. Mishra, T.M. Masters. 2012. The influence of bite size on quantity of food consumed: A field study. *Journal of Consumer Research*. **38**(5) 791-795.
Moher, D., A. Liberati, J. Tetzlaff, D.G. Altman, P. Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLoS Med*. **6**(7) e1000097.
* Mollen, S., R.N. Rimal, R.A. Ruiter, G. Kok. 2013. Healthy and unhealthy social norms and food selection. Findings from a field-experiment. *Appetite*. **65** 83-89.
* Morizet, D., L. Depezay, P. Combris, D. Picard, A. Giboreau. 2012. Effect of labeling on new vegetable dish acceptance in preadolescent children. *Appetite*. **59**(2) 399-402.
Murimi, M.W., M. Kanyi, T. Mupfudze, M.R. Amin, T. Mbogori, K. Aldubayan. 2017. Factors influencing efficacy of nutrition education interventions: A systematic review. *J Nutr Educ Behav*. **49**(2) 142-165.e141.
Nikolaou, C.K., C.R. Hankey, M.E.J. Lean. 2014. Calorie-labelling: Does it impact on calorie purchase in catering outlets and the views of young adults? *Int J Obes*. **39** 542.
* Nikolova, H.D., J.I. Inman. 2015. Healthy choice: The effect of simplified pos nutritional information on consumer food choice behavior. *J Marketing Res*. **7**(December) 817–835.

Nornberg, T.R., L. Houlby, L.R. Skov, F.J. Perez-Cueto. 2016. Choice architecture interventions for increased vegetable intake and behaviour change in a school setting: A systematic review. *Perspectives in public health*. **136**(3) 132-142.

* Ogawa, Y., N. Tanabe, A. Honda, T. Azuma, N. Seki, T. Suzuki, H. Suzuki. 2011. Point-of-purchase health information encourages customers to purchase vegetables: Objective analysis by using a point-of-sales system. *Environ Health Prev Med*. **16**(4) 239-246.

Full publication date: 1999 Oliver, R.L. 1999. Whence consumer loyalty? *J Marketing*. **63** 33-44.

* Olstad, D.L., L.A. Goonewardene, L.J. McCargar, K.D. Raine. 2014. Choosing healthier foods in recreational sports settings: A mixed methods investigation of the impact of nudging and an economic incentive. *Int J Behav Nutr Phys Act*. **11**(1) 6.

* Olstad, D.L., J. Vermeer, L.J. McCargar, R.J. Prowse, K.D. Raine. 2015. Using traffic light labels to improve food selection in recreation and sport facility eating environments. *Appetite*. **91** 329-335.

Ouellette, J.A., W. Wood. 1998. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychol Bull*. **124**(1) 54-74.

* Perry, C.L., D.B. Bishop, G.L. Taylor, M. Davis, M. Story, C. Gray, S.C. Bishop, R.A.W. Mays, L.A. Lytle, L. Harnack. 2004. A randomized school trial of environmental strategies to encourage fruit and vegetable consumption among children. *Health Educ Behav*. **31**(1) 65-76.

* Policastro, P., Z. Smith, G. Chapman. 2017. Put the healthy item first: Order of ingredient listing influences consumer selection. *J Health Psychol*. **22**(7) 853-863.

Prelec, D., G. Loewenstein. 1998. The red and the black: Mental accounting of savings and debt. *Marketing Sci*. **17**(1) 4-28.

Pulos, E., K. Leng. 2010. Evaluation of a voluntary menu-labeling program in full-service restaurants. *Am J Public Health*. **100**(6) 1035-1039.

* Redden, J.P., T. Mann, Z. Vickers, E. Mykerezi, M. Reicks, S. Elsbernd. 2015. Serving first in isolation increases vegetable intake among elementary schoolchildren. *PLoS One*. **10**(4) e0121283.

* Reicks, M., J.P. Redden, T. Mann, E. Mykerezi, Z. Vickers. 2012. Photographs in lunch tray compartments and vegetable consumption among children in elementary school cafeterias. *JAMA*. **307**(8) 784-785.

