Combining Systematic Reviews with Modeling to Inform Public Health Decision-Making: Case Study of the Childhood Obesity Intervention Cost-Effectiveness Study

Stephen C. Resch (Harvard University)*, Y. Claire Wang (Columbia University), Michael Long (Harvard University), Catherine Giles (Harvard University), Angie Cradock (Harvard University), Steve Gortmaker (Harvard University)

Working Paper prepared for:

**Methods for Research Synthesis:
A Cross-Disciplinary Workshop**

Harvard Center for Risk Analysis

October 3, 2013

www.hcra.harvard.edu

*Corresponding author: sresch@hsph.harvard.edu

TITLE: Combining Systematic Reviews with Modeling to Inform Public Health Decision-Making: Case Study of the Childhood Obesity Intervention Cost-Effectiveness Study

AUTHORS:

*Stephen Resch, MPH PhD Center for Health Decision Science, Harvard School of Public Health, 718 Huntington Ave, Boston MA 02115

Y. Claire Wang, MD ScD, Department of Health Policy and Management, Columbia Mailman School of Public Health

Michael Long, PhD, Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Ave, 7th Floor, Boston, MA 02115

Catherine Giles, Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Ave, 7th Floor, Boston, MA 02115

Angie Cradock, Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Ave, 7th Floor, Boston, MA 02115

Steve Gortmaker, Department of Social and Behavioral Sciences, Harvard School of Public Health, 677 Huntington Ave, 7th Floor, Boston, MA 02115

*Corresponding author: sresch@hsph.harvard.edu

Abstract

The Childhood Obesity Intervention Cost-Effectiveness Study (CHOICES) provides a useful case study of the adaptation, integration, and application of evidence synthesis methods for public health decision making. CHOICES is conducting systematic reviews tightly coupled with decision modeling to evaluate about 40 population-level childhood obesity prevention strategies. Focusing on evidence from actual interventions, rather than hypothetical behavior or BMI changes, the project aims to synthesize existing evidence and transform it into information sought by policymakers and other stakeholders. Intervention logic models--which detail the intermediate steps between intervention recruitment and health outcomes--provide structure for both the systematic review process and decision modeling. Drawing on a broader range of studies than most systematic reviews would include, evidence of intervention effect along multiple pathways is stitched together and converted to a standard set of outcome measures with a model. The systematic review process pays particular attention to external validity of studies and the contextual factors relevant to the implementation or adoption of particular strategies so that these can be incorporated into modeling scenarios. CHOICES modeling scenarios will estimate the potential and likely population-level health impact and cost-effectiveness of each strategy at a national scale, for a common time horizon, and under a consistent set of alternative assumptions regarding key unknown factors.

1. Introduction

The well-documented health impacts, societal costs, and worsening trends of obesity in the United States create a strong rationale for public health intervention [1, 2]. A number of programs and policies have been implemented or proposed to counter the prevailing energy imbalance during childhood. Potentially, large amounts of money and human effort could be invested scaling up such approaches. While many studies have been conducted, little evidence of these obesity prevention strategies' potential for large-scale impact over time, or their cost-effectiveness, is available to support decision making. The Institute of Medicine recently highlighted the need for an evidenced-based public health approach to childhood obesity that undertakes evidence synthesis with a broad 'systems' perspective and a greater emphasis on contextual factors affecting impact of potential strategies [3].

The Childhood Obesity Intervention Cost-Effectiveness Study (CHOICES) provides a useful case study of the adaptation, integration, and application of evidence synthesis methods for public health decision making. CHOICES is conducting systematic reviews tightly coupled with decision modeling to evaluate about 40 population-level childhood obesity prevention strategies such as excise tax on sugar-sweetened beverages, menu labeling in restaurants, healthy eating incentives in food stamp programs, regulation of screen time in early education centers, and after-school programs to boost physical activity. As shown in Figure 1, for each of these strategies selected and specified with the aid of experts participating in a structured stakeholder process, the totality of what is known about its effectiveness from past observational and experimental studies is being gathered and summarized through systematic reviews conducted by CHOICES research teams. Subsequently, most strategies will be evaluated in a common decision analytical framework using a model to integrate various observed effectiveness outcomes into simulation based projections of population-level health impact and cost-effectiveness.
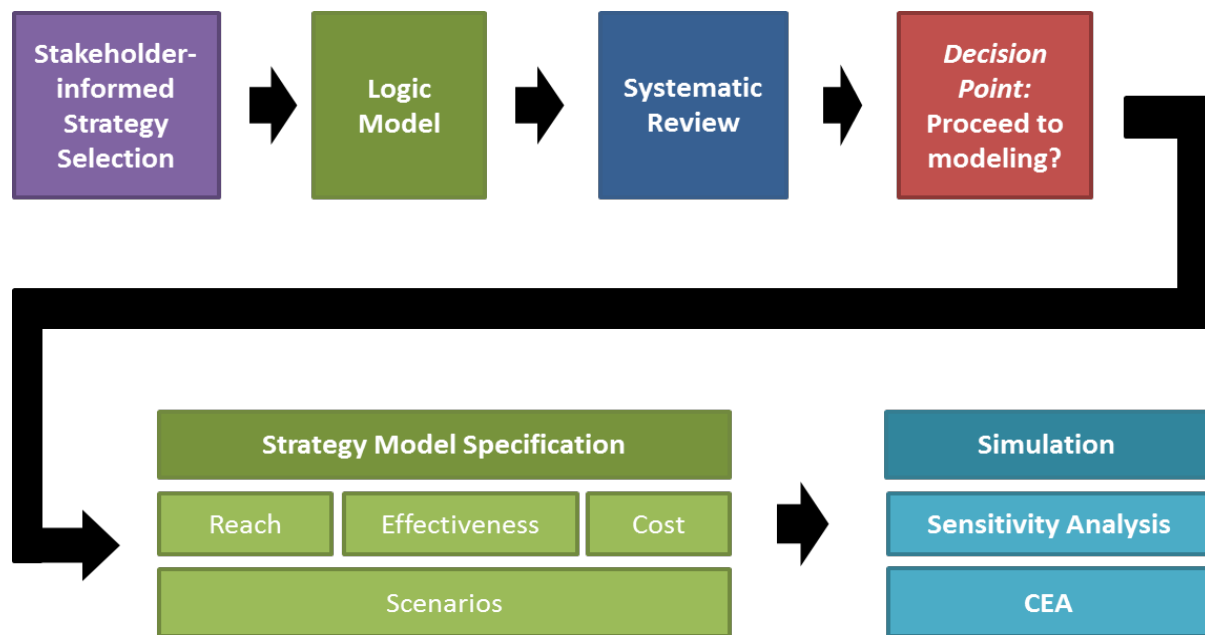
Figure 1. The CHOICES Evidence Synthesis Process.

