2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 774210

CO-CREATE

D7.2: Articles of simulation of policy effects

University of Oslo
31.01.2023

Deliverable administration and summary			
Due date	31.01.2023		
Submission date	31.01.2023		
Deliverable type	Report		
Contributors:	Name	Organisation	Role / Title
Deliverable Leader	Nanna Lien	UoO	WP Leader/Professor
Contributing Author(s)	Eduard Romanenko	UoO	Researcher
Reviewer(s)	Pål Davidsen Andrew Brown	UoB Global Centre for Preventive Health and Nutrition (Deakin University, Australia)	Prof Emeritus Research Fellow
Final review and approval			

Document change history				
Version	Release date	Reason for Change	Status (Draft/In-review/Submitted)	Distribution
V1	09.01.23		Draft	Internal reviewers
V2	26.01.23	Internal review	Submitted	

Dissemination level		
CO	Confidential, only for members of the consortium (including the Commission Services) - Appendix 2 - article confidential until published in peer-review journal	X

Executive Summary

Deliverable 7.2 is tasked to provide 1-3 articles of simulation of policy effects, including policy actions that were most commonly suggested by the youth in CO-CREATE. For this purpose, system dynamics (SD) method was employed and, in particular, an approach that builds on previous modelling work that utilized survey data to explore and simulate the impacts of intertwined social determinants on health. In line with this approach, a combination of statistical analysis and simulation modeling was used to develop a parsimonious SD model based on the data from the Health Behavior in School-aged Children (HBSC) study. The model was then used to identify the five most influential points of intervention that were compared to the policy ideas suggested by the youth. Three of the model-based priority areas were in line with the policy ideas, and two of the priority areas extended beyond the policy ideas prioritized by the youth. These modelling results are reported in an article published in Obesity Reviews in November 2022, as part of CO-CREATE supplement. A manuscript of the second article exploring the sensitivity of the results to some of the modelling decisions was submitted to System Dynamics Review on December 13 2022.

Table of content

List of acronyms / abbreviations

AdOWOB	Adolescent Overweight and Obesity
CEIDSS	Centro de Estudos e Investigação em Dinâmicas Sociais e Saúde
HBSC	Health Behaviour in School-aged Children
LSHTM	London School of Hygiene and Tropical Medicine
NIPH	Norwegian Institute of Public Health
PI	Principal Investigator
RCT	Randomized Controlled Trial
SD	System Dynamics
SES	Socioeconomic Status
SWPS	SWPS University of Social Sciences and Humanities
UCT	University of Cape Town
UK	United Kingdom
UoB	University of Bergen
UoO	University of Oslo
UvA	University of Amsterdam
WP	Work Package

Introduction

Work package 7 (WP7) has the overarching aim to evaluate the project using process, output and impact data. This aim is broken down into three objectives with corresponding tasks. This deliverable is part of Objective 7.1: *To develop an adaptable system dynamics (SD) core model (knowledge repository) for quantitative modelling of the system structure governing the development of obesity and the model-based assessment of selected policies, - both based upon state-of the art evidence (WP3) and the system maps (WP4*), and the corresponding Task 7.1. *To develop an adaptable SD core model (knowledge repository)*. This task has the following three deliverables, and involves seven of the CO-CREATE-partners (Lead: UoO/UoB; Participants: UvA, LSHTM, CEIDSS, SWPS, UCT):

D7.1: Review of existing SD models on overweight/obesity in children.

D7.2: Articles of simulation of policy effects.

D7.3: An SD model (knowledge repository).

The formulation of Objective 7.1 and Task 7.1 suggests that the initial – intuitive – strategy for WP7 was to develop one SD model, both to serve as knowledge repository and to be used for simulation of direct and indirect, short- and long-term consequences of 1-3 of the most commonly suggested co-created policies. In the process of fulfilling the objective and the corresponding task, we learnt more about various approaches to using SD in public health applications and their strengths and weaknesses depending on model purpose and available evidence and data. More specifically, we learnt that using one SD model to serve both as knowledge repository and as a tool for simulation and assessment of policy effects was not the strongest approach in our case; we could achieve a better quality of D7.2 and D7.3 by applying SD method to each deliverable differently. We, therefore, carried out the part of task 7.1 that corresponds to D7.2 by developing and using an additional – smaller – SD model, tailored specifically to simulation and analysis of policy effects. The knowledge repository model is the subject of D7.3.

The present deliverable contains 2 articles on simulation of policy effects (the first article published in Obesity Reviews in November 2022 as part of CO-CREATE supplement 1 and the second article was submitted to System Dynamics Review on December 13 2022), the description and rationale for our approach to using SD method in fulfilling the task of simulating and evaluating co-created policies, and our reflections on the chosen approach and suggestions for further work in the direction of using SD for evaluation of policies to prevent adolescent overweight and obesity.

Deliverable description

In accordance with the grant agreement:

1-3 articles of simulation of policy effects will be provided, including the most commonly suggested policy actions.

Background

CO-CREATE has taken on systems approach to confronting the rise in adolescent overweight and obesity (AdOWOB). The approach is a response to a growing recognition of the need to move beyond interventions targeting the individual level and towards comprehensive packages of policies which address the epidemic as a result of an obesity system (Roberto et al 2015, IOM 2012, WHO 2008) comprised of a wide array of interconnected factors, both individual and environmental. As part of the approach, the project engaged adolescents in systems mapping to identify various factors potentially contributing to AdOWOB (Savona et al 2021). A related process engaged the youth in developing recommendations for potential comprehensives policies to prevent AdOWOB and resulted in 4 overarching policy ideas suggested by the youth representatives from the 5 partner countries. These policy ideas are: (1) Stop all marketing of unhealthy foods to children under the age of 18 years; (2) secure high-quality practical food and nutrition education in school and a healthy school cafeteria for all children; (3) implement a sugar-sweetened beverage (SSB) tax to make unhealthy foods more expensive; and (4) offer free, organized physical activities at least once every week for all children and adolescents (Co-Create Youth Declaration 2020). This way, CO-CREATE combined both participatory (including the voice of the youth) and systems (comprehensive view of many interconnected factors at multiple levels) approaches in the process of developing policy ideas.

The present deliverable deals with the next logical step, which is to evaluate the proposed policy ideas. Assessing the effectiveness of comprehensive public health policies, however, is a well-recognized challenge, since the gold standard for effect evaluation - a randomized controlled trial (RCT) - is not readily applicable for policy evaluations (IOM 2013). Therefore, to address the task of evaluating youth-generated policy ideas, CO-CREATE chose to employ simulation modeling and, specifically, SD method.

SD method is well suited for studying complex dynamic problems, such as rising AdOWOB, that are driven by multiple factors connected through behavioral and psychological feedback loops (that is, circular chains of influence). An SD model is a dynamic hypothesis, which is comprised of hypothesized causal influences. As part of SD modeling process, a dynamic hypothesis (model) is validated in two ways: structurally (i.e., demonstrating correspondence to the relevant literature and expert knowledge) and behaviorally (i.e., being able to reproduce reasonably well historical trend data) (Sterman 2000). In SD method, model validation is a process of building confidence in a dynamic hypothesis. Once considered to be sufficiently validated, a model can serve as a useful tool for simulating and consequently assessing policies and interventions of interest (Barlas 1996).

When it comes to using SD for policy evaluation, the behavioral aspect of model validation is particularly important. Typically, dynamic models comprised of multiple feedbacks are insensitive to a wide range of their parameters, which makes it hard to force such models to replicate historical data for several variables simultaneously with realistic parameter values and meaningful formulation of the hypothesized causal links (Forrester 1961). Therefore, if a model is strong on the merits of both behavioral and structural validity (reproducing historical data “for the right reasons”), such model can be considered reliable enough for simulating policies (which is essentially running a model under conditions for which there is no data available). When it is not possible to validate a model behaviorally against multiple time series data (often due to lack of data), such model can still be very useful (including for the purpose of understanding the feedback mechanisms in question or even the response of the model variables to key assumptions about the model structure) but should be considered as more of an exploratory model (Homer 2014).

In typical situations where SD is used to study policy impacts, most of the causal influences can be established with reasonable degree of certainty prior to behavioral validation (based on the literature, prior knowledge, expert opinion or even modeler judgement). Only a few formulations and parameter values are highly uncertain and, therefore, explored further through model calibration to the available time series data (Homer 2012). The context for developing a system model of AdOWOB is different in this respect, as the population-level dynamics of AdOWOB is the result of multiple complex mechanisms operating at the level of an individual. An appropriate quantification of such mechanisms for an aggregated population-level model is challenging, because the results of focused studies (i.e., RCTs or longitudinal designs) are often context-specific and difficult to generalize. Furthermore, there is a significant diversity of evidence and opinions associated with the determinants of AdOWOB (Vandebroeck and Goossens 2007; Allender et al 2015). Therefore, the degree to which a modeler can be confident in a-priori formulations and parameter values in an SD model of AdOWOB at a population (for example, country) level is limited.

We have observed two approaches that have been used in SD practice to address the described challenge in development and validation of models that can be used for evaluation of policies targeting AdOWOB or other health conditions.

One approach is to employ participatory methods (community-based approach to SD and group model building) whereby the combined knowledge of diverse group of experts and stakeholders is used to inform the majority of the model's structure. Then, as in more "typical" SD applications, only most uncertain formulations and parameter values are tested through the calibration of the model to a few historical data series. Triangulation of various data sources is also actively used in this approach, but higher confidence in the hypothesized model structure comes mostly from extensive involvement of a well-represented group of experts in the process of model development and validation (the process of data triangulation is often facilitated through expert engagement as well). For the examples of using such an approach in evaluating policies to reduce AdOWOB, see Roberts et al 2019, Freebairn et al 2016, and Atkinson et al 2017.

Another approach focuses on using survey data to explore and simulate the impacts of intertwined social determinants on health conditions (for examples, see Mahamoud et al 2013 and Milstein and Homer 2020). Such an approach uses multifactorial statistical analysis to aid in the development of a dynamic hypothesis which could then be tested using dynamic simulation. The strength of this approach lies in its ability to perform appropriate quantification of the hypothesized causal links among various factors at a population level. In this sense, statistical analysis allows to curate many possible variables and links suggested by literature and experts and identify those that are significant and should be included in a dynamic hypothesis. In practice, the studies that followed this approach relied on cross-sectional data for one or two periods of data, which allowed estimating causal link strengths to be fed into a simulation model (these studies did not attempt to replicate historical behavior over time, and rather assumed an equilibrium baseline). In principle, however, if survey data measures several collections of individuals (for example, population segments, such as gender-socio-economic status (SES) or gender-country) at multiple points in time, the SD model can be calibrated to time series data for many, if not all, variables and across several cases (corresponding to population segments, for example). Such calibration would provide a very strong test of behavioral validation (a model is able to reproduce time series data for multiple variables and for multiple cases) and, therefore, could be considered sufficiently reliable to be used for evaluation of policy effects.

As discussed under the "Description of activities" section below, we chose to utilize the second approach for the present deliverable. Under the "Discussion/Reflections" section, we explain how the first approach can be potentially employed for further work with modeling AdOWOB.

Collaboration among partners/relation to other project activities

The deliverable assesses policy effects, including those that were most commonly suggested by the youth in CO-CREATE, as part of WP5 work.