* Roberto, C.A., P.D. Larsen, H. Agnew, J. Baik, K.D. Brownell. 2010. Evaluating the impact of menu labeling on food choices and intake. *Am J Public Health*. **100**(2) 312-318.

Robinson, E. 2017. The science behind smarter lunchrooms. *PeerJ Preprints*. **5** e3137v3131.

Robinson, E., S. Nolan, C. Tudur-Smith, E.J. Boyland, J.A. Harrold, C.A. Hardman, J.C. Halford. 2014. Will smaller plates lead to smaller waists? A systematic review and meta-analysis of the effect that experimental manipulation of dishware size has on energy consumption. *Obes Rev*. **15**(10) 812-821.

Rothstein, H.R., A.J. Sutton, M. Borenstein. 2006. *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Roy, R., B. Kelly, A. Rangan, M. Allman-Farinelli. 2015. Food environment interventions to improve the dietary behavior of young adults in tertiary education settings: A systematic literature review. *J Acad Nutr Diet*. **115**(10) 1647-1681.e1641.

Rozin, P., C. Fischler, S. Imada, A. Sarubin, A. Wrzesniewski. 1999. Attitudes to food and the role of food in life in the u.S.A., japan, flemish belgium and france: Possible implications for the diet-health debate. *Appetite*. **33**(2) 163-180.

Rozin, P., K. Kabnick, E. Pete, C. Fischler, C. Shields. 2003. The ecology of eating: Smaller portion sizes in france than in the united states help explain the french paradox. *Psychol Sci*. **14**(5) 450-454.

* Rozin, P., S. Scott, M. Dingley, J.K. Urbanek, H. Jiang, M. Kaltenbach. 2011. Nudge to nobesity i: Minor changes in accessibility decrease food intake. *Judgm Decis Mak*. **6**(4) 323-332.

Schwartz, J., D. Mochon, L. Wyper, J. Maroba, D. Patel, D. Ariely. 2014. Healthier by precommitment. *Psychol Sci*.

* Schwartz, J., J. Riis, B. Elbel, D. Ariely. 2012. Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. *Health Aff*. **31**(2) 399-407.

* Schwartz, M.B. 2007. The influence of a verbal prompt on school lunch fruit consumption: A pilot study. *Int J Behav Nutr Phys Act*. **4**(1) 6.

Seymour, J.D., A. Lazarus Yaroch, M. Serdula, H.M. Blanck, L.K. Khan. 2004. Impact of nutrition environmental interventions on point-of-purchase behavior in adults: A review. *Prev Med*. **39**(2) 108-136.

* Shah, A.M., J.R. Bettman, P.A. Ubel, P.A. Keller, J.A. Edell. 2014. Surcharges plus unhealthy labels reduce demand for unhealthy menu items. *J Marketing Res*. **51**(6) 773-789.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014a. P-curve and effect size: Correcting for publication bias using only significant results. *Perspect Psychol Sci*. **9**(6) 666-681.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014b. P-curve: A key to the file-drawer. *J Exp Psychol Gen*. **143**(2) 534.

Sinclair, S.E., M. Cooper, E.D. Mansfield. 2014. The influence of menu labeling on calories selected or consumed: A systematic review and meta-analysis. *J Acad Nutr Diet*. **114**(9) 1375-1388.e1315.

Skov, L.R., S. Lourenco, G.L. Hansen, B.E. Mikkelsen, C. Schofield. 2013. Choice architecture as a means to change eating behaviour in self-service settings: A systematic review. *Obes Rev*. **14**(3) 187-196.

Sniehotta, F.F., J. Presseau, V. Araújo-Soares. 2014. Time to retire the theory of planned behaviour. *Health Psychol Rev*. **8**(1) 1-7.

Snook, K.R., A.R. Hansen, C.H. Duke, K.C. Finch, A.A. Hackney, J. Zhang. 2017. Change in percentages of adults with overweight or obesity trying to lose weight, 1988-2014. *JAMA*. **317**(9) 971-973.

Srinivasan, S., M. Vanhuele, K. Pauwels. 2010. Mind-set metrics in market response models: An integrative approach. *J Marketing Res*. **47**(4) 672-684.