The coupling of systematic review and decision modeling to assess cost-effectiveness is a distinguishing feature of the CHOICES project. Other initiatives to systematically gather and evaluate evidence regarding community preventive services have stopped with recommendations flowing directly from systematic reviews (e.g. CDC Guide to Community Preventive Services [4], and AHRQ's recent comparative effectiveness review of childhood obesity interventions [5]). While a number of simulation models have been developed to predict future prevalence and consequences of the obesity epidemic as well as a more limited set of interventions [6, 7], no other project has attempted to use decision modeling as a way to translate evidence across a wide range of childhood obesity strategies into a common set of indicators most meaningful for policy in the United States.

In Australia, the 'Assessing Cost-effectiveness' (ACE) program has undertaken evidence synthesis and cost-effectiveness modeling of hundreds of health interventions including several childhood obesity interventions using a common framework [8, 9]. Building on the ACE-Prevention and ACE-Obesity approaches, but utilizing a more flexible microsimulation model, CHOICES modeling scenarios will estimate the potential and likely population-level health impact and cost-effectiveness of each strategy at a national scale, for a common time horizon,

and under a consistent set of alternative assumptions regarding key uncertain factors such as the 'tracking' of obesity from childhood into adulthood and the effect of BMI on disease risk and survival.

2. Systematic Reviews and Decision Modeling as tools of Evidence-based Public Health

Limits of traditional systematic reviews

Traditional systematic reviews provide a summary of what is known about the effect size of an intervention, an indication of the uncertainty about that effect size, and information about the quality of past studies measuring the interventions' effect. When a group of studies on a particular research question are sufficiently similar, more statistically rigorous meta-analysis techniques are possible to quantitatively summarize intervention effect. With an adequate number of studies, random effects meta-regression approaches can be used to explore issues of study quality and subgroup heterogeneity [10, 11]. Still, systematic reviews are essentially structured summaries of the existing evidence base, constrained by the content of the underlying studies and strict rules to protect against subjective value judgments of the reviewer. Differences in the study design, measured outcomes, and populations of those studies pose challenges for systematic reviews. Moreover, with the exception of network meta-analysis of randomized clinical trials [12], traditional systematic review methods are not well suited to making comparisons across different interventions that have not been compared in the context of the original studies. As such, traditional systematic reviews by themselves often cannot provide the key information sought by decision makers on a particular issue. Strengthening the link between evidence synthesis and policy making requires systematic reviews that are more inclusive with regard to study design, and which pay greater attention to contextual factors that influence generalizability [13, 14]. These reviews would be further enhanced by linking them to other evidence through decision modeling, which extends the range of outcomes that can be considered, and allows the integration of effectiveness data with evidence or assumptions regarding reach, effect-moderating heterogeneity, and cost [15].

Importance of external validity when evaluating public health strategies

As others have noted, the principles of evidence-based medicine that have driven the development of systematic review and meta-analysis methods are frequently not applicable to public health, where interventions are 'programs' delivered to groups of people (e.g. in schools, workplaces, neighborhoods) , or policies (e.g. taxes) applied to broad populations [16]. The recent focus on 'patient-centered outcomes research' in clinical medicine may signal a shift in the evidence-based medicine paradigm; but, traditionally, concern about population characteristics, stakeholder preferences, required human and financial resources, and environmental or organizational context have weighed more heavily on public health decision making than on decision regarding medical treatments for individual patients. Accordingly, evidence synthesis methods for public health must place greater emphasis on external validity.