The results of systems mapping with adolescents in WP4 were used to inform a part of structural validation of the SD model (whether the model reflects the views of adolescents as key stakeholders).

Anne-Siri Fismen (NIPH/WP3) and Professor Harry Rutter (University of Bath, UK/PI) contributed to article 1 as co-authors.

Description of activities

Work with the knowledge repository model: September 2020 – June 2021
In accordance with the formulation of Objective 7.1 and Task 7.1, we began with exploring how an adaptable SD model (knowledge repository model), which had already been under development, could be used for the task of policy evaluation. At that stage, model still required quantification for many of its parts. As the process of model quantification was moving along, we monitored what the model would need to be able to serve as a useful tool for simulation of policy ideas. Throughout this process and, in particular, during a series of three workshops, where various sub-models of the model were presented to and discussed with the experts (February-May 2021), we learnt that the breadth and scope of the model (physical and food environment, individual-level mechanisms through mental health) could not be matched with the available data or numerical estimates from the literature (evidence) and, therefore, left many of the model formulations uncertain. Additionally, and again due to the scope of the model relative to the available data, we could not include socio-economic inequality and country-specific dimension, both of which are of interest to CO-CREATE. Incorporating the dimension of socio-economic inequality is particularly important for the task of

policy evaluation, since the literature emphasizes that some interventions might be less effective for lower income groups.

Work with the HBSC data: July-September 2021

In our work on quantifying the knowledge repository model, we identified the dataset from the Health Behavior in School-aged Children (HBSC) study as the only available dataset that reports a broad array of relevant variables consistently across many European countries and over multiple time periods (only the data from survey years 2002, 2006, 2010 and 2014 were open access at the time; the 2018 data became available end of October 2022 and was not included in the analysis), including CO-CREATE partner countries.

The availability of such data opened a possibility to employ the approach focusing on using survey data to explore and simulate the impacts of intertwined social determinants on health conditions (the second of the two approaches described in the Background section). Moreover, the availability of data for four survey periods opened a possibility to rigorously validate a model against the time series specific to country, gender, and perceived family wealth (a marker of socioeconomic status) – all three dimensions of interest for CO-CREATE.

We, therefore, decided to carry out the task of evaluating policy ideas by developing a new, parsimonious model (smaller "policy" model) comprised of the variables that could be directly related to the variables in the HBSC dataset. In fact, this decision allowed us to avoid compromising between the two parts of task 7.1: (1) *quantitative modelling of the system structure governing the development of obesity* and *(2) the model-based assessment of selected policies*. By delegating (1) to the initial knowledge repository model enabled us to keep all of the knowledge about various mechanisms governing AdOWOB system which had been gained within CO-CREATE, instead of "trimming" the model down for the purpose of complying with requirements of the policy evaluation task. In this sense, the initial model could truly serve as a knowledge repository and the smaller "policy" model could serve as a reliable tool for policy assessment.

We then proceeded with the analysis of the HBSC data, which included selection and formulation (dichotomization) of potential variables, analysis of probabilistic odds ratios for potential explanatory variables of AdOWOB, analysis of correlations between all the variables, and stepwise multivariate linear regressions for all endogenous variables (that is, generated within the model rather than used as model inputs) to be considered for a dynamic hypothesis.

Developing an SD model for policy evaluation (smaller "policy" model): October 2021 – February 2022

The statistical analysis of the HBSC data resulted in a dynamic hypothesis, which was then converted into a simulation SD model. Since the model was informed by the data analysis across 31 European countries, it could be considered as a generalized European model. This generalized model was then tested by using automated calibration to the HBSC data for 24 different cases based on 4 gender and perceived wealth segments for each of the 5 CO-CREATE countries (The Netherlands, Norway, Poland, Portugal, and the UK) and for Europe overall. Our dynamic hypothesis demonstrated high explanatory power across the tested cases, which meant that it could be used for further policy analysis. For each case, we tested 10 factors of AdOWOB in the model as potential points of intervention and ranked them by projected reduction of AdOWOB by 2026. We used our model-based findings to support or supplement the policies suggested by the adolescent participants who were part of CO-CREATE.

Development of Article 1: March – June 2022

Once the policy analysis was completed, we moved to the stage of drafting the article. Eduard Romanenko (UoO/WP7), Jack Homer (Homer Consulting and MIT Research Affiliate, USA; Eduard Romanenko's mentor within the SD Society's mentorship program) and Professor Nanna Lien (UoO/WP7) developed the first draft of the article. Anne-Siri Fismen (NIPH/WP3) and Professor Harry Rutter (University of Bath, UK/PI) were invited to contribute to the article; they reviewed the first draft and introduced important changes that ensured an appropriate description of the use and interpretation of the HBSC data and the presentation of SD method to the public health readership. The article was planned to be submitted to Obesity Reviews, as part of the first CO-CREATE supplement. The final version of the manuscript "Assessing policies to reduce adolescent overweight and obesity: Insights from a system dynamics model using data from the Health Behavior in School-aged Children study" was submitted to the journal on June 15, 2022. The revised version (minor revisions) was submitted on August 10, 2022 and accepted for publication on October 12, 2022.

Development of Article 2: June – December 2022

Our approach to developing an SD model for policy evaluation in CO-CREATE had implications for which variables and causal links we did or did not include in the model. Expert reviewers generally agreed with our modeling decisions, but two decisions did raise questions: (1) excluding the influences of food environment and built environment, for which we had no data; and (2) including five causal links (from School Pressure and Feel Nervous directly to AdOWOB and three links capturing the effect of environmental variables on fruit and vegetables consumption) that were supported statistically but might be considered disputable (since they required implicit intermediate variables, for which there was no data). To address the reviewers' questions, we created four possible model structures and performed automated calibration with them followed by intervention testing and ranking. We then compared the goodness of fit and intervention results.

Based on these results, we developed the draft of the second article entitled "As Simple as Possible but not Simpler: structural sensitivity testing of a dynamic model of adolescent overweight and obesity" (authors: Eduard Romanenko, Jack Homer and Professor Nanna Lien). In this article we used the analysis of the alternative model structures to discuss implications for how to move forward with the AdOWOB model, including through additional data gathering. The targeted journal for article 2 is System Dynamics Review, since the type of analysis that we performed for this article formalizes structural sensitivity testing and, thereby, provides an important contribution to SD modeling practice. The results included in article 2 provide a formal way to support further the modeling approach we chose for the policy assessment in CO-CREATE. The article was submitted to System Dynamics Review on December 13 2022.

Results

The journal article entitled "Assessing policies to reduce adolescent overweight and obesity: Insights from a system dynamics model using data from the Health Behavior in School-aged Children study" is attached (Appendix 1: D7.2 Articles of simulation of policy effects). The article was published in November 2022 in Obesity Reviews (early view; to be included in the CO-CREATE supplement).

The manuscript of journal article entitled "As Simple as Possible but not Simpler: structural sensitivity testing of a dynamic model of adolescent overweight and obesity" is attached (Appendix 2: D7.2 Articles of simulation of policy effects). The article was submitted to System Dynamics Review on December 13 2022.

Discussion/reflections

In our work, we utilized a combination of literature review, statistical screening procedures, and SD modelling to build a strongly evidence-based model with only 12 major variables (8 of them endogenous and 4 exogenous) and 30 causal links (with corresponding strengths known as hazard ratios). The approach that we chose allowed us to utilize fully the available HBSC dataset on a wide array of health behaviours and health determinants to support the task of simulation of policy effects. The approach also allowed us to include socio-economic inequality and country-specific dimension, both of which are of interest to CO-CREATE.

Our analysis identified five intervention points as most impactful across the studied county-gender-perceived well-off cases: exercise, fruit, life dissatisfaction, school pressure, and skipping breakfast. Three of these priority areas (exercise, fruit, and skipping breakfast) correspond to the four policy ideas suggested by adolescents themselves in the CO-CREATE project (CO-CREATE Youth Declaration 2020), as those are all related to either nutrition or physical activity (specifically: 1) marketing of unhealthy foods; (2) nutrition education in school and healthy school cafeteria; (3) SSB tax; and (4) free organized physical activities). Two of our top intervention priorities (reducing life dissatisfaction and school pressure) were not prioritized by the CO-CREATE Youth Task Force, but correspond to mental health and social factors, the importance of which have been raised by the adolescents during systems mapping (Savona et al 2021).

Our experience of applying SD method to simulating policy effects that target AdOWOB demonstrates importance of access to high quality data, measuring various health behaviors and health determinants and spanning multitude of countries and multiple time periods. Our work also demonstrated that a useful parsimonious model, which avoids speculative relationships, can be developed (as described in Appendix 1), in spite of the challenges of quantifying uncertain cross impacts of multiple determinants of AdOWOB at a population level, and then used for robust policy analysis.

At least two useful directions for those wishing to build upon our experience could be suggested. Progressing in either or both of these directions may result in a model with a more expanded boundary and a higher level of detail, which among other things, may include appropriate entry points for simulating more specific types of policies (such as, marketing, education and SSB tax, as suggested by the adolescents in CO-CREATE).

First, gathering more evidence (data) is important for expanding the boundary of the model beyond what could be supported by one dataset alone. This task may require triangulating evidence from multiple datasets, which may not be directly compatible with each other. In our second article, we (1) identified which of the “disputable” links and “missing” variables our AdOWOB model appeared to be sensitive to and (2) provided examples of potential sources of evidence for such variables and links.

Second, and related to the first direction, once a validated model based on the best available data is developed, such a model can be used as a starting point for further modeling work together with the diverse group of AdOWOB experts (as in the first of the two approaches described in the Background section). The experts may be able to assess whether reasonable formulations for the concepts for which less or little evidence is available can be incorporated into the model. It is important, however, that the experts understand well the logic of the model and how changes in model formulations affect the resulting dynamics of its key variables. Achieving such understanding is an absolute pre-requisite for a successful group-model building process, but usually requires substantial commitment in terms of time and motivation (see, for example, Freebairn et al 2019), the resources for which need to be budgeted into a modeling project.

Conclusion/recommendations

Our work demonstrated that SD modelling and simulation is well-suited for assessing policies to reduce AdOWOB. We followed an approach that suited the needs of the project and capitalized on the available data and evidence. Building on the previous modelling work that utilized survey data to explore and simulate the impacts of intertwined social determinants on health, we developed a parsimonious model that avoided speculative formulations and could be appropriately validated, in line with the best SD practices. We then used the model to identify five most influential points for intervention, three of which were in line with the exercise and nutrition-related policy ideas prioritized by the youth in CO-CREATE, and two of which extended beyond those. We then explored the sensitivity of our policy conclusions to our more disputable modelling decisions and identified those variables and causal links for which more evidence needs to be gathered in the future. Article 1 reports on the baseline model and the associated policy analysis. Article 2 reports the analysis of alternative model structures and discusses the implications for future data needs. For the modelling teams involved in future assessment of policies targeting AdOWOB, we recommend beginning with identifying and utilizing high quality data sources and, if possible, expanding the boundary by gathering and incorporating more evidence and engaging a diverse group of experts.

References

Allender S, Owen B, Kuhlberg J et al. A community based systems diagram of obesity causes. *PLoS One*. 2015; 10(7): e0129683.