* Steenhuis, I., P. van Assema, G. van Breukelen, K. Glanz, G. Kok, H. de Vries. 2004. The impact of educational and environmental interventions in dutch worksite cafeterias. *Health Promot Int*. **19**(3) 335-343.

Sunstein, C.R. 2017. 'Better off, as judged by themselves': A comment on evaluating nudges. *International Review of Economics*. **In press**.

Sutin, A.R., L. Ferrucci, A.B. Zonderman, A. Terracciano. 2011. Personality and obesity across the adult life span. *J Pers Soc Psychol*. **101**(3) 579.

* Tal, A., B. Wansink. 2015. An apple a day brings more apples your way: Healthy samples prime healthier choices. *Psychol Mark*. **32**(5) 575-584.

* Tandon, P.S., C. Zhou, N.L. Chan, P. Lozano, S.C. Couch, K. Glanz, J. Krieger, B.E. Saelens. 2011. The impact of menu labeling on fast-food purchases for children and parents. *Am J Prev Med*. **41**(4) 434-438.

Thaler, R.H., C.R. Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books, New York.

Thapaa, J.R., C.P. Lyford. 2014. Behavioral economics in the school lunchroom: Can it affect food supplier decisions? A systematic review. *Int Food Agribus Man*. **17**(A) 187-208.
* Thorndike, A.N., J. Riis, L.M. Sonnenberg, D.E. Levy. 2014. Traffic-light labels and choice architecture: Promoting healthy food choices. *Am J Prev Med*. **46**(2) 143-149.
* Thorndike, A.N., L. Sonnenberg, J. Riis, S. Barraclough, D.E. Levy. 2012. A 2-phase labeling and choice architecture intervention to improve healthy food and beverage choices. *Am J Public Health*. **102**(3) 527-533.
* van Ittersum, K., B. Wansink. 2013. Extraverted children are more biased by bowl sizes than introverts. *PLoS One*. **8**(10) e78224.
van Kleef, E., O. van den Broek, H.C.M. van Trijp. 2015. Exploiting the spur of the moment to enhance healthy consumption: Verbal prompting to increase fruit choices in a self-service restaurant. *Appl Psychol Health Well Being*. **7**(2) 149-166.
* Vanderlee, L., D. Hammond. 2014. Does nutrition information on menus impact food choice? Comparisons across two hospital cafeterias. *Public Health Nutr*. **17**(06) 1393-1402.
Viechtbauer, W. 2010. Conducting meta-analyses in r with the metafor package. *J Stat Softw*. **36**(3) 1-48.
Viechtbauer, W., M.W.L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. **1**(2) 112-125.
Walsh, J.W. 1995. Flexibility in consumer purchasing for uncertain future tastes. *Marketing Sci*. **14**(2) 148-165.
Wansink, B. 2015. Change their choice! Changing behavior using the can approach and activism research. *Psychol Mark*. **32**(5) 486-500.
* Wansink, B., H. Bhana, M. Qureshi, J.W. Cadenhead. 2016. Using choice architecture to create healthy food interventions in food pantries. *J Nutr Educ Behav*. **48**(7) S35-S36.
Wansink, B., P. Chandon. 2006. Can 'low-fat' nutrition labels lead to obesity? *J Marketing Res*. **43**(4) 605-617.
Wansink, B., P. Chandon. 2014. Slim by design: Redirecting the accidental drivers of mindless overeating. *J Consum Psychol*. **24**(3) 413-431.
* Wansink, B., A.S. Hanks. 2013. Slim by design: Serving healthy foods first in buffet lines improves overall meal selection. *PLoS One*. **8**(10) e77055.
* Wansink, B., D.R. Just, A.S. Hanks, L.E. Smith. 2013. Pre-sliced fruit in school cafeterias: Children's selection and intake. *Am J Prev Med*. **44**(5) 477-480.
Wansink, B., D.R. Just, C.R. Payne. 2012. Can branding improve school lunches? *Arch Pediatr Adolesc Med*. **166**(10) 967-968.
* Wansink, B., J. Kim. 2005. Bad popcorn in big buckets: Portion size can influence intake as much as taste. *J Nutr Educ Behav*. **37**(5) 242-245.
* Wansink, B., C.R. Payne. 2007. Counting bones: Environmental cues that decrease food intake. *Percept Mot Skills*. **104**(1) 273-276.
Wansink, B., C.R. Payne, P. Chandon. 2007. Internal and external cues of meal cessation: The french paradox redux? *Obesity*. **15**(12) 2920-2924.
* Wansink, B., K. van Ittersum. 2013. Portion size me: Plate-size induced consumption norms and win-win solutions for reducing food intake and waste. *J Exp Psychol Appl*. **19**(4) 320-332.
* Wansink, B., K. van Ittersum, J.E. Painter. 2006. Ice cream illusions: Bowls, spoons, and self-served portion sizes. *Am J Prev Med*. **31**(3) 240-243.
* Wansink, B., K. van Ittersum, C.R. Payne. 2014. Larger bowl size increases the amount of cereal children request, consume, and waste. *J Pediatr*. **164**(2) 323-326.