While the power of experimental designs is acknowledged by public health researchers, randomized control trials to evaluate preventative health programs and policies that are delivered at the community or population level are challenging and often infeasible. Effects may take many years to manifest, and the ability to control (or measure) exposure to other factors that influence the outcomes is often limited. Moreover, the transferability of a successful public health program from one setting to another may depend on factors that are not easily detected through an evaluation designed to maximize internal validity [17]. For these reasons, systematic reviews of public health programs and policies need to accommodate other study designs, such as natural experiments and change-in-change observational studies, and be attentive not only to observed effectiveness, but also generalizability, including issues of acceptability, potential reach, and cost [13, 14]. In the ACE-Prevention and ACE-Obesity studies from which much of the CHOICES methodology is derived, these contextual factors, referred to as 'second-stage filters', were 'equity', 'strength of the evidence', 'feasibility of implementation', 'acceptability to stakeholders', 'sustainability' and 'potential for side-effects' [8, 9]. The evidence synthesis considered these factors when extracting evidence from published studies during systematic reviews and also through a stakeholder process that informed the assessment of each strategy.

Role of decision models

Adding decision modeling to a systematic review allows the synthesized evidence on various related intermediate effect measures (e.g. short-term weight change, body mass index change,

energy intake changes, physical activity changes, food purchase changes) to be extended to key common outcomes of interest to patients and policymakers (e.g. long-term body mass index changes, changes in BMI-attributable disease, survival, and quality-adjusted life years). As a result, a larger number of empirical studies can contribute to the estimate of an intervention's impact, and several interventions can be compared using common metrics. Further, modeling can be used to explore heterogeneity across subpopulations when detected in the empirical evidence. Models can account for this heterogeneity in scale-up scenarios that predict the impact of interventions when implemented broadly.

Decision modeling is a method of evidence synthesis that reaches beyond the scope of traditional systematic review or meta-analysis. While it is not uncommon for systematic reviews to conclude that the evidence is 'insufficient' for making recommendations (e.g. [18, 19]), decision analysis starts from the premise that a decision must be made given the available evidence. Decision analysis has an obligation to predict the best course of action based on a set of (evidence-informed) assumptions. To do so, it codifies evidence about the effect and cost of different courses of action, as well as the decision maker's values and preferences. The resulting "manufactured" output is generally viewed as being in a different class than the results produced in systematic review [15].

Skepticism about the credibility of model-based evidence stems from a perception that modeling methods are not sufficiently standardized, complex models are not sufficiently transparent, and are vulnerable to conscious manipulation or unconscious propagation of the modeler's own biases. Indeed, models typically combine information from multiple evidence synthesis efforts corresponding to different model parameters, and must fill in evidence gaps regarding model structure or parameter values with uncertain assumptions. Ideally, the assumptions used to fill evidence gaps are obtained through formal expert elicitation, tested in sensitivity analysis, and communicated to consumers of model output. But modelers often fail to conduct *systematic* reviews or use rigorous expert elicitation to inform key model parameters, and instead make arbitrary decisions about what studies to use to inform the model. Even when multiple studies are found through review, modelers often just use the effect size from the "best" study (i.e. the largest, most precise, strongest design) as their base case model input [20]. The interdisciplinary

team supporting the tight linkage between the systematic review, stakeholder process, and decision modeling activities of CHOICES is intended to avoid some of these pitfalls.

3. CHOICES Microsimulation Model

There are five challenging features of childhood obesity and strategies to prevent it that make a decision modeling a particularly useful tool in the evidence synthesis process. First, estimating the likely national impact of particular strategies that have only been tested on a small or modest scale, requires mapping them to target populations at the national and state level and making assumptions about their reach within these populations. Second, the 'natural' relationship between childhood obesity and adult obesity is a critical factor in estimating the value of prevention since many of the health effects occur at older ages [21-23]. Third, many studies of childhood prevention strategies have shorter time horizons and measure intermediate outcomes. The model is important for linking intermediate outcomes such as increases in physical activity or decreases in energy intake to weight change. Even when weight change is a measured outcome, studies are often too short for the full impact on bodyweight to manifest, because the dynamics of energy imbalance result in slow, non-linear relationship between energy intake and bodyweight change. Fourth, when a strategy does impact physical activity or leads to a change in weight, it is beneficial to decision makers to translate this impact into health effects since inactivity and excess bodyweight are not intrinsically bad. Fifth, there is no direct evidence that the weight-reducing impact of strategies implemented in childhood will be maintained. Necessary assumptions about the 'maintenance of effect' must be made drawing on theory, parallel evidence, and expert opinion. The sensitivity of conclusions to these assumptions must be tested. The overarching advantage of the modeling approach is that it provides a framework that enables consistency in the approach to each of these challenges, thus improving the credibility of *comparisons* between strategies. In this subsequent sections, we briefly describe the CHOICES approach and how it addresses each of these five challenges.

Population and natural history

Data from the United States Census, American Communities Survey (ACS), National Health and Nutrition Survey (NHANES), National Study of Children's Health (NSCH), National Longitudinal Study of Youth (NSLY), National Longitudinal Study of Adolescent Health

(AddHealth) are used to create the model's synthetic population of individuals that accurately represent the covariance of age, sex, race, household income, education, geographical location, height, weight, food consumption patterns, physical activity level, obesity-related disease risk, and survival probability observed in the current US population. Natural history height and weight 'trajectories' are generated for simulated individuals from analysis of longitudinal data on growth and calibrated to reflect observed cross-sectional data and population-level secular trends.