Atkinson JA, O'Donnell E, Wiggers J, McDonnell G, Mitchell J, Freebairn L, Indig D, Rychetnik L. Dynamic simulation modelling of policy responses to reduce alcohol-related harms: rationale and procedure for a participatory approach. *Public Health Res. Pract*. 2017; 27(1).

Barlas Y. Formal aspects of model validity and validation in system dynamics. *Syst Dyn Rev*. 1996; 12(3):183-210.

CO-CREATE Youth Declaration: Time to Act and Ensure Good Health for All. November 2020. https://www.fhi.no/contentassets/0a74196d35c64da89d337e25af982f5f/co-create-youth-declaration-on-ending-childhood-and-adolescent-obesity.pdf

Forrester JW. *Industrial Dynamics*. Cambridge, MA: MIT Press. 1961.

Freebairn L, Atkinson J, Kelly P, McDonnell G, Rychetnik L. Simulation modelling as a tool for knowledge mobilisation in health policy settings: a case study protocol. *Health Res. Policy Syst*. 2016. 14(1).

Freebairn L, Atkinson J-A, Osgood ND, Kelly PM, McDonnell G, Rychetnik L. Turning conceptual systems maps into dynamic simulation models: An Australian case study for diabetes in pregnancy. *PLoS One*. 2019; 14(6): e0218875.

Homer JB. Partial-model testing as a validation tool for system dynamics (1983). *Syst Dyn Rev*. 2012; 28(3):281-94.

Homer J. Levels of evidence in system dynamics modeling. *Syst Dyn Rev*. *2014;* 30:75-80.

Institute of Medicine 2012. Accelerating progress in obesity prevention: Solving the weight of the nation. Washington, DC: The National Academies Press.

Institute of Medicine 2013. Evaluating obesity prevention efforts: A plan for measuring progress. Washington, DC: The National Academies Press.

Mahamoud A, Roche B, Homer J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. *Soc Sci Med*. 2013; 93:247-255.

Milstein B, Homer J. Which priorities for health and well-being stand out after accounting for tangled threats and costs? Simulating potential intervention portfolios in large urban counties. *Milbank Q*. 2020; 98(2):372-398.

Roberto CA et al. Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking. Lancet 2015; 385: 2400-2409.

Roberts N, Li V, Atkinson JA, Heffernan M, McDonnell G, Prodan A, Freebairn L, Lloyd B, Nieuwenhuizen S, Mitchell J, Lung T. Can the target set for reducing childhood overweight and obesity be met? A system dynamics modelling study in New South Wales, Australia. *Syst Res Behav Sci*. 2019; 36(1):36-52.

Savona N, Macauley T, Aguiar A, et al. Identifying the views of adolescents in five European countries on the drivers of obesity using group model building. *Eur J Public Health*. 2021; 31(2):391-396.

Sterman JD. *Business dynamics: systems thinking and modeling for a complex world*. Boston, MA: Irwin McGraw-Hill. 2020 (Ch. 21, "Truth and beauty: validation and model testing", pp. 845-891.)

Vandenbroeck IP, & Goossens JMC. Foresight tackling obesities: future choices—building the obesity system map. Government Office for Science, UK Government's Foresight Programme. 2007; Retrieved 2 December 2022 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/295154/07-1179-obesity-building-system-map.pdf

World Health Organization. WHO global strategy on diet, physical activity and health: a framework to monitor and evaluate implementation. 2008.

Appendix 1

The journal article entitled "Assessing policies to reduce adolescent overweight and obesity: Insights from a system dynamics model using data from the Health Behavior in School-aged Children study" (published in Obesity Reviews on November 22 2022) is attached below.

Received: 15 June 2022 | Revised: 10 August 2022 | Accepted: 12 October 2022
DOI: 10.1111/obr.13519

SUPPLEMENT ARTICLE

OBESITY Reviews WILEY

Assessing policies to reduce adolescent overweight and obesity: Insights from a system dynamics model using data from the Health Behavior in School-Aged Children study

Eduard Romanenko[1] | Jack Homer[2] | Anne-Siri Fismen[3,4] | Harry Rutter[5] | Nanna Lien[1]

[1]Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
[2]Homer Consulting and MIT Research Affiliate, Barrytown, New York, USA
[3]Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway
[4]Center for Evaluation of Public Health Measures, Norwegian Institute of Public Health, Bergen, Norway
[5]Department of Social and Policy Sciences, University of Bath, Bath, UK

Correspondence
Eduard Romanenko, Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, PO Box 1046, Blindern, N-0316, Oslo, Norway.
Email: eduard.romanenko@medisin.uio.no

Funding information
The CO-CREATE project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 774210. The content of this article reflects only the authors' views, and the European Commission is not liable for any use that may be made of the information it contains.

Summary

Adolescent overweight and obesity (AdOWOB) in Europe has proven to be a persistent and complex problem, and appropriate systems methods may help in evaluating potential policy options. This paper describes the development of a system dynamics model of AdOWOB as part of the EU-funded CO-CREATE project. The model was developed using literature and data from the Health Behavior in School-Aged Children (HBSC) study across 31 European countries. We identified 10 HBSC variables that were included as direct or indirect drivers of AdOWOB in the dynamic model, seven at the level of the individual, and three related to the social environment. The model was calibrated to 24 separate cases based on four gender and perceived wealth segments for each of the five CO-CREATE countries (The Netherlands, Norway, Poland, Portugal, and the UK) and for Europe overall. Out of 10 possible intervention points tested, exercise, fruit, life dissatisfaction, school pressure, and skipping breakfast were identified as the top five most influential ones across the 24 cases. These model-based priorities can be compared with the policy ideas suggested by the CO-CREATE adolescents.

KEYWORDS
HBSC, obesity prevention, quantitative modeling, system dynamics, youth

1 | INTRODUCTION

Obesity in adults is associated with an increased risk of serious health conditions, such as Type 2 diabetes, cardiovascular disease, and several cancers, and is a leading risk factor for death in high-income and some middle-income countries.[1] Adolescents living with overweight or obesity may experience adverse physical and mental health effects and are also at increased risk for adult obesity.[2–4] In Europe, one in seven young people aged 15 years lives with

Abbreviations: AdOWOB, adolescent overweight and obesity; AdOWOBY, AdOWOB of Youngest; BMI, body mass index; EU, European Union; HBSC, Health Behavior in School-Aged Children; HR, hazard ratio; LWOB, less well-off boys; LWOG, less well-off girls; MAPE, mean absolute percentage error; MWOB, more well-off boys; MWOG, more well-off girls; NL, The Netherlands; PA, physical activity; SD, system dynamics; SSB, sugar-sweetened beverage; WHO, World Health Organization.

overweight or obesity,[5] a fraction that is projected to increase to one in five by 2025.[6]

The Health Behavior in School-Aged Children (HBSC), a World Health Organization (WHO) collaborative cross-national survey, collects self-reported data on health, health behaviors, and social determinants of health from nationally representative samples of adolescents in Europe and North America. HBSC data suggest that the prevalence of adolescent obesity across Europe increased gradually from 2002 to 2014, with boys showing consistently greater prevalence than girls, and adolescents from less affluent families with greater prevalence than those from more affluent.[7] These patterns are consistent with other European data using objective measurements of weight and height.[8] HBSC data from 2018 indicate that the prevalence of adolescent overweight and obesity (AdOWOB) continued to increase in about a third of the survey countries and declined only for some gender–age groups in a small number of countries relative to 2014.[5] Reversing the rising trend in childhood and adolescent obesity has been declared by the WHO and the European Union (EU) as an important public health priority in key documents.[9–13] The global policy target set by the WHO is to halt the increase in obesity prevalence by 2025.[10]

Many government responses to AdOWOB prevalence have emphasized health education programs aiming to influence individual's eating and physical activity (PA) choices, the two behavioral factors most directly determining energy balance and weight change.[14,15] Examples of such interventions are the Change for Life campaign in the UK[16] and the AMEA TEENS program in Portugal.[17] The persistence of high AdOWOB prevalence, however, has led to a growing recognition of the importance of social, physical, and economic environments in shaping an individual's diet and activity behaviors, and thus health outcomes including mental health.[18–20] Some have suggested that the role of environmental factors is particularly salient in adolescents who, compared with adults, have lower levels of behavioral autonomy and for lower income groups who face greater barriers to adopting healthy behaviors.[21]

To identify potentially effective policy interventions among a wider array of interconnected factors, both individual and environmental, an EU-funded project, "Confronting obesity: Co-creating policy with youth" (CO-CREATE),[22] has engaged adolescents in systems mapping as part of a process of developing recommendations for potential policies to reduce AdOWOB. In November 2020, a task force consisting of youth from the five participating countries (The Netherlands, Norway, Poland, Portugal, and the UK) adopted a declaration that listed four policy ideas: (1) Stop all marketing of unhealthy foods to children under the age of 18 years; (2) secure high-quality practical food and nutrition education in school and a healthy school cafeteria for all children; (3) implement a sugar-sweetened beverage (SSB) tax to make unhealthy foods more expensive; and (4) offer free, organized physical activities at least once every week for all children and adolescents.[23]

Here, we report on an effort within CO-CREATE to evaluate a variety of intervention points, including those related to the policy ideas suggested by the adolescents, using system dynamics (SD) modeling and simulation. The SD method is well suited for studying complex dynamic problems, such as rising AdOWOB, that are driven by multiple factors connected both directly and indirectly and through behavioral and psychological feedback loops.

In SD modeling, a model of hypothesized causal influences (dynamic hypothesis) is developed and then validated structurally (e.g., demonstrating correspondence to available literature and subject matter expert knowledge) and behaviorally (e.g., closely reproducing historical trend data).[24] For our SD model of AdOWOB, structural validation has also meant ensuring that the model reflects the views of the CO-CREATE adolescents themselves, whose views are reported elsewhere.[25] The steps of structural and behavioral validation help build confidence in a model as a useful tool for the assessment of policies and interventions.[26]

No previously validated dynamic model of AdOWOB has been published that considers the wide range of behavioral, psychological, and social issues described in the AdOWOB literature.[27,28] Socioecological models of overweight and obesity have been proposed[29,30] but never before quantified or tested. Our goal was to develop a parsimonious model that could be rigorously validated against the time series data, in line with SD best practices.[31] The AdOWOB literature includes so many possible variables and links that some way was needed to sort through them and identify which were significant. To do so, our study builds on previous work that utilized multifactorial data analysis and simulated the impacts of interconnected social determinants on health conditions at a population level.[32,33] Multifactorial data analysis is important for appropriate quantification in the case of AdOWOB, because the results of focused studies (i.e., randomized control trials or longitudinal designs) are often context-specific and difficult to generalize.