* Webb, K.L., L.S. Solomon, J. Sanders, C. Akiyama, P.B. Crawford. 2011. Menu labeling responsive to consumer concerns and shows promise for changing patron purchases. *Journal of Hunger and Environmental Nutrition*. **6**(2) 166-178.

Wertenbroch, K. 1998. Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Sci*. **17**(4) 317-337.

Wilson, A.L., E. Buckley, J.D. Buckley, S. Bogomolova. 2016a. Nudging healthier food and beverage choices through salience and priming. Evidence from a systematic review. *Food Qual Prefer*. **51** 47-64.

* Wilson, N.L., D.R. Just, J. Swigert, B. Wansink. 2016b. Food pantry selection solutions: A randomized controlled trial in client-choice food pantries to nudge clients to targeted foods. *J Public Health*. fdw043.

Zlatevska, N., C. Dubelaar, S.S. Holden. 2014. Sizing up the effect of portion size on consumption: A meta-analytic review. *J Marketing*. **78**(May) 140-154.

Online Appendix to

"Which Healthy Eating Nudges Work Best?

A Meta-Analysis of Field Experiments"

This online appendix contains the following sections:

Appendix A: Search strategy, keyword selection, and flow diagram

This meta-analysis focuses on articles describing nudge interventions, without restriction on the population. Specific search terms were developed in accordance with the SPICE (Setting, Population, Intervention, Comparison, Evaluation) framework (Booth 2006) (see Table A1). We mainly searched for interventions that involved nudging, choice architecture, or behavioral economics. We considered all articles published in the English language that reported a nudge intervention in a field setting. As shown in Table A2, key terms within the SPICE framework were combined using the Boolean operators "AND" and "OR."