An individuals' 'body mass index' (BMI) can be calculated from their weight and height (BMI = weight (in kilograms) divided by height (in meters) squared). The prevalence of obesity in the US is defined as a body-mass index of more than 30 kg/m² in adults and according to standards for children specific to age and gender [24]. Of particular interest is the 'tracking' of childhood BMI into adulthood. A frequently cited rationale for childhood obesity prevention is the downstream health impact of high BMI in adulthood. Certainly, childhood obesity, especially when it develops or persists into adolescents, is a strong risk factor for adult obesity. However, many obese children (especially young ones) will *not* become obese adults [21-23]. Thus, the distribution of BMI trajectories across the simulation population must not only reflect population averages by age, but also the 'noisiness' of fluctuations in BMI in individuals over their life.

Intervention evidence pathways

In the model, the program or policy's effect is simulated by adjusting parameters regarding consumption of particular foods (e.g. sugar sweetened beverages), total energy intake (calories per day), physical activity level (MET-hours per day). Studies evaluating the interventions often report their effect on anthropomorphic outcomes directly in terms of weight change or BMI change. In these cases, measured outcomes could be used fairly directly in the model. However, much of the evidence of intervention effect focuses on more proximate intermediate outcomes. For example, studies of "front of package nutrition labeling" policies may measure impact on purchase choices in particular venues such as supermarkets or fast food restaurants. Additional evidence for the downstream portion of the pathway from purchases to actual energy intake may only be available from separate studies or missing entirely.

Dynamics of energy balance

Changes in individual BMI in the CHOICES model are estimated based on daily energy balance equations applied for the duration of each time period (e.g. each simulated month)[25]. Most previous obesity policy evaluation models have relied on simplifications of the relationship between changes in energy intake and weight change (e.g. a reduction of 3,500 kcal leads to a 1 lb weight loss). Unfortunately, this linear approximation substantially overestimates both the speed and magnitude of weight changes due to sustained reductions in energy balance [26]. Earlier models did not recognize that as body weight increases, the energy required to maintain that weight also increases. Likewise, the amount of energy consumed during a fixed amount of physical activity (say, jogging for 20 minutes) increases with bodyweight. Thus for a person 'at equilibrium' at some weight level, a fixed decrease in daily caloric intake with no change in physical activity level will generate a finite amount of weight loss over a period of about three years, with 50% of the loss realized in the first year [25]. Equations for calculating bodyweight change in adults of particular height and starting weight for a fixed daily energy intake and physical activity level over a fixed time period have been developed and validated [27]. Similar models for children, which account for growth-related energy needs, have also been developed [28]. These algorithms have been embedded in the CHOICES model so that the dynamics of energy imbalance are accurately reflected when predicting the impact of obesity prevention strategies.

Disease risk and health outcomes

The modeling of BMI's effect on disease, survival, and disability-adjusted life years (DALYs) is common to all interventions. The incidence of 9 BMI-related diseases is explicitly modeled (including stroke, ischemic heart disease, hypertensive heart disease, diabetes mellitus, osteoarthritis, post-menopausal breast cancer, colon cancer, endometrial cancer, and kidney cancer), and virtual individuals in the simulation with elevated BMI (and other risk factors such as age and sex) will face higher risk of contracting and dying from those diseases. Accordingly, interventions that change BMI lead to lower incidence of certain diseases, improved survival and averted treatment costs. Age- and sex-specific disease incidence and BMI relative risks were generated by synthesizing evidence from numerous sources including the CDC Wonder databases, American Heart Association, National Heart Lung and Blood Institute, and published

literature. Disease-related medical costs were obtained primarily from statistical analysis of the Medical Expenditure Panel Survey (MEPS). The model accounts for competing (non-obesity related) health risks in the computation of net survival gains, disability-adjusted life years (DALYs), and net costs. Uncertainty in this portion of the model is propagated through to final results, but is common to all interventions.

Consistent assumptions in policy-relevant scenarios

To estimate the impact of each childhood obesity prevention strategy on the full US population (or selected subsets) over a long time horizon (e.g. 30 calendar years from 2013- 2042) or alternatively, over the *lifetime* of individuals in 30 birth cohorts (e.g. those born between 1998 and 2027), the model requires data indicating the likely reach of the strategy, the short-term effect of the strategy on individuals' energy balance, assumptions about how the effect will be maintained over a person's lifetime (especially if individuals' exposure to the program or policy is time-limited), and a projection of what would happen in absence of intervention.