Informed by the literature, we considered more than 20 potential drivers of AdOWOB from the four rounds of the HBSC survey spanning the period 2002–2014. In line with the previous SD work on social determinants of health that utilized survey data, each of these variables was expressed as a population prevalence fraction.[32,33] These drivers included adverse behaviors (e.g., inadequate exercise), psychological states (e.g., nervousness), and social determinants (e.g., school pressure) that can lead to AdOWOB, either directly or through other such variables, based on plausible causal mechanisms. The hypothesized causal links in many cases subsume implicit intermediate variables (e.g., caloric intake) not identified in the HBSC survey nor, therefore, in the dynamic model. In other words, the dynamic model collapses many real-world mechanisms into a smaller number for the sake of parsimony with respect to available data. Although the HBSC survey does not contain all the possible variables that might be used for modeling AdOWOB, it is the only dataset that reports a broad array of relevant variables consistently across many European countries. This breadth and consistency of the HBSC data allowed us to validate a generalized model against 24 different cases that vary by country, gender, and perceived family wealth (a marker of socioeconomic status), three dimensions of interest to the CO-CREATE project. In particular, we tested the potential reduction in AdOWOB that might be achieved through intervention at 10 different points in the modeled system.

2 | METHODS

2.1 | Study design, locations, and data sources

The study employed a combination of statistical analysis and SD modeling. The aim of the statistical analysis was to explore the regularities in the associations between AdOWOB and various health behaviors for adolescents across 31 countries and over time. This exploratory data analysis allowed us to curate the plausible causal influences suggested by the literature[25,34–39] and formulate a dynamic hypothesis (model). This model provides a multivariate causal explanation of the trajectories of AdOWOB for 24 cases, defined by dividing each of the five CO-CREATE countries and Europe overall (the weighted average of 31 countries) into four population segments related to gender and perceived family wealth. By calibrating to 24 different cases, we followed the SD tradition of gaining confidence in a model through "family member" analysis, in which one tests the model's ability to reproduce the behavior of multiple instances of the same system.[24] Figure 1 summarizes the steps involved in the statistical analysis and SD modeling (Figure S1 provides further details about each step of the analysis).

The data on body mass index (BMI) and health behaviors come from the HBSC survey and cover the period from 2002 to 2014, with the survey conducted every fourth year. These data are open access and were obtained from the HBSC Data Management Centre.[40] Only the WHO European region countries participating in all the four survey years were included in our statistical analysis. The 31-country HBSC dataset provides a large sample size (around 30 thousand observations per segment for each survey year) for exploring statistical regularities. Though typically used for cross-sectional analysis, HBSC data have also been used for trend analysis.[7]

2.2 | Potential variables

Our statistical analysis explored 24 variables from the HBSC dataset, including BMI, gender, perceived wealth, and 21 potential explanatory factors of adolescence as guided by the public health literature.[25,34–39] BMI was based on self-reported weight and height and used to calculate the prevalence of AdOWOB based on the international standardized age- and gender-specific cut-off points proposed by Cole et al for the International Obesity Task Force.[41] The 21 factors of adolescence included the following:

1. eating habits (fruit, vegetable, and SSB consumption; skipping breakfast, and dieting);
2. PA and sedentary behavior (moderate-to-vigorous PA, vigorous exercise, watching television, and computer use);
3. substance use (beer consumption and smoking) and other risk behaviors (being bullied);
4. other health-related conditions (feeling low, feeling nervous, self-rated health, difficulty in sleeping, body image, and life satisfaction); and
5. social context at the level of family (communication with mother or father), peers (perceived peer support), and school (school pressure).

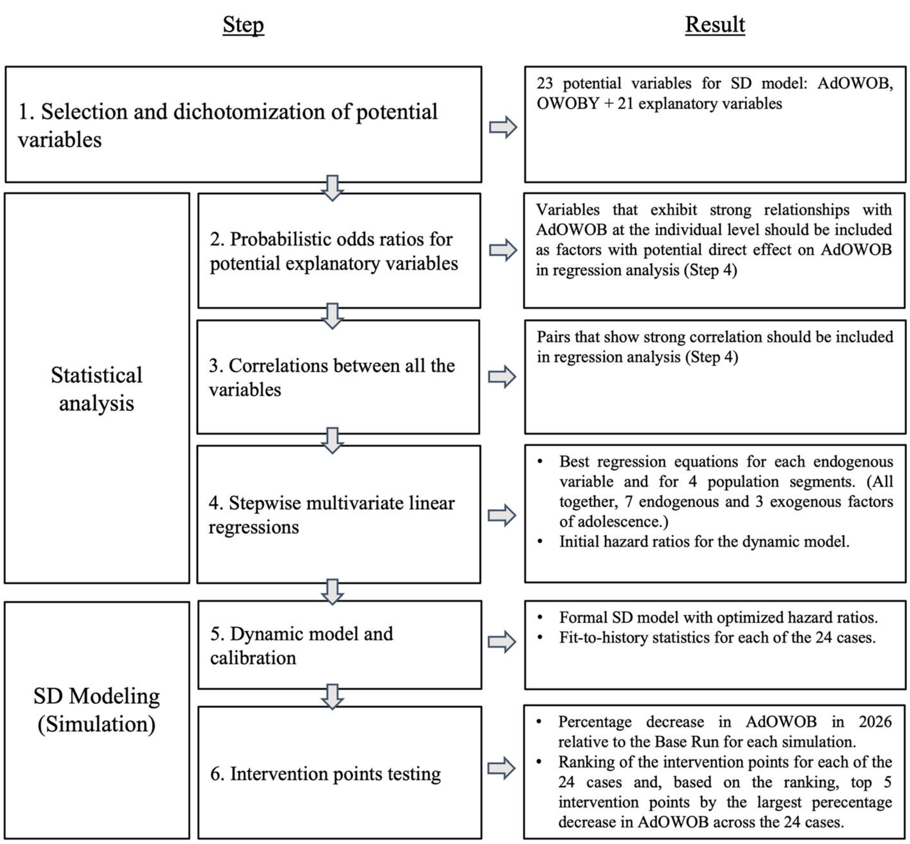

FIGURE 1 Steps of analysis

Only those variables that were asked and phrased consistently in all the four survey years were included, with the exception of the vigorous exercise variable, which was not asked in the first survey year (2002). Each of the 21 potential explanatory variables, originally recorded as categorical, was dichotomized to represent a prevalence fraction of an adverse condition. We dichotomized in order to make the analysis tractable and recognize inherent nonlinearities (adverse vs. positive or normal) in the categorical data. Our cut-offs were chosen (a) for nutritional and PA variables, in recognition of minimally adequate amounts, and (b) for all other variables where the categories (based on the wording) changed from normal to adverse. (Table S1 reports the cut-offs that we used together with those used by the HBSC team; our adjusted cut-offs give improved discriminatory power for the purposes of dynamic modeling.) For example, the original HBSC variable "Eat fruit" was recorded using seven categories, from "never" to "more than once daily". Our variable "Inadequate Fruit" represents the fraction of those who consume fruit less than 5 days a week (the first four of the seven categories). The prevalence fractions were calculated as population-level fractions for two perceived wealth categories (less and more well-off) for each gender (boys and girls) and for each of the 31 countries. The perceived wealth categories themselves were formulated by dichotomizing the family well-off variable (question: "How well off do you think your family is?", with the answers recorded using five categories, from "very well off" to "not at all well off").

2.3 | Statistical analysis

The statistical analysis was carried out in three screening steps (see Figure 1), always keeping separate the four gender-perceived wealth segments.

First, for each of the 21 potential explanatory factors, we calculated probabilistic odds ratios of having AdOWOB, for example, the probability of having AdOWOB if eating inadequate fruit compared with the probability of having AdOWOB if eating adequate fruit. These odds ratios were calculated at the level of individual subjects across the 31 countries and for each survey year. This screening step allowed us to identify strong relationships at the individual level that might otherwise become disguised by aggregation.

Second, for each survey year, we calculated correlations between AdOWOB and the 21 potential explanatory factors across the 31 countries. This step allowed us to investigate the relations of the factors not only to AdOWOB but to each other. When considering influences on AdOWOB, we used the next survey year value for AdOWOB (or "AdOWOB-Next") in the correlation matrices. With AdOWOB-Next, we were able to represent the delayed effect of influences on the gradually changing stock of AdOWOB, reflecting the time required to move from one BMI category to another.[42] All other influences were considered to occur within the current survey period, that is, over a period of 1 year or less. Our use of AdOWOB-Next allowed us to statistically "break" all feedback loops going through the stock of AdOWOB and thereby avoid a problem of estimation in the case of bidirectional influences.

The first two screening procedures informed the third step, where we performed a series of stepwise multivariate linear regressions across the 31 countries and across the survey years 2006, 2010, and 2014, performed separately for each of the four gender-perceived wealth segments. For each endogenous variable, we began with a full set of plausible independent variables suggested by the literature. In some cases, we initially included independent variables even if they had not shown a strong correlation in the first two screening steps, based on strength of support from the literature. We used AdOWOB-Next when regressing for the effect of factors of adolescence on AdOWOB and used AdOWOB (the current survey year value) when regressing for the effects on endogenous drivers. Using stepwise regression, we eliminated factors with regression polarities not supported by the literature (e.g., minus rather than plus) or with overly large P-values (greater than 0.2) indicating low statistical significance. The accepted regression equations were the ones that maximized the adjusted R-squared.[33,43]

The resulting dynamic hypothesis consisted of all the influences inferred from the accepted regressions for any of the four gender-perceived wealth segments. These included 10 factors (hereafter, "factors of adolescence"), seven of them endogenous (affected by one another or by AdOWOB) and three exogenous. Taken together, some of the endogenous influences (and AdOWOB) formed reinforcing feedback loops. The regressions were also used to provide an initial quantification of the impacts between the variables included in the model. To do so, we algebraically derived hazard ratios (HRs) corresponding to the estimated regression coefficients. With HRs, one may express the impacts of multiple factors on a dependent variable as the product of influences as affected by changes in prevalence fractions for the independent variables (as done by Milstein and Homer[33]). The HRs varied by the four gender-perceived wealth segments.

In addition to the 10 regression-based factors of adolescence, we included in our dynamic hypothesis preadolescent OWOB called "AdOWOB Youngest" (AdOWOBY), which we calculated as OWOB prevalence of the youngest part of the HBSC sample (aged 11.6 or younger). The impact of AdOWOBY was formulated through a well-defined flow into the stock of AdOWOB, with a diluting time constant that corresponds to the period of adolescence surveyed by the HBSC; this formulation, therefore, did not require an HR.

2.4 | Simulation modeling

Our statistical procedure produced a dynamic hypothesis that could be converted into a simulation model. Simulation is critical, as it is needed to test the dynamic hypothesis and ensure that it is capable of reproducing historical trends and producing plausible futures. Only simulation modeling can provide a proper dynamic test of the dynamic hypothesis.

The key uncertain parameters in our model are the HRs described above. The model contains 30 such HRs, and the regressions suggested different HR values for the four gender-perceived wealth segments. These regression-based values gave us promising starting points for an estimation process that involved automated model calibration for each of the 24 cases described above. This automated calibration was performed using Powell optimization (as implemented in the Stella Architect software for SD simulation, with automated weighting based on absolute error terms for each endogenous variable). As a result of optimization, some of the HRs were estimated as having values equal to or very close to 1.0, thereby effectively eliminating that influence for the case in question. Thus, for any particular case, the optimizer reduced the model's complexity to some extent relative to the initial dynamic hypothesis. Using the optimized HRs, we simulated a base run for each country-segment case from 2002 to 2026 in one-eighth year time increments. We assumed the values of the exogenous variables were unchanging after 2014, the last available HBSC data point.