Table A1: Application of the SPICE framework for keyword selection

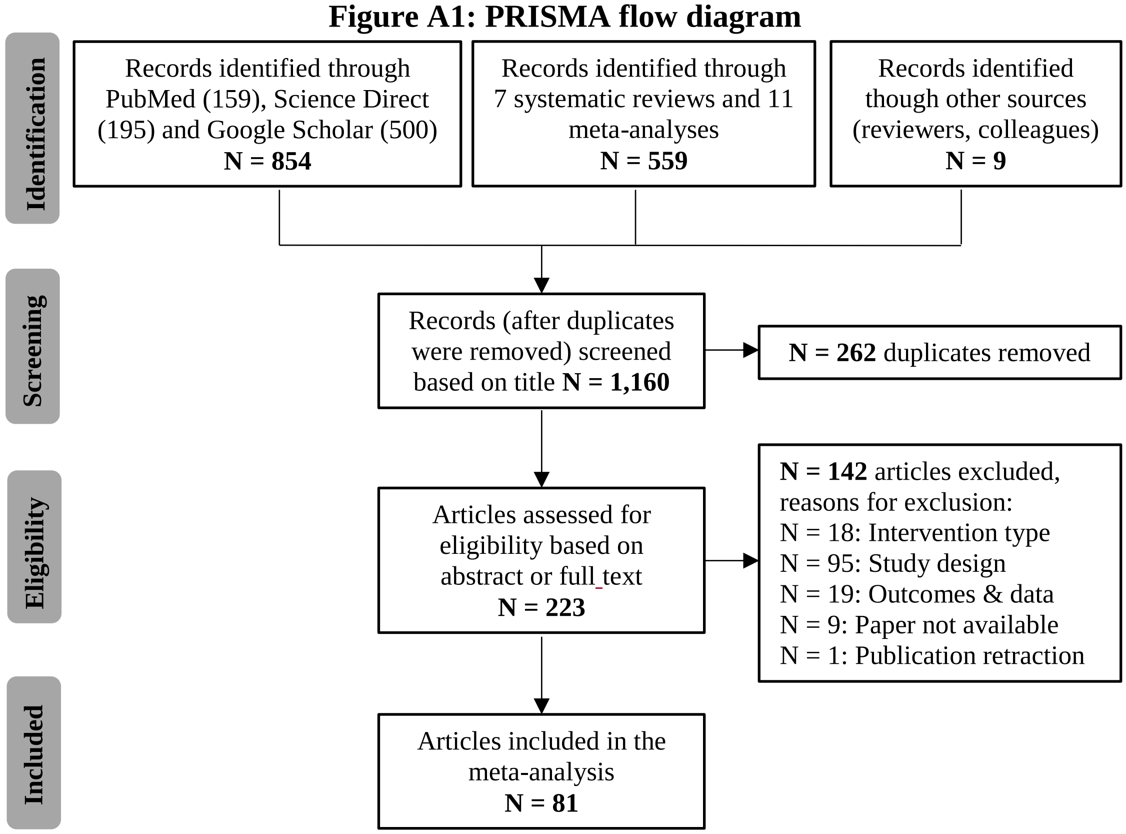
SPICE	Keywords
Setting	Food; eat*; fruit; vegetable; drink; beverage; diet; nutriti*; (un)healthy; calorie
Population	None assigned, interested in all populations
Intervention	Nudg*; choice architect*; behavioral economics; behavioral intervention
Comparator	Field study; field experiment
Evaluation	Selection; consumption; sales; choice

Table A2: Specific keywords used in database search using Boolean operators

Database	Keywords
Science Direct	("food" OR "eat" OR "fruit" OR "vegetable" OR "drink" OR "beverage" OR "diet" OR "nutriti" OR "calorie") AND ("nudg" OR "choice architect" OR "behavioral economics" OR "behavioral intervention") AND ("field study" OR "field experiment") AND NOT ("lab study" OR "lab experiment") AND ("selection" OR "consumption" OR "sales" OR "choice")
PubMed	("food" OR "eat*" OR "fruit" OR "vegetable" OR "drink" OR "beverage" OR "diet" OR "nutriti*" OR "calorie") AND ("nudg*" OR ("choice" AND "architect*"))
Google Scholar	"Nudge behavioral intervention food selection consumption choice"

Figure A1 reports the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Moher et al. 2009) flow diagram. The search strategy was first applied to three electronic databases: PubMed, Science Direct, and Google Scholar. The search was initially conducted in January 2016 and updated on October 2017. Within these first 854 articles, we focused on intervention-based articles as well as review-based articles. In fact, we found 7 systematic reviews and 11 meta-analyses on topic. We included in our identification base all 559 references cited in these review-based articles. Last, we also included 9 other references identified through other sources. After removing duplicates and references based on titles, we evaluated 223 articles. Of these, 142 articles were excluded for various reasons (see Figure 1), and 81 articles were included in the meta-analysis.