Key unknown parameters such as those related to the maintenance of intervention effect over time (if an intervention lowers a teenager's BMI by 10%, what impact does this have on his or her BMI 30 years hence) are handled with baseline, optimistic, and pessimistic scenarios. A major strength of the project is that all 40 interventions will be evaluated using consistent assumptions regarding these uncertainties, such that comparability of results and utility for informing policy is maximized.
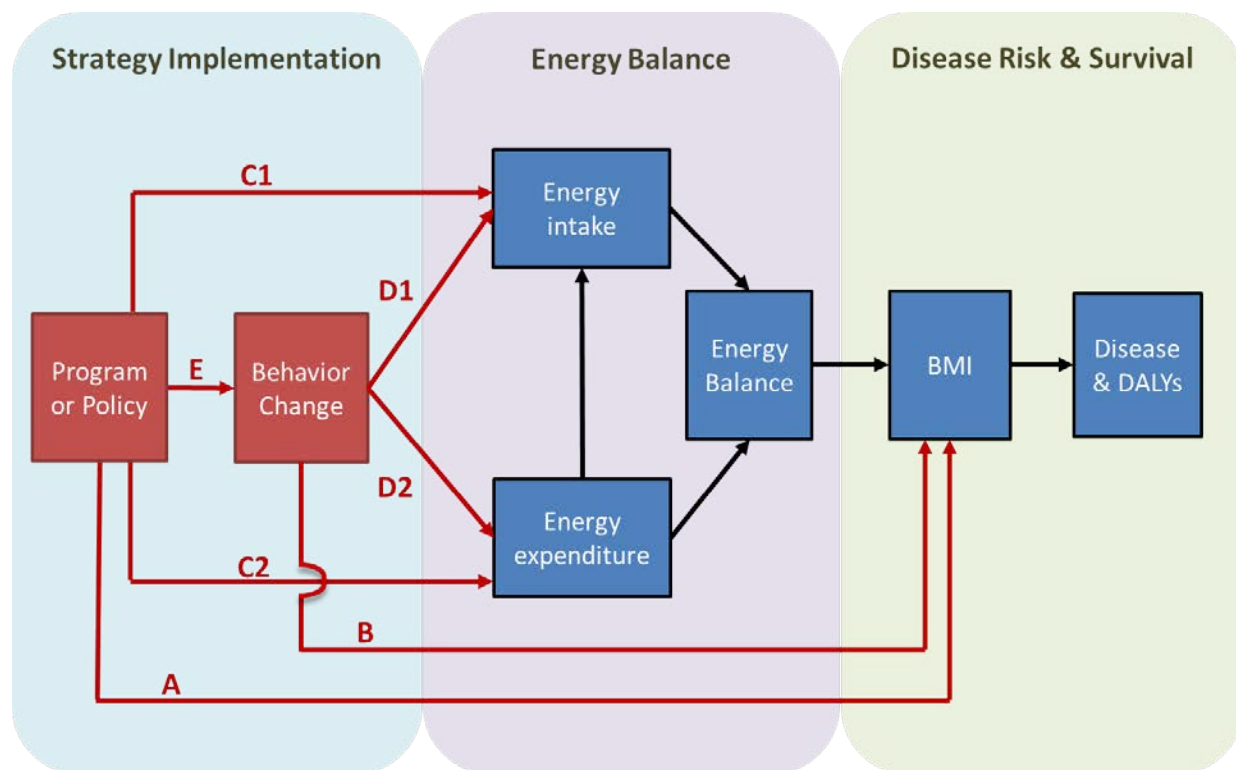
4. Systematic Review Process

The CHOICES process for systematic review of the effectiveness of childhood obesity prevention strategies is tailored to supporting the decision modeling activity. Recognizing that modeling allows the chaining together of evidence along an 'evidence pathway' between the program or policy implementation and key intermediate outcomes of interest, the systematic review search strategies are guided by logic models mapping possible evidence pathways for each strategy. Inclusion criteria allow studies with the wide range of outcome measures that the model can accommodate and translate to common outcomes of interest. Perhaps in contrast to many clinically-oriented systematic reviews, the CHOICES reviews gathers evidence regarding the preventions strategies' potential 'reach', cost, and side effects (i.e. non-obesity specific

benefits or harms). Access to a regularly convened group of expert stakeholders ensures the relevance of model output, due attention to contextual factors, and a source of expert opinion to fill evidence gaps.

Logic models

To guide each systematic review, a logic model is constructed, mapping the intermediate steps between the intervention implementation and final outcomes of interest (generally BMI change). Figure 1 depicts a general version of these logic models. In it, the blue boxes connected with black arrows indicate aspects of the simulation model's structure for handling dynamics of weight gain and disease risk over the simulated lives of individuals in the virtual population. These are handled in a consistent way common to all interventions. The red boxes connected

Figure 2. General logic model mapping potential evidence pathways. The relationship between the outcomes in blue boxes represented by the black arrows are modeled in a consistent way common to all interventions. CHOICES systematic reviews focus on identifying and extracting evidence from studies that inform the red arrows, with the goal of understanding the linkage between the policy or program and one or more of the blue outcomes.

by red arrows indicate the program or policy's effects along possible evidence pathways. Systematic reviews focus on identifying and extracting evidence from studies that inform the red arrows, with the goal of understanding the linkage between the policy or program and one or more of the blue outcomes.

An intervention will achieve a certain level of participation and individual exposure, which may stimulate behavior change such as increasing physical activity or reducing intake of energy dense food of low nutritional value. Studies of the intervention may measure a variety of relationships represented by the lettered red arrows. Evidence along Pathway A or a combination of Pathway E and B is most preferred, as it provides direct evidence using an intermediate outcome that is an anthropomorphic measurement. For example, a randomized trial measuring exposure to an intervention in a group and the difference in their BMI change over the study horizon, compared to the BMI change in a control group (Pathway A) would be given substantial weight in the evidence synthesis assuming it was well-executed and judged to have reasonable external validity. Studies measuring body weight are nearly as good, but require making assumptions about height (e.g. based on age and sex) to estimate BMI change.

Some studies might only measure the effect of the behavior change (e.g. increased minutes of moderate-to-vigorous physical activity or reduction in daily consumption of sugar-sweetened beverages). In these cases, the scope of the measurement is important since physical activity in one part of the day (e.g. during an intervention) may reduce (or increase) physical activity, or energy consumption at a later part of the day. Studies that cannot account for these possibilities in their design are considerably weaker in terms of their ability to inform the effect on BMI, given the phenomenon of compensation is well recognized [29].