In the optimization settings, we specified the maximum number of simulations at 200,000 ("Opt200k"). As a matter of sensitivity analysis, we repeated the optimization specifying the maximum number of simulations at 50,000 ("Opt50k"). Here, we primarily report the results of Opt200k, with some Opt50k results shown graphically to demonstrate the model's insensitivity to parameter uncertainty.

We calculated two types of goodness-of-fit statistics for the 24 cases, for all eight of the model's endogenous variables. The first of these statistics is the mean absolute percentage error (MAPE) between simulated output and data, a well-known measure that indicates how well the model replicates the general magnitude of the data.[24] The second statistic is a customized R-squared measure (range 0 to 1) of how well the model predicts changes (variance) away from the initial data point in 2002; we call this statistic "R^2i." We have found that these two statistics together give a more accurate sense of goodness-of-fit than either one of them alone.

2.5 | Interventions points

The literature suggested that all 10 of the model's factors of adolescence were plausible points of direct intervention.[44–48] In order to facilitate the comparison of intervention results, we applied an effect size of 25% for each intervention starting in 2018. In particular, an intervention was assumed to produce a 1-year ramp-wise reduction in the prevalence fraction of the target variable by 25% from 2018 to 2019 (after which feedback loops might lead to further changes in that variable if it is endogenous). Literature-based estimates of effect size, often expressed in terms of continuous individual-level impacts (e.g., grams increase in daily fruit consumption), are notoriously difficult to translate into the terms needed for a dynamic model dealing with population prevalence fractions.[31,32] Instead, our choice of the same 25% effect size for all 10 intervention points was guided by examination of historical variations (the ratio of minimum to maximum value) across all 24 country-segment cases in the HBSC data. We simulated the model by subjecting each variable to the 25% reduction separately, as well as performing a test combining all 10 intervention points. For each test, we calculated the percentage change in AdOWOB in 2026 relative to the base run.

3 | RESULTS

3.1 | Statistical analysis and dynamic hypothesis

Our statistical analysis led to a dynamic hypothesis explaining changes in AdOWOB as being driven by OWOBY (an exogenous factor described above) plus 10 factors of adolescence, seven of them endogenous and three exogenous. Table 1 lists the model's endogenous, exogenous, and excluded variables.

Variables were excluded either because the data reflect a true (real life) lack of significance, or in some cases perhaps because of how the variable is defined in the HBSC survey. An example of the latter is SSB consumption, which several studies have found to be a risk factor for OWOB,[49] but which our statistical analysis did not reveal to be even moderately associated with AdOWOB. The reason could be one of the definitions in the HBSC survey. The HBSC survey team themselves recognize that the phrasing of the question asking about SSB consumption limits the variable to mostly soda rather than the full variety of SSBs (e.g., juices from concentrate and sweetened milk drinks) that are popular among adolescents.[7]

The seven endogenous factors shown in Table 1 include four related to nutrition (fruits, vegetables, dieting, and skipping breakfast), one related to exercise, and two related to mental health (feeling nervous and feeling low). The three exogenous factors are more reflective of the social environment surrounding the individual: school pressure, excess computer (and smartphone) use, and life dissatisfaction. The overuse of computers and smartphones by adolescents, although traditionally used as a measure of sedentary behavior, also

TABLE 1 HBSC variables included in model or excluded based on statistical analysis; 31 countries × 4 gender-perceived wealth segments, 2002–2014

Endogenous (8)	Exogenous (4)	Excluded (10)
AdOWOB/BMI, Inad Fruit, Inad Vegetables, Dieting, Inad Breakfast, Inad Exercise/PA[a], Feel Nervous, Feel Low	School Pressure, Computer Overuse, Life Dissatisfaction, AdOWOBY	SSBs, Beer, Smoking, Sleep Difficulty, Been Bullied, Excess TV, Body Image, Inad Family Support, Inad Friends Support, Self-Rated Health

Note: A variable was included in the model if it showed promising correlations and proved significant in regressions yielding a causal pathway leading to AdOWOB, for any of the four gender-perceived wealth segments. Inad, inadequate.

[a]Vigorous exercise (h/week) question introduced in 2006 and dominates PA (days/week) statistically, but PA 2002–06 ratio useful for synthetic estimation of Inad Exercise 2002.

reflects the larger social trend toward the use of the internet and social media (the HBSC question "How many hours a day, in your free time, do you usually spend using electronic devices …" has broadened over the analyzed period to incorporate a greater variety of new online activities, including chatting and tweeting). Life dissatisfaction is often related to family circumstances, more often affecting adolescents from lower-income households.[50]

Figure 2 portrays a complete set of the influences inferred from the HBSC data analysis. Table S2 provides an explanation of potential causal mechanisms for each of the hypothesized links from Figure 2 and documents the supporting literature. Six factors of adolescence influenced AdOWOB directly: inadequate fruits, vegetables, exercise, and breakfast; feeling nervous; and school pressure. The remaining factors influenced AdOWOB indirectly: dieting affecting breakfast; feeling low affecting exercise and a decision to diet; computer overuse (social media in particular) affecting dieting; and life dissatisfaction affecting fruits, vegetables, exercise, and feeling nervous. School pressure also had indirect influences on AdOWOB through fruits, vegetables, exercise, and feeling nervous.

A number of reinforcing feedback loops may be found in Figure 2, which may be described as follows (with implicit intermediate variables in parentheses):

1. Inadequate exercise may lead to AdOWOB, and AdOWOB in turn may further inhibit willingness to exercise, especially in public.
2. Nervousness may lead to AdOWOB (because of high-calorie consumption), and AdOWOB may in turn lead to greater nervousness.
3. Inadequate breakfast or skipping breakfast may lead to AdOWOB (because of high-calorie consumption during the rest of the day), and AdOWOB in turn may cause some adolescents to skip breakfast (a disruption of normal eating patterns, perhaps an informal or unreported form of dieting).
4. Inadequate breakfast may also lead to less consumption of vegetables during the day (because of high-calorie consumption in place of vegetables), which may result in AdOWOB, and in turn back to skipping breakfast.
5. Dieting may cause some adolescents to skip breakfast, which may lead to AdOWOB, in turn leading to even stricter dieting.
6. Lack of exercise may lead to lack of care about healthy eating—less vegetables, less fruits, and skipping breakfast—which may lead to AdOWOB, and AdOWOB in turn may inhibit exercise in public.
7. Persistent nervousness may lead to feeling low, which may suppress the desire to exercise, which, in turn, may fairly quickly lead to even greater nervousness. (This is the one feedback loop in Figure 2 that is purely cognitive-behavioral and does not include AdOWOB.)
8. The preceding cognitive-behavioral loop fans out to cause worse nutrition—vegetables, fruits, breakfast—which can lead to AdOWOB, which in turn may lead back to greater nervousness.

AdOWOB and Feel Low are the model's two stock variables, both formulated as simple first-order adjustment (balancing loop) processes with appropriate time constants. The stock of AdOWOB is affected by exogenous AdOWOBY with an adjustment time of 2.5 years and by the factors of adolescence with an adjustment time of 2.0 years.

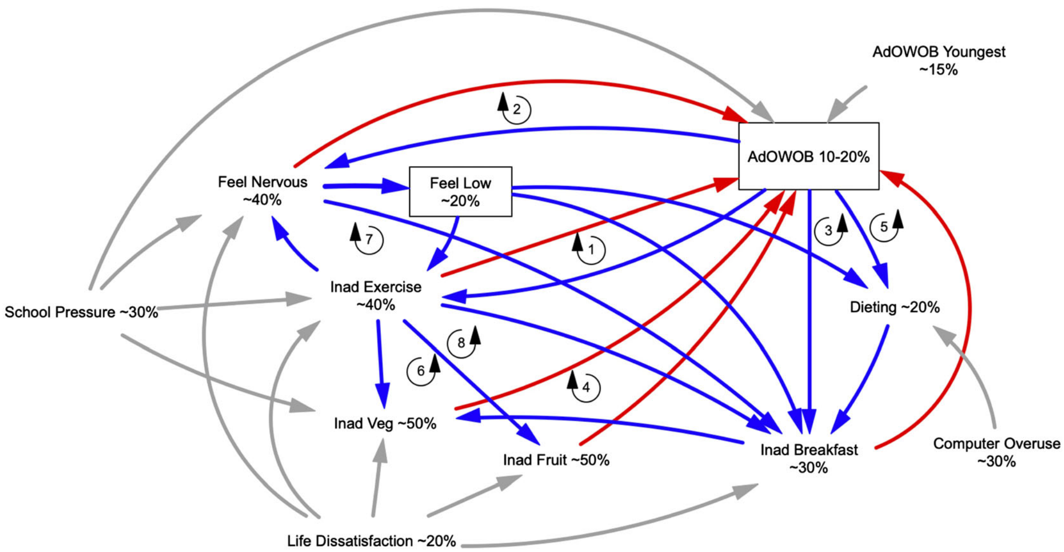

FIGURE 2 Dynamic hypothesis suggested by statistical analysis of HBSC data and supported by AdOWOB literature; 31 countries × 4 gender-perceived wealth segments, 2002–2014. Veg, vegetables. Typical prevalence percentages shown for eight endogenous and four exogenous variables. All the links have positive polarity, i.e., for pair variables, an increase in an input variable, all else equal, leads to an increase in an output variable (the same for the case of decrease). Red, endogenous influence direct to AdOWOB; blue, other endogenous influence; grey, exogenous influence. The numbered feedback loops (all reinforcing) are described in the main text. Rectangle, stock variable (including simple adjustment balancing loops not shown in the figure). The stock of AdOWOB is affected by exogenous AdOWOB Youngest and by the factors of adolescence over several years. The stock of feel low changes much more quickly, in under 1 year.

The stock of Feel Low changes much faster, with an adjustment time of only one-quarter of a year.

Complete model documentation is provided in the Supporting Information (Table S3: equations listing for a generic case model; Tables S4–9: data for running case-specific simulations; Table S10: initial HR values prior to running optimization).

3.2 | Base run: Optimized hazard ratios and fit to history

Table 2 reports the estimates of the 30 HRs from optimizing the generalized model to the 24 country-segment cases (using Opt200k). To illustrate specific cases, the middle columns list the estimates for less well-off boys in Norway (Norway LWOB) and less well-off girls in the Netherlands (NL LWOG). For each HR, the last two columns report (a) the number of cases where the HR is significant (above 1.05) and (b) the maximum estimated value of the HR across all the cases. Table S11 reports this information for the Opt50k optimization; the estimates under Opt50k are numerically different from those under 200 k, but generally close in value.

Table 3 reports the summary fit statistics across the optimized cases. We find that the model provides good explanatory value, in terms of MAPE and R^2i, for the great majority of cases. Only three of the 24 cases (England LWOG, NL MWOB, and Norway LWOG) do not show a good fit to history (MAPE > 15% and R^2i < 21%).