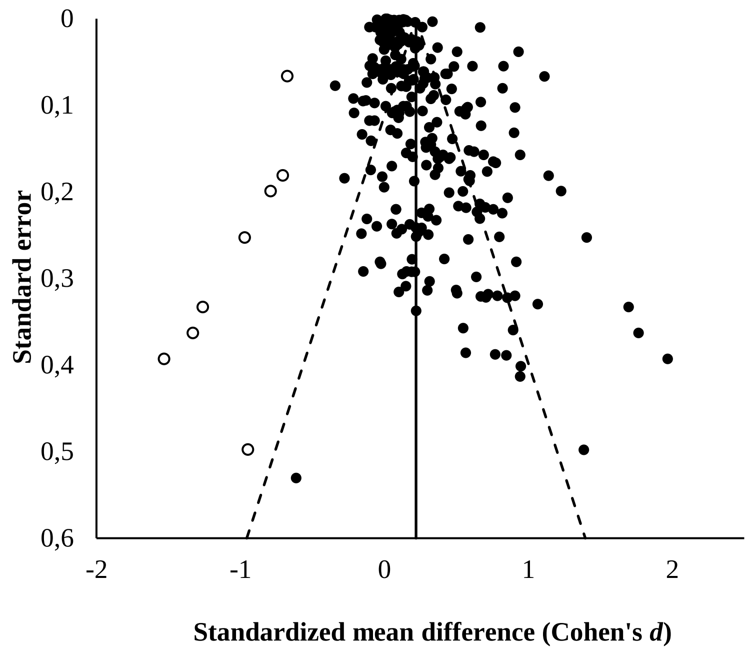
Figure A1: PRISMA flow diagram



Appendix B: Publication bias

In Figure A2, the funnel plot displays each observation as a function of the effect size (standardized mean difference or Cohen's *d*) and the standard error. Several observations appeared outside of the funnel on the right-hand side, suggesting a potential publication bias.

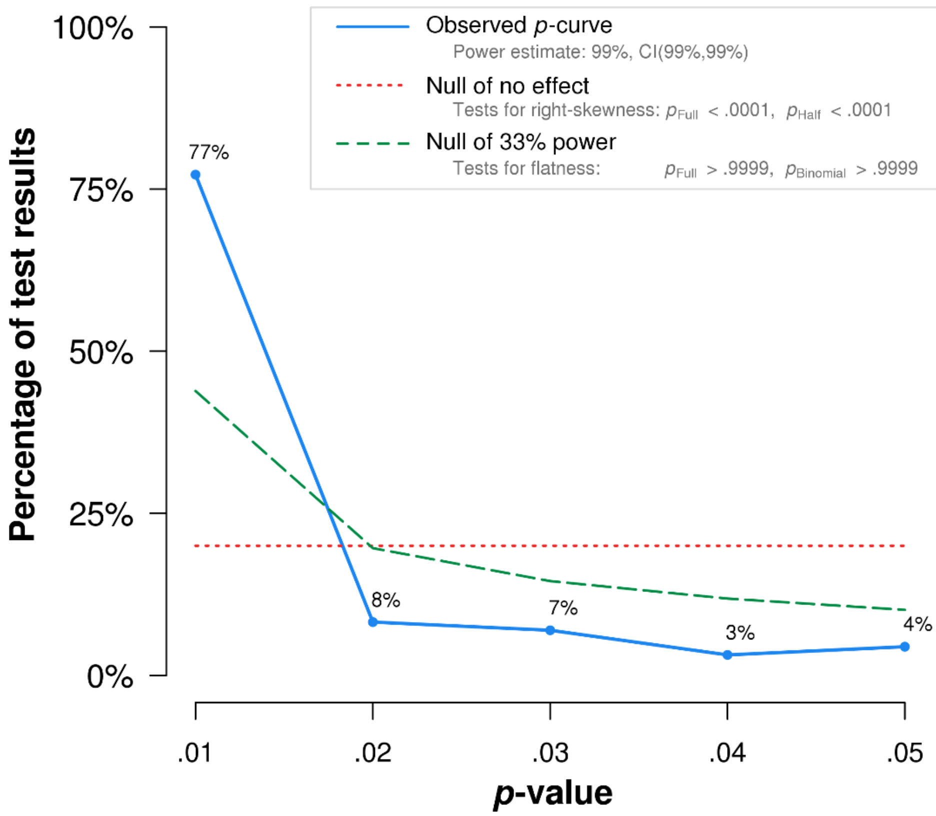
Figure A2: Funnel plot



First, using Duval and Tweedie's (2000) trim and fill method (not available for a three-level analysis), we included 8 missing studies (represented by the white dots in Figure 2), resulting in a slightly lower estimated effect size (*d* = .21, *se* = 0.2, *p* < .001) than in the unadjusted two-level analysis (*d* = .22, *se* = .02, *p* < .001). However, both estimated effect sizes are positive, considered of medium magnitude, and significantly different from zero.