When studies do not directly measure an impact on BMI, but do measure an impact on total daily energy intake or energy expenditure (e.g. C1), that effect can be converted in the model to corresponding BMI change. However, recent experience shows that a naïve approach could produce misleading conclusions. For example, the expected weight change from reduction in daily energy intake of one serving of sugar-sweetened beverage (~140 kcal) over four years would lead to an average reduction of approximately 14 pounds bodyweight in a population with starting bodyweight distribution typical of US adults [25, 26]. But, change-in-change analysis of large longitudinal datasets (e.g. Nurses' Health Study), found this reported change in intake led

to only about 1 pound in bodyweight over four years [30]. Thus, naively analyzing an intervention such as an excise tax on sugar-sweetened beverages using evidence pathway 2→4→5 in Figure 3, will lead to a much large estimated impact than using evidence along pathway 1. The analogous problem has been observed for television viewing in children. A controlled trial found that a difference in television viewing of 1 hour per day was associated with an increase of 106 kcal per day and a 0.84 pound change in weight over 21 months[31]; whereas, dynamic energy balance models predict that a daily change of 106 kcal would lead to a bodyweight change of 10 pounds.
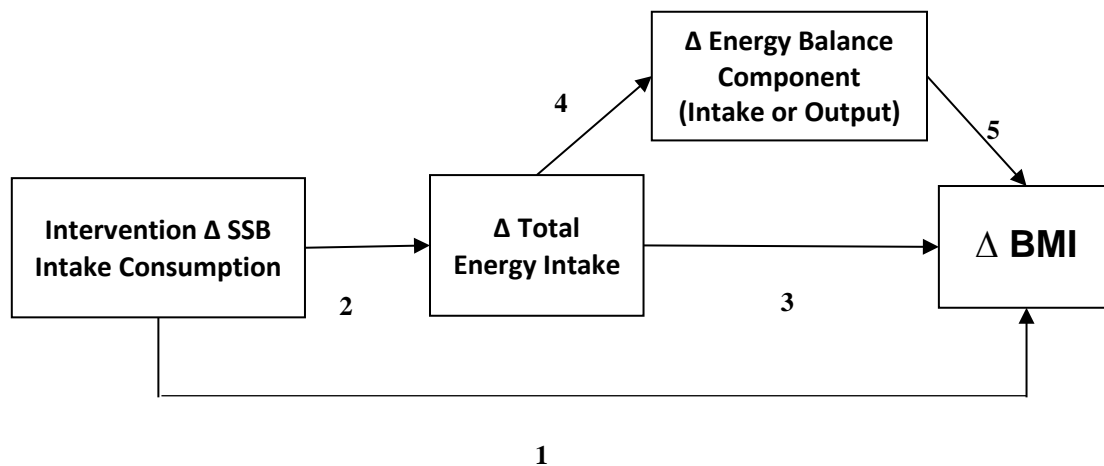
Figure 3. Example of an Intervention Logic Model for a policy intervention to change sugar-sweetened beverage (SSB) intake.

The flexibility of the model to allow several impact measures enables more inclusiveness in the systematic review and consideration of a wider evidence base. Still, a challenge for CHOICES will be to develop methods to avoid the potential pitfalls of making naïve assumptions connecting evidence of intermediate outcomes to bodyweight and BMI. Fortunately, the modeling approach supports the evaluation of uncertainty by using the full range of evidence to test alternative evidence pathways in scenario analyses. A consistent finding across a range of

evidence pathways would reveal consensus across disparate studies that may be difficult to identify using standard systematic review and meta-analysis techniques. Likewise, if across numerous interventions, a consistent bias is observed in the estimated impact modeled through "stitched" energy balance pathways relative to direct pathways between the intervention and BMI (or weight), it may be possible to develop correction factors that would improve the accuracy of estimates based on indirect evidence.

Evidence typology

While the use of a microsimulation model allows for incorporation of information from a broad range of studies, the CHOICES systematic review process prioritizes studies with strong research designs in order to limit the risk of bias. Reviewed studies are assigned an evidence level using the CHOICES Evidence Typology (Figure 4), which is used to compare evidence across evidence pathways for a given intervention and overall evidence level across interventions. Randomized trials are generally viewed as having the best potential to provide the strong evidence of intervention efficacy. In some cases, natural experiments may provide the only available evaluation of obesity prevention policies or programs. Change-in-change observational studies that utilize appropriate within-person analyses to control for time-invariant confounding are often the best available source of the best long-term evidence of the relationship between behavior change and weight change. While theoretically less appealing that randomized trials in terms of internal validity, natural experiments and longitudinal observational studies often have the advantage of measuring effects of more realistic intervention implementations over a larger, less contrived target population. In this sense, they may provide more generalizable effect estimates.

Given the high risk of bias due to reverse causation in chronic disease epidemiology, the CHOICES project does not use cross-sectional associations identified during systematic reviews in effectiveness models. When no evidence exists at Level 5 or higher, decision modeling becomes highly speculative, and is only carried out for interventions with a high level of stakeholder interest. In these cases, the modeling is described as exploratory and used primarily to give direction to future primary research.