TABLE 2 Estimated HRs using Opt200k. HR values are shown for two of the 24 cases and are summarized for all 24 cases

HR parameter		HR values for two example cases		Summary across 24 cases	
#	Name	Norway LWOB	NL LWOG	# of cases with HR ≥ 1.05	Max HR of 24 cases
1	HR of Dieting for InadBkfast	8.00	1.00	17	8.00
2	HR of CompOveruse for Dieting	4.37	2.26	21	7.88
3	HR of FeelLow for Dieting	5.00	6.26	15	6.26
4	HR of FeelLow for InadBkfast	4.33	1.00	12	5.00
5	HR of FeelLow for InadEx	1.00	2.38	15	10.00
6	HR of FeelNerv for FeelLow	28.76	6.25	23	30.96
7	HR of FeelNerv for InadBkfast	4.77	1.00	17	4.77
8	HR of FeelNerv for AdOWOB	2.17	1.00	8	5.00
9	HR of InadBkfast for InadVeg	1.89	2.60	16	4.30
10	HR of InadBkfast for AdOWOB	5.00	1.26	14	5.15
11	HR of InadEx for FeelNerv	1.00	2.98	19	8.00
12	HR of InadEx for InadBkfast	2.98	1.96	18	5.00
13	HR of InadEx for InadFruit	1.36	4.99	20	6.65
14	HR of InadEx for InadVeg	7.00	1.00	20	7.00
15	HR of InadEx for AdOWOB	1.00	1.61	11	6.11
16	HR of InadFruit for AdOWOB	1.00	4.98	19	4.98
17	HR of InadVeg for AdOWOB	1.00	2.61	12	3.05
18	HR of LifeDissat for FeelNerv	1.00	5.35	17	9.99
19	HR of LifeDissat for InadBkfast	1.38	1.00	17	3.74
20	HR of LifeDissat for InadEx	16.50	1.00	20	16.50
21	HR of LifeDissat for InadFruit	10.00	1.00	17	10.00
22	HR of LifeDissat for InadVeg	8.12	5.21	16	8.12
23	HR of AdOWOB for Dieting	1.26	3.99	22	9.08
24	HR of AdOWOB for FeelNerv	2.92	2.75	16	8.06
25	HR of AdOWOB for InadBkfast	9.03	1.00	18	10.90
26	HR of AdOWOB for InadEx	6.22	1.00	14	11.37
27	HR of SchoolPr for FeelNerv	3.21	8.14	21	8.43
28	HR of SchoolPr for InadEx	5.93	1.00	7	5.93
29	HR of SchoolPr for InadVeg	1.47	1.00	9	8.11
30	HR of SchoolPr for AdOWOB	3.18	1.00	11	6.89
	# of HRs with estimated value ≥ 1.05:	24	17	Green: 16+ cases have estimated value of this parameter ≥ 1.05	
		Pink: HR value < 1.05			

Bkfast, Breakfast; Ex, Exercise; Veg, Vegetables; LifeDissat, Life Dissatisfaction; SchoolPr, School Pressure.

TABLE 3 Summary fit-to-history statistics for the 24 cases for Opt200k

Country and Segment	Model MAPE (mean absolute % error[a])		Model R^2i (% of data variance vs. initial value explained)	
	MAPE of AdOWOB	Mean across 8 variables	R^2i of AdOWOB	Mean across 8 variables
Avg31 LWOB	4%	3%	95%	82%
Avg31 LWOG	4%	2%	94%	75%
Avg31 MWOB	5%	3%	80%	81%
Avg31 MWOG	7%	4%	75%	44%
England LWOB	8%	8%	95%	69%
England LWOG	25%	10%	14%	41%
England MWOB	26%	11%	73%	69%
England MWOG	21%	10%	74%	57%
NL LWOB	10%	9%	65%	48%
NL LWOG	11%	12%	6%	27%
NL MWOB	15%	15%	0%	25%
NL MWOG	1%	10%	100%	46%
Norway LWOB	2%	3%	90%	89%
Norway LWOG	19%	8%	0%	50%
Norway MWOB	4%	5%	39%	75%
Norway MWOG	7%	7%	58%	55%
Poland LWOB	7%	7%	97%	67%
Poland LWOG	25%	8%	48%	47%
Poland MWOB	16%	10%	67%	42%
Poland MWOG	8%	8%	96%	62%
Portugal LWOB	3%	11%	87%	49%
Portugal LWOG	7%	5%	42%	61%
Portugal MWOB	1%	10%	99%	69%
Portugal MWOG	1%	5%	99%	60%

Note: MAPE is the mean absolute percentage error. R^2 is the fraction of variance from the initial value explained. The statistics are calculated for AdOWOB as well as for the model's other seven endogenous variables. Good fit: MAPE < 15%; R^2i > 21%. Avg31, HBSC weighted average for 31 countries; Veg, Vegetables.

[a]MAPE basis: 2006–2014 for AdOWOB, 2002–2014 for all other variables.

We present two time graphs to illustrate the model's ability to replicate historical trajectories of AdOWOB, again using the example cases of Norway LWOB (Figure 3) and NL LWOG (Figure 4). These graphs compare four different simulations with the HBSC data (dotted red line) during 2002–2014 and simulate forward to 2022. The simulations all assume exogenous factors remaining constant after 2014 at their 2014 values and no interventions. They may be viewed as a sequence from the crudest to the most refined. First, "AdOWOBY-only" (black dashed line) is a simulation driven only by AdOWOBY with all HR values set to 1.0, thus neutralizing the influence of all factors of adolescence. Second, "Regression-based" (green dashed line) is a simulation using HRs set to the values based on the statistical regression for the gender-perceived wealth segment in question but across all 31 countries, without optimization for the country in question. Third, "Opt50k" (blue dotted line) uses the HR values from the 50 k optimization for the specific country and segment. Fourth, "Opt200k" (blue solid line) uses the HR values from the 200 k optimization for the specific country and segment.

The graphs show how greater refinement leads generally to a closer fit to the historical data. Changes in AdOWOBY help explain the data trajectories, but the factors of adolescence add more explanatory power, first without country-specific optimization in the regression-based simulation and then with country-specific optimization in Opt50k and Opt200k. The results of Opt50k and Opt200k are so close as to be indistinguishable, which means that the model is insensitive to the differences in their estimated HR values.

Note that even with country-specific optimization, the model may occasionally miss a data point; see especially 2010 for NL LWOG in Figure 4. Such a miss can happen when none of the model's explanatory variables anticipates the observed change in AdOWOB. When this occurs, it will tend to worsen one or both of the summary statistics in Table 3. In the case of NL LWOG, the MAPE statistic is satisfactory, but the R^2i for AdOWOB is poor because of the miss in 2010.

From a review of all case-specific graphs (24 cases × 8 endogenous variables; see Figures S2–9), we find that the same basic model is able to replicate a wide variety of trajectories and patterns seen in

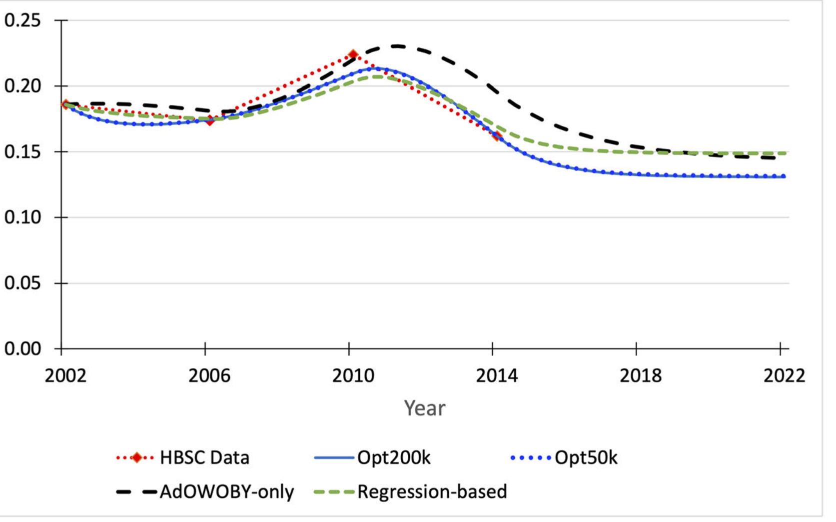

FIGURE 3 Simulated AdOWOB (under four parameter settings) vs. HBSC data, for the case of Norway LWOB. The four parameter settings, from crudest to most refined: "AdOWOBY-only" (black dashed line) is a simulation driven only by AdOWOBY with all HR values set to 1.0, thus neutralizing the influence of all factors of adolescence. "Regression-based" (green dashed line) is a simulation using HRs set to the values based on the statistical regression for the gender-perceived wealth segment in question but across all 31 countries, without optimization for the country in question. "Opt50k" (blue dotted line) uses the HR values from the 50 k optimization for the specific country and segment. "Opt200k" (blue solid line) uses the HR values from the 200 k optimization for the specific country and segment.

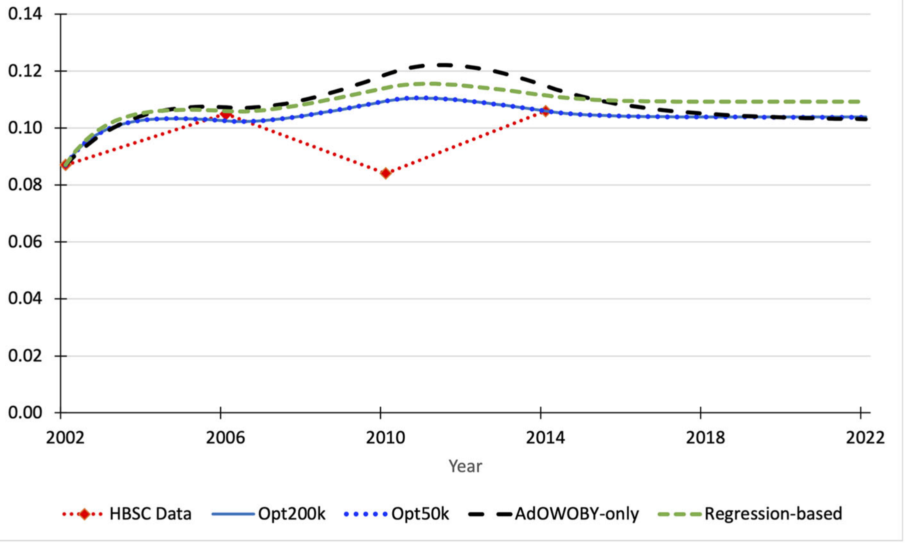

FIGURE 4 Simulated AdOWOB (under four parameter settings) vs. HBSC data, for the case of the Netherlands LWOG. The four parameter settings, from crudest to most refined: "AdOWOBY-only" (black dashed line) is a simulation driven only by AdOWOBY with all HR values set to 1.0, thus neutralizing the influence of all factors of adolescence. "Regression-based" (green dashed line) is a simulation using HRs set to the values based on the statistical regression for the gender-perceived wealth segment in question but across all 31 countries, without optimization for the country in question. "Opt50k" (blue dotted line) uses the HR values from the 50 k optimization for the specific country and segment. "Opt200k" (blue solid line) uses the HR values from the 200 k optimization for the specific country and segment.

the data. Table 3 tells us that the model's explanatory power is generally strong, with only a minority of exceptions.

3.3 | Intervention points testing

Intervention testing results for each of the 24 country-segment cases are reported in Table 4 in terms of percentage decrease in AdOWOB as of 2026 relative to the base run. The results seen in this table are for the Opt200k calibrations of the model, but the same testing using the Opt50k calibrations gives very similar results (see Tables S12 and S13).