Second, following Viechtbauer and Cheung (2010), we performed sensitivity analyses by removing 30 (30) observations for which the leverage (Cook's distance) was more than twice the average leverage (Cook's distance) in the overall sample. In both cases, the adjusted effect size is only slightly lower (respectively *d* = .27, *se* = .03, *p* < .001; and *d* = .23, *se* = .03, *p* < .001) compared to the original three-level analysis (*d* = .27, *se* = .03, *p* < .001).

Figure A3: *P*-curve



Note: The observed *p*-curve includes 158 statistically significant ($p < .05$) results, of which 143 are $p < .025$. There were 119 additional results entered but excluded from *p*-curve because they were $p > .05$.

Third, we also produced the *p*-curve (Simonsohn et al. 2014b) for all effect sizes included in our sample. As shown in Figure F3, the *p*-curve is strongly right-skewed ($p < .001$) and indicates the presence of evidential value. Note that 119 out of 277 effect sizes (43%) were not significant. A closer examination of non-significant effects shows that only 16 out of 81 articles (19%) published only non-significant results, 31 papers (38%) published significant and non-significant effect sizes (e.g., the effect size is significant for fruits but not significant for vegetables) and 31 papers (38%) published only significant results. Using the procedure presented in Simonsohn et al. (2014a), we also estimated the corrected *p*-curve effect size on the sample of 158 significant effect sizes. The corrected effect size estimated through bootstrap ($d = .21$, $se = .03$, $p < .001$) was lower than the "naïve" estimate on the 158 significant effect sizes without the *p*-curve correction ($d = .39$, $se = .03$, $p < .001$), and lower than the "earnest" overall average effect size including all observations

($d = .27$, $se = .03$, $p < .001$). Last, the *p*-curve effect size is similar to the average effect size from the overall model including all observations and covariates ($d = .22$, $se = .04$, $p < .001$). All four estimates are positive, considered of medium magnitude, and significantly different from zero.

Last, in light of the criticisms regarding some of the studies conducted by the Cornell Food and Brand Lab (Robinson 2017) included in this meta-analysis, we added to the multivariate model a binary variable controlling for the 32 observations originating from this lab. This variable was highly insignificant ($p = .75$). Moreover, we do not include Wansink et al. (2012) in our meta-analysis because the paper was recently retracted. Including the study would not affect the results in any way.

Appendix C: Subcategory analyses

In Table A3, we report the estimates for the multivariate model using the detailed categorization for intervention type. On average, the estimates of effect sizes are 12% lower in the multivariate analysis than in the univariate ones.

Table A3: Estimates for detailed intervention types

		Univariate		**Multivariate**		**Difference**	
	k	***d***	***se***	***d***	***se***	**Raw**	**Percent**
Cognitive interventions							
Descriptive labeling	31	.05	.06	.05	.09	-.01	-17%
Evaluative labeling	34	.14*	.06	.13*	.06	-.01	-6%
Salience enhancements	25	.19*	.08	.10	.08	-.09	-46%
Mixed: Cognitive only	23	.15*	.06	.18*	.08	.02	14%
Affective interventions							
Healthy eating cues	35	.32*	.06	.25*	.07	-.07	-21%
Sensory cues	12	.29*	.10	.22*	.10	-.07	-23%
Behavioral interventions							
Convenience enhancements	56	.41*	.05	.33*	.06	-.08	-18%
Plate & portion size changes	17	.71*	.09	.55*	.10	-.16	-22%
Mixed interventions							
Mixed: Cognitive present	37	.20*	.07	.24*	.08	.04	20%
Mixed: Cognitive absent	7	.29*	.10	.25*	.11	-.04	-14%
Weighted average	277					-.04	-12%

$^{*}p < .05$

Appendix D: Interaction analyses

The multivariate model shown in equation 7 and Table 4 does not lend itself to interaction analyses because it already requires estimating 16 coefficients. Adding interactions would require estimating more than 30 coefficients, which would quickly exhaust the available degrees of freedom, especially for some combinations with few or no data. Hence, we added interactions to a simplified model with only one coefficient for each of the 8 types of predictors. Essentially, we recoded or grouped categories together to provide a single (linear or binary) coefficient per predictor.