	Level 1:	Meta-analyses or Systematic Reviews of <u>experimental and/or quasi-experimental</u> studies with a <u>pretest-posttest</u> and <u>control or comparison group</u>.
	Level 2:	A Single <u>Experimental</u> or <u>Quasi-Experimental</u> Study with a <u>pretest-posttest</u> and <u>control or comparison group</u>.
	Level 3:	Meta-analyses or Systematic Reviews of <u>Observational Studies</u> (with no intervention or manipulation) that analyze <u>change in the exposure</u> predicting <u>change in the outcome</u> ("change in change").
	Level 4:	A Single <u>Observational Study</u> (no intervention or manipulation) that analyzes <u>change in the exposure</u> predicting <u>change in the outcome</u> ("change in change").
	Level 5:	Meta-Analyses or Systematic Reviews with <u>less "robust" designs</u> (including cross-sectional) using <u>direct evidence</u> and found in the <u>peer-reviewed</u> literature.
	Level 6:	Single studies with a <u>less "robust" design</u> (including cross-sectional) or pilot studies with no control group results, found published in the <u>peer-reviewed</u> literature.
	Level 7:	Lowest level of evidence, including indirect evidence that is <u>not peer-reviewed</u> (<u>i.e.</u> Effectiveness Analyses, Indirect (or assumed) evidence, Parallel Evidence, Theory and Program Logic, Informed Opinion, Gray Literature)

Figure 4. CHOICES Evidence Typology. See Annex 1 for a detailed description for the 7 levels of evidence developed for the CHOICES project and definitions for all levels and various study designs.

By using common evidence ratings across interventions, the CHOICES project provides a qualitative assessment of uncertainty that supplements model-based probabilistic and deterministic sensitivity analyses. In some cases, there may be a tradeoff between level and directness of evidence. For example, for a particular intervention, experimental studies may show effect on intermediate outcomes such as energy intake, while the only direct evidence of effect on BMI is available from observational studies.

External validity assessment

Ultimately, the model is used to transfer observed effect from study populations to a national scale for the United States. The CHOICES systematic review process extracts information from studies regarding the generalizability of their findings. Although not purely quantitative, the attention to features of the study population and features of interventions' implementation (such as resource intensity and uniqueness of study sites) is systematic. Assessment of factors affecting generalizability is standardized and designed to ensure consistency across the interventions that

will be compared with modeling. Uncertainty about the generalizability of findings regarding interventions' potential population-level reach or individual-level effect are can then at least be addressed in sensitivity analysis.

5. Discussion

This case study in adapting evidence synthesis methods for a project comparing a large number of childhood obesity interventions illustrates several evidence synthesis challenges and some solutions. Methodological choices have been driven by a decision-analytic perspective that is obligated to suggest the best course of action given current information. This perspective prioritizes the inclusion of a broad information base. Augmenting traditional systematic review (and meta-analysis) with modeling enabled information from a wide range of studies to be combined along evidence pathways and translated to salient health outcomes such as BMI change, disease incidence reduction, survival gains, DALYs, and cost-effectiveness. Not only does modeling provide a solution for linking evidence from groups of studies focused on different parts of a long causal chain, it also allows for speculative assumptions based on theory or expert opinion where stronger evidence is lacking.

To a considerable extent, a unified summary of many points of uncertainty can be generated with model-based simulation. But there are uncertainties and policy-relevant factors that are difficult to capture in a model. The CHOICES study team has not found a quantitative way to combine evidence along two or more different pathways to BMI change (e.g. as shown in Figure 3). Instead, the pathway with the strongest evidence is selected as a base case, and other evidence pathways are considered in alternative scenarios. While this approach is reasonable, when the choice of evidence pathway leads to different conclusions, there is no clear way to determine how much 'weight' to put on each result. Moreover, the identification of the best pathway to serve as the base case is not necessarily straightforward since the choice is based both on the design of studies along each pathway and the directness of the pathway. A large number of studies with strong design and fewer 'steps' between the intervention and BMI change is preferable. In some cases, studies with better designs may lie along less direct pathways. The relative importance of design and directness is difficult to quantify, making the evaluation of evidence tricky.

Another area where the quantification of uncertainty is difficult relates to generalizability of reviewed evidence. In CHOICES, childhood interventions are modeled at the national and state level for the United States. For most interventions, evidence is draw from much more limited implementations. When transferring evidence of effectiveness to the U.S. population level, effort is made to adjust for setting differences. For instance, a school-based physical education (PE) intervention may have predictable variation in effectiveness attributable to state-level differences in baseline PE programming and use of specialist PE teachers. Systematic reviews rarely provide information to quantify uncertainty in the generalization to a different population.

Finally, information regarding an intervention's reach, acceptability, feasibility, sustainability, and impact on equity are often not easy to extract from studies in the course of systematic review. In CHOICES, assumptions regarding an intervention's reach are transparently reported and informed by empirical evidence where possible. But, other factors moderating the expected value of an intervention such as 'feasibility' are considered only qualitatively outside of the model simulations.

Despite the challenges of "weighting" evidence and characterizing some types of uncertainty, the CHOICES approach to evidence synthesis--combining systematic review and decision modeling, with due consideration to generalizability and implementation factors--promises a fair comparison of childhood obesity interventions using criteria and measures most salient to policy makers.