Testing the impact of all 10 intervention points combined resulted in substantial (more than 8%) reductions in AdOWOB in 19 cases, moderate (2%–8%) reductions in two cases, and negligible changes in three cases (Netherlands LWOB, Norway MWOB, and Norway MWOG). In these three cases, there were no strong causal paths (i.e., with HRs greater than 1.0 all along the path) leading to AdOWOB.

Figure 5 shows the trajectories of simulated AdOWOB for a single case, Norway LWOB, under the tested intervention points. (Fruit and vegetable interventions did not reduce AdOWOB for this segment, because of HR values of 1.0 as reported in Table 2, HR's #16 and 17 for Norway LWOB.) These trajectories all follow a declining goal-seeking pattern, with the majority of the reduction occurring by

TABLE 4 Intervention points testing results (Opt200k): Percentage decrease in AdOWOB

Country & Segment	AdOWOB % decrease from base run as of 2026 after 0.75x intervention, by intervention point									
	Computer overuse	Life Dissat	School pressure	Inad breakfast	Dieting	Inad exercise	Feel low	Feel nervous	Inad fruit	Inad veg	All 10 combined
Avg31 LWOB	1.0%	10.2%	0.5%	9.2%	2.0%	12.8%	1.0%	2.0%	5.1%	0.0%	28.1%
Avg31 LWOG	0.0%	18.3%	3.1%	5.3%	0.0%	19.8%	0.0%	9.9%	9.9%	6.9%	42.0%
Avg31 MWOB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.6%	0.0%	0.0%	5.6%
Avg31 MWOG	0.0%	3.9%	0.0%	0.0%	0.0%	2.9%	0.0%	0.0%	8.7%	0.0%	12.6%
England LWOB	0.0%	1.4%	2.8%	2.3%	0.5%	1.4%	0.0%	0.5%	0.0%	1.8%	8.7%
England LWOG	0.0%	1.3%	4.0%	0.0%	0.0%	2.2%	0.9%	0.9%	3.5%	0.0%	9.7%
England MWOB	0.0%	0.7%	6.3%	0.0%	0.0%	0.7%	0.0%	0.0%	2.1%	0.0%	9.1%
England MWOG	0.0%	2.5%	5.6%	0.0%	0.0%	1.9%	0.0%	0.0%	5.6%	1.9%	14.8%
NL LWOB	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	0.7%
NL LWOG	0.0%	2.9%	1.0%	1.9%	0.0%	7.7%	1.9%	1.9%	9.6%	4.8%	22.1%
NL MWOB	0.0%	1.3%	9.3%	0.0%	0.0%	2.7%	1.3%	1.3%	5.3%	0.0%	18.7%
NL MWOG	0.0%	14.6%	22.8%	8.9%	0.0%	27.6%	22.8%	26.0%	16.3%	8.9%	45.5%
Norway LWOB	2.3%	2.3%	10.7%	6.1%	3.8%	2.3%	3.8%	8.4%	0.0%	0.0%	21.4%
Norway LWOG	0.0%	0.0%	3.8%	0.0%	0.0%	3.8%	0.0%	0.0%	9.1%	2.3%	15.9%
Norway MWOB	0.0%	0.0%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	1.6%
Norway MWOG	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Poland LWOB	0.5%	3.5%	0.5%	4.5%	1.5%	3.0%	0.0%	3.0%	1.0%	4.5%	14.6%
Poland LWOG	0.0%	2.2%	3.3%	2.2%	1.1%	4.3%	0.0%	1.1%	6.5%	0.0%	16.3%
Poland MWOB	0.0%	5.5%	0.0%	4.3%	0.6%	6.7%	0.0%	1.2%	6.7%	0.0%	19.0%
Poland MWOG	0.0%	3.3%	4.1%	0.8%	0.0%	3.3%	0.8%	0.8%	5.7%	2.5%	14.8%
Portugal LWOB	0.0%	0.5%	0.0%	0.9%	0.0%	0.0%	0.0%	0.0%	2.3%	0.0%	3.7%
Portugal LWOG	0.7%	1.3%	0.7%	2.0%	0.7%	3.4%	0.7%	1.3%	10.1%	0.0%	15.4%
Portugal MWOB	0.0%	0.0%	2.6%	0.5%	0.0%	0.5%	0.0%	0.0%	5.2%	3.6%	12.0%
Portugal MWOG	0.0%	0.6%	0.0%	0.0%	0.0%	1.3%	0.0%	0.0%	7.6%	1.9%	11.5%

Note: Blue: >2%, Green: >8%. Avg31, HBSC weighted average for 31 countries; Veg, Vegetables.

2020 and stabilizing by 2024. The single most effective intervention point in 2026 for this case is school pressure (10.7%), followed by feeling nervous (8.4%) and inadequate breakfast (6.1%). Adding up the 10 individual impact fractions gives 39.7%, but the combined intervention simulation shows an impact of only 21.4%. This negative synergy is the result of diminishing returns as the opportunity for further improvement in AdOWOB (and all endogenous variables) declines after each individual intervention. This is only one example of 24, and the results can vary in magnitude and ranking from one case to another.

Based on the percentage decreases in AdOWOB, we ranked the intervention points from 1 to 10 for each of the 24 cases, as reported in Table 5. The last two rows of this table provide a count for the number of cases in which an intervention point was ranked #1 or 2 and also a count for the number of cases in which it was ranked #3 or 4. Five intervention points stand out as most impactful across the 24 cases based on the number of top-four rankings: exercise (18 top-four rankings, including 4 at #1 and 10 at #2), fruit (16, with 11 at #1 and 3 at #2), life dissatisfaction (16, with 3 at #2), school pressure (11, with 6 at #1 and 2 at #2), and skipping breakfast (10, with 2 at #1 and 2 at #2).

Looking at the details, we note the following:

- Increasing exercise and fruit were generally in the top four for all gender-perceived wealth segments, for Europe overall (Avg31) and in all CO-CREATE countries except Norway;
- Reducing life dissatisfaction was generally in the top four for all gender-perceived wealth segments, for Europe overall and in all CO-CREATE countries except Norway;
- Reducing school pressure was a top four intervention for all segments of England, two segments each of the Netherlands and Norway, and one segment each of Poland and Portugal, but not for Europe overall;
- Skipping breakfast was a top four intervention for three of the four segments of Poland and Portugal, two of the segments in Norway, and for one segment of England and Europe overall. For all countries except the Netherlands, skipping breakfast appears to be a particularly significant intervention point for LWOB.

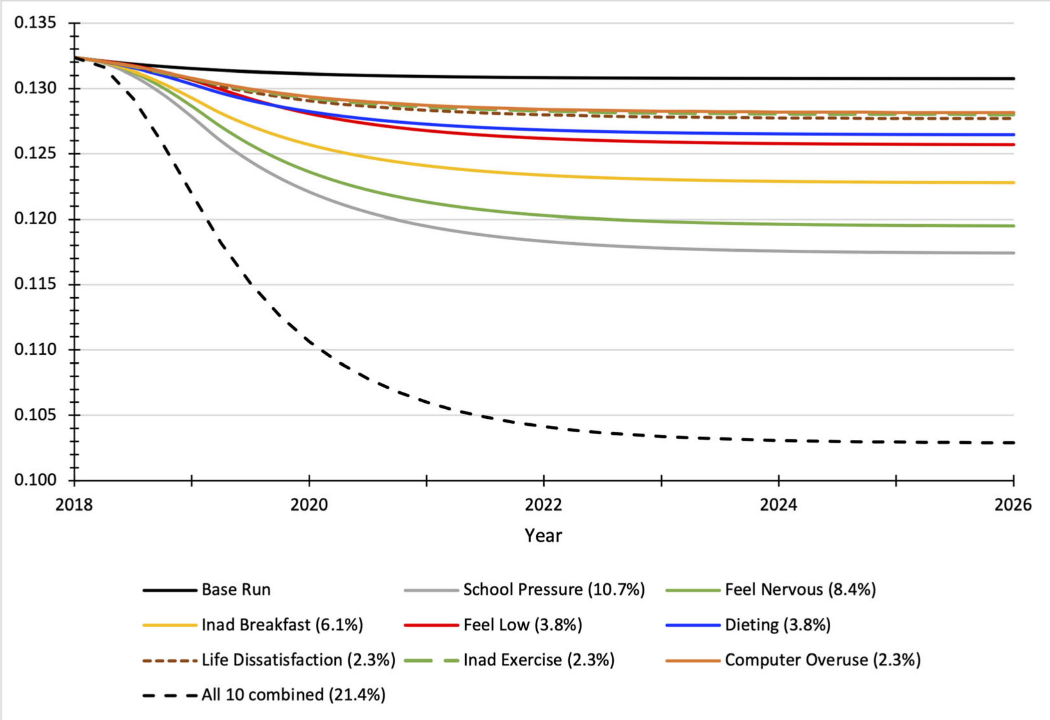

FIGURE 5 Simulated AdOWOB under the tested intervention points, for the case of Norway less well-off boys (Opt200k). Numbers in brackets show % decrease in AdOWOB from base run as of 2026; intervention runs for Inad fruit and Inad vegetables are the same as base run and not shown here.

Among the less impactful intervention points, there are two that made the top four in an intermediate number of cases (vegetables[8] and feeling nervous[7]), and three that made the top four in very few cases (feeling low,[3] dieting,[1] and computer overuse [0]).

4 | DISCUSSION

4.1 | Significance

Persistently high prevalence of AdOWOB in Europe, as well as a growing recognition that this problem is driven by a wide array of interconnected factors, calls for the use of systems methods to evaluate potential intervention and policy options.[51] Such methods often identify a few higher-leverage intervention points that translate to priorities for allocating limited public health resources. Here, we have used a combination of data analysis and SD modeling to analyze potential intervention points in AdOWOB as part of the CO-CREATE project. We believe that this work contributes to the literature both for its results and as an advance in methodology for public health analysis.

Our study builds on previous modeling work that utilized data to explore and simulate the impacts of intertwined social determinants on health conditions.[32,33] These earlier studies analyzed cross-sectional data for one or two periods of data and estimated causal link strengths, but they did not attempt to replicate historical behavior over time and rather assumed a steady-state base run. Thus, our SD study of social determinants is the first to utilize multiperiod data for dynamic model validation.