Simplified model. First, we used linear coding for intervention type (-1 = cognitive, 0 = affective, and 1 = behavioral). We included the mixed interventions into the cognition-affect-behavior categorization according to the first stage that they sought to influence. Hence, mixed interventions including at least one cognitive intervention ("mixed: cognitive present") were included among cognitive interventions. Mixed interventions excluding a cognitive one ("mixed: cognitive absent") were included among affective ones (no study used a mixed intervention of behavioral nudges). We also used a linear coding for outcome measure (-1 = total eating, 0 = healthy eating and mixed eating, and 1 = unhealthy eating), while healthy and mixed eating were grouped together because of a small difference in effect sizes (see Table 2). Third, location was ANOVA coded (-½ for grocery stores vs. ½ for onsite or offsite cafeterias), while offsite and onsite locations were grouped together because of a small difference in effect sizes (see Table 2). Fourth, study design was ANOVA coded (-½ for pre-post and ½ for single and double difference), while single and double difference were grouped together because of a small difference in effect sizes (see Table 2). The rest of the variables were unchanged; that is, outcome behavior (selection vs. consumption), age (children vs. adults), and country (other vs. US) were ANOVA coded, while duration was measured in weeks and mean centered.

Results. We first estimated a model with only main-effects (Model 1). Its R^2 decreased only slightly compared to the model with the categorical coding reported in the main text (from .46 to .44), and none of the hypotheses tests changed. To examine when each type of intervention is most effective, we included an interaction term between intervention type and all the other predictors. Table A4 shows that the effects of intervention type are stronger 1) for consumption than for selection, 2) for adults than for children, and 3) in the US than in other countries.

Table A4: Interactions analyses

	Model 1		**Model 2**	
Predictor	$\boldsymbol{\beta}$	**se**	$\boldsymbol{\beta}$	**se**
Constant	.21***	.04	.20***	.04
Intervention type (cognitive to affective to behavioral)	.12***	.03	.10*	.05
Outcome behavior (selection vs. consumption)	.02	.04	.05	.04
Outcome measure (total to healthy/mixed to unhealthy)	.11***	.03	.08*	.04
Population setting (cafeterias vs. grocery stores)	.13*	.06	.09	.07
Population age (adults vs. children)	.04	.05	.07	.04
Population country (US vs. other)	.11*	.05	.13**	.05
Study duration (duration in weeks, mean centered)	-.002	.001	-.004	.003
Study design (single & double difference vs. pre-post)	.12**	.04	.09*	.04
Outcome behavior × intervention type			.14**	.05
Outcome measure × intervention type			-.02	.04
Population setting × intervention type			-.06	.07
Population age × intervention type			.16***	.05
Population country × intervention type			.16**	.06
Study duration × intervention type			-.003	.003
Study design × intervention type			-.01	.05
K (observations)	277		278	
N (studies)	87		87	
R^2	44%		56%	
LR test vs. intercept-only model	65***		87***	

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

References

Booth, A. 2006. Clear and present questions: Formulating questions for evidence based practice. *Library hi tech*. **24**(3) 355-368.

Duval, S., R. Tweedie. 2000. Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. **56**(2) 455-463.

Moher, D., A. Liberati, J. Tetzlaff, D.G. Altman, P. Group. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. **6**(7) e1000097.

Robinson, E. 2017. The science behind smarter lunchrooms. *PeerJ Preprints*. **5** e3137v3131.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014a. P-curve and effect size: Correcting for publication bias using only significant results. *Perspect Psychol Sci*. **9**(6) 666-681.

Simonsohn, U., L.D. Nelson, J.P. Simmons. 2014b. P-curve: A key to the file-drawer. *J Exp Psychol Gen*. **143**(2) 534.

Viechtbauer, W., M.W.L. Cheung. 2010. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. **1**(2) 112-125.

Wansink, B., D.R. Just, C.R. Payne. 2012. Can branding improve school lunches? *Arch Pediatr Adolesc Med*. **166**(10) 967-968.