***

6. References

1. Finkelstein, E.A., et al., *The lifetime medical cost burden of overweight and obesity: implications for obesity prevention.* Obesity (Silver Spring), 2008. **16**(8): p. 1843-8.
2. Ogden, C.L., et al., *Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010.* JAMA, 2012. **307**(5): p. 483-90.
3. Committee on an Evidence Framework for Obesity Prevention Decision Making, F.a.N.B., *Bridging the evidence gap in obesity prevention: a framework to inform decision making.* 2010, Washington DC: The National Academies Press.
4. Briss, P.A., et al., *Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services.* Am J Prev Med, 2000. **18**(1 Suppl): p. 35-43.
5. Wang, Y., et al., in *Childhood Obesity Prevention Programs: Comparative Effectiveness Review and Meta-Analysis.* 2013: Rockville (MD).
6. Levy, D.T., et al., *Simulation models of obesity: a review of the literature and implications for research and policy.* Obes Rev, 2011. **12**(5): p. 378-94.
7. Haby, M.M., et al., *A new approach to assessing the health benefit from obesity interventions in children and adolescents: the assessing cost-effectiveness in obesity project.* Int J Obes (Lond), 2006. **30**(10): p. 1463-75.
8. Vos, T., et al., *Assessing cost-effectiveness in prevention (ACE-Prevention): Final Report.* 2010.
9. Carter, R., et al., *Assessing cost-effectiveness in obesity (ACE-obesity): an overview of the ACE approach, economic methods and cost results.* BMC Public Health, 2009. **9**: p. 419.
10. Juni, P., et al., *The hazards of scoring the quality of clinical trials for meta-analysis.* JAMA, 1999. **282**(11): p. 1054-60.
11. Thompson, S.G. and J.P. Higgins, *How should meta-regression analyses be undertaken and interpreted?* Stat Med, 2002. **21**(11): p. 1559-73.
12. Naci, H., et al., *Comparative benefits of statins in the primary and secondary prevention of major coronary events and all-cause mortality: a network meta-analysis of placebo-controlled and active-comparator trials.* Eur J Prev Cardiol, 2013. **20**(4): p. 641-57.
13. Green, L.W. and R.E. Glasgow, *Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology.* Eval Health Prof, 2006. **29**(1): p. 126-53.
14. Victora, C.G., J.P. Habicht, and J. Bryce, *Evidence-based public health: moving beyond randomized trials.* Am J Public Health, 2004. **94**(3): p. 400-5.
15. Sainfort, F., et al., *Adding decision models to systematic reviews: informing a framework for deciding when and how to do so.* Value Health, 2013. **16**(1): p. 133-9.
16. Brownson, R.C., J.E. Fielding, and C.M. Maylahn, *Evidence-based public health: a fundamental concept for public health practice.* Annu Rev Public Health, 2009. **30**: p. 175-201.
17. Pawson, R., et al., *Realist review--a new method of systematic review designed for complex policy interventions.* J Health Serv Res Policy, 2005. **10 Suppl 1**: p. 21-34.
18. Whitlock, E.P., et al., *Screening and interventions for childhood overweight: a summary of evidence for the US Preventive Services Task Force.* Pediatrics, 2005. **116**(1): p. e125-44.

19. Katz, D.L., et al., *Public health strategies for preventing and controlling overweight and obesity in school and worksite settings: a report on recommendations of the Task Force on Community Preventive Services.* MMWR Recomm Rep, 2005. **54**(RR-10): p. 1-12.

20. Braithwaite, R.S., M.S. Roberts, and A.C. Justice, *Incorporating quality of evidence into decision analytic modeling.* Ann Intern Med, 2007. **146**(2): p. 133-41.

21. Freedman, D.S., et al., *The relation of childhood BMI to adult adiposity: the Bogalusa Heart Study.* Pediatrics, 2005. **115**(1): p. 22-7.

22. Guo, S.S., et al., *Predicting overweight and obesity in adulthood from body mass index values in childhood and adolescence.* Am J Clin Nutr, 2002. **76**(3): p. 653-8.

23. Whitaker, R.C., et al., *Predicting obesity in young adulthood from childhood and parental obesity.* N Engl J Med, 1997. **337**(13): p. 869-73.

24. Centers for Disease Control and Prevention, *Overweight and obesity.*

25. Hall, K.D., et al., *Quantification of the effect of energy imbalance on bodyweight.* Lancet, 2011. **378**(9793): p. 826-37.

26. Lin, B.H., et al., *Measuring weight outcomes for obesity intervention strategies: the case of a sugar-sweetened beverage tax.* Econ Hum Biol, 2011. **9**(4): p. 329-41.

27. Chow, C.C. and K.D. Hall, *The dynamics of human body weight change.* PLoS Comput Biol, 2008. **4**(3): p. e1000045.

28. Hall, K.D., et al., *Dynamics of childhood growth and obesity: development and validation of a quanitative mathematical model.* The Lancet Diabetes & Endocrinology, 2013.

29. Melzer, K., et al., *Effects of physical activity on food intake.* Clin Nutr, 2005. **24**(6): p. 885-95.

30. Mozaffarian, D., et al., *Changes in diet and lifestyle and long-term weight gain in women and men.* N Engl J Med, 2011. **364**(25): p. 2392-404.

31. Sonneville, K.R. and S.L. Gortmaker, *Total energy intake, adolescent discretionary behaviors and the energy gap.* Int J Obes (Lond), 2008. **32 Suppl 6**: p. S19-27.