We identified five intervention points as most impactful across the 24 cases based on the number of top-four rankings. These top intervention points were exercise, fruits, life dissatisfaction, school pressure, and skipping breakfast. These priority areas can be compared with the four policy ideas suggested by adolescents themselves in the CO-CREATE project, which related to (1) marketing of unhealthy foods, (2) nutrition education in school and healthy school cafeteria, (3) SSB tax, and (4) free organized physical activities. The fourth of these clearly corresponds to our priority area of exercise. The other three, dealing with nutrition, correspond to our priority areas of encouraging fruit consumption and regular breakfast. Unfortunately, we could not consider an SSB tax directly, because of problems with the HBSC SSB variable as discussed earlier in this paper. Based on our dynamic hypothesis, one would expect the variables with direct, stronger, and multiple pathways to AdOWOB to be more effective. Indeed, lacking fruits had more counts of significant HRs than any other direct driver of AdOWOB. Skipping breakfast and lack of exercise showed strong direct links to AdOWOB in fewer cases but often exhibited significant indirect pathways leading to AdOWOB. These priorities are also in agreement with the most common areas for interventions identified by the systematic reviews, including the recent ones by the EU-funded STOP (Science and Technology in childhood Obesity Policy) project.[52]

TABLE 5 Intervention points testing results (Opt200k): Rankings (1–10)

Country & Segment	Ranking of intervention points by AdOWOB % decrease from base run as of 2026 (1 = best)									
	Computer overuse	Life Dissat	School pressure	Inad breakfast	Dieting	Inad exercise	Feel low	Feel nervous	Inad fruit	Inad veg
Avg31 LWOB	7	2	9	3	5	1	7	5	4	--
Avg31 LWOG	--	2	7	6	--	1	--	3	3	5
Avg31 MWOB	--	--	--	--	--	--	--	1	--	--
Avg31 MWOG	--	2	--	--	--	3	--	--	1	--
England LWOB	--	4	1	2	6	4	--	6	--	3
England LWOG	--	4	1	--	--	3	5	5	2	--
England MWOB	--	3	1	--	--	3	--	--	2	--
England MWOG	--	3	1	--	--	4	--	--	1	4
NL LWOB	--	--	--	--	--	--	--	--	--	1
NL LWOG	--	4	8	5	--	2	5	5	1	3
NL MWOB	--	4	1	--	--	3	4	4	2	--
NL MWOG	--	6	3	7	--	1	3	2	5	7
Norway LWOB	6	6	1	3	4	6	4	2	--	--
Norway LWOG	--	--	2	--	--	2	--	--	1	4
Norway MWOB	--	--	--	1	--	--	--	--	--	--
Norway MWOG	--	--	--	--	--	--	--	--	--	--
Poland LWOB	8	3	8	1	6	4	--	4	7	1
Poland LWOG	--	4	3	4	6	2	--	6	1	--
Poland MWOB	--	3	--	4	6	1	--	5	1	--
Poland MWOG	--	3	2	6	--	3	6	6	1	5
Portugal LWOB	--	3	--	2	--	--	--	--	1	--
Portugal LWOG	6	4	6	3	6	2	6	4	1	--
Portugal MWOB	--	--	3	4	--	4	--	--	1	2
Portugal MWOG	--	4	--	--	--	3	--	--	1	2
Total count, any 1–2	0	3	8	4	0	8	0	3	14	4
Total count, any 3–4	0	13	3	6	1	10	3	4	2	4

Note: Blue: rank 1–2, Green: rank 3–4. Avg31, HBSC weighted average for 31 countries; Veg, Vegetables.

Two of our top intervention points were not suggested by the CO-CREATE adolescents, namely, reducing life dissatisfaction and reducing school pressure. Both of these refer to social factors (influenced, for example, by family background and educational policies) that lie beyond the usual discussion of exercise and nutrition with respect to AdOWOB. However, the literature does support the importance of these two intervention points. With regard to the first, studies indicate that better life satisfaction may lead to a healthier diet[53] and greater levels of participation in structured extracurricular sports activities and may also act as a buffer against the negative effects of stress.[54] With regard to school pressure, studies indicate that it can cause stress, interfere with cognitive processes, and trigger psychological and biochemical processes that lead to AdOWOB and also can have adverse effects on exercising and nutrition.[55] Our identification of these two priority intervention points also lends empirical support to those who argue for a prominent role of environmental factors in AdOWOB[29] as the adolescents' perceptions of these two factors may be seen as their responses to how society is organized.

Our analysis found that interventions addressing vegetable consumption, dieting, feeling nervous, feeling low, and computer overuse were less impactful than the top five intervention points. That is not to say they are unimportant but simply that they appear to have less to contribute, based on this dynamic model calibrated to the HBSC data. The rest of the less impactful intervention points have not been among the priority areas in systematic reviews, although there is an emerging interest in the role of mental health as a driver of AdOWOB.[56] Our analysis suggests that the mental health variables (feeling nervous and feeling low) and dieting are important conduits for causal pathways from other variables but are not in themselves the most effective places to intervene in the modeled system. Computer overuse only impacted dieting in our model and is, therefore, constrained by the effectiveness of reducing dieting as an intervention point.

Besides providing a richer collection of intervention points, the approach taken in this study allowed us to explore possible variations by gender, perceived wealth, and country. Three of the top interventions—increasing exercise and fruit and reducing life

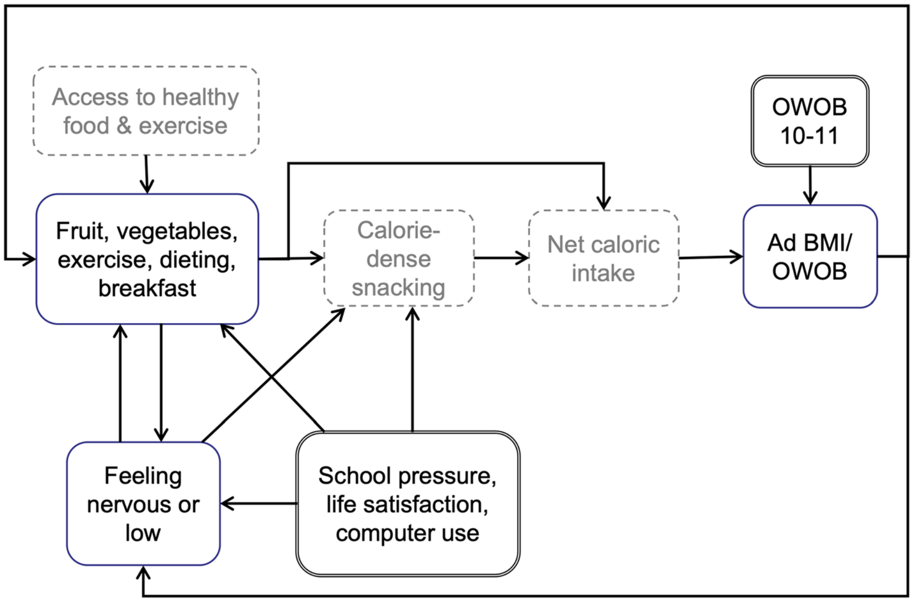

FIGURE 6 Expanded view of a future model requiring additional data sources. Dashed boxes indicate "hidden" variables (lacking data).

dissatisfaction—proved effective for all gender-perceived wealth segments, for Europe overall and in all CO-CREATE countries except Norway. The results are less uniform for reducing school pressure and promoting breakfast. However, we did find that for LWOB, promoting breakfast was a particularly significant intervention point for Europe overall and in all CO-CREATE countries except the Netherlands. This finding can be viewed in the light of systematic reviews on AdOWOB interventions, which note that some interventions are less effective for lower-income groups.[52] Our analysis suggests an exception: when it comes to LWOB, promoting a regular breakfast may be an effective intervention.

4.2 | Limitations and extensions

Our analysis has a few noteworthy limitations. First, we had only four survey data points and eight endogenous variables to help with the estimation of the 30 uncertain HRs. Even though the behavior of the model and the results of intervention testing did not appear to be sensitive to uncertainty in parameter estimates, additional data points would have improved the robustness of our analysis. Our model will benefit from further parameter refinement and validation when the HBSC 2018 dataset is publicly released (expected in Fall 2022). Structural sensitivity analysis is another important evaluative technique[24] but was beyond the scope of this paper.

Second, it is not necessarily the case that all 10 intervention points are, in real life, equally amenable to the 25% reduction we assumed. The choice of the same 25% effect size for all interventions was guided by an examination of historical changes from the HBSC data. However, this procedure was not exact and did not consider the impacts or costs of specific policies or programs. Our approach examined different areas of intervention broadly rather than intervention details, as considered by some other public health modeling studies.[57,58]

Third, our analysis was limited by the variables that were available in the HBSC dataset. In order to explore the stability of associations between various factors, as well as to be able to validate our model using historical time series, we were limited to only those variables that were asked consistently over all the four survey periods (e.g., the HBSC variable "Have dinner with family" was excluded for this reason). Also, we could not include variables with possible definitional difficulties, as in the case of SSBs. The findings of the analysis are also limited by potential biases and weaknesses of the HBSC data itself (such as the data being self-reported), yet the ability of the generalized dynamic model to reproduce reasonably well the data trajectories of eight endogenous variables for 24 cases speaks to the apparent power of our dynamic hypothesis.

For future work, we may want to identify other sources of data that could supplement the HBSC dataset and perhaps expand the boundary of our analysis slightly. Figure 6 portrays a simplified view of a possible expanded model. In this figure, the variables in black boxes are the ones included in the current model, and the variables in dashed boxes indicate "hidden" variables requiring more data. These hidden variables start with net caloric intake, which has been considered by other modeling studies using objective survey data for the United States.[42,59,60] We would also benefit from data on snacking and access to healthy food and exercise. The local food environment and built environment play key roles in other systems frameworks of obesity,[21,61] but their quantification remains a challenge.

5 | CONCLUSIONS

This study demonstrates that SD modeling and simulation, supported by the appropriate use of multiperiod data, is well-suited for analyzing AdOWOB as a dynamic system (comprising multiple interacting behavioral and psychological factors of adolescence) and identifying the most influential points for intervention. Three of the top intervention priorities identified by our analysis (exercise, eating fruits, and eating breakfast) are in line with the exercise and nutrition-related policy ideas suggested by youth from the EU's CO-CREATE project, whereas two other priorities (reducing life dissatisfaction and school pressure) extend beyond those. These additional priority intervention points can be used to enrich future policy discussions among youth and other stakeholders. This work also contributes to the growing literature linking empirical data to dynamic socioecological modeling of health-related conditions like AdOWOB.

ACKNOWLEDGMENTS

We would like to thank Knut Inge Klepp (Norwegian Institute of Public Health) and Birgit Kopainsky (University of Bergen) for their useful feedback and Trond Helland (University of Bergen) for his support and guidance in using the HBSC data. HBSC is an international study carried out in collaboration with WHO/EURO. The International Coordinator of the 2001/02, 2005/05, 2009/10, and 2013/14 surveys was Prof. Candace Currie at the University of St. Andrews, Scotland, and the Data Bank Manager was Prof. Oddrun Samdal at the University of Bergen, Norway. For details of participating countries, see http://www.hbsc.org.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Eduard Romanenko https://orcid.org/0000-0002-0776-1780
Jack Homer https://orcid.org/0000-0003-3690-9892
Anne-Siri Fismen https://orcid.org/0000-0001-6711-5767
Harry Rutter https://orcid.org/0000-0002-9322-0656
Nanna Lien https://orcid.org/0000-0003-1486-4769

REFERENCES

1. Lobstein T, Jackson-Leach R. Planning for the worst: estimates of obesity and comorbidities in school-age children in 2025. *Pediatr Obes*. 2016;11(5):321-325. doi:10.1111/ijpo.12185
2. World Health Organization. Consideration of the evidence on childhood obesity for the Commission on Ending Childhood Obesity: report of the ad hoc working group on science and evidence for ending childhood obesity, Geneva, Switzerland. 2016.
3. Quek YH, Tam WW, Zhang MW, Ho RC. Exploring the association between childhood and adolescent obesity and depression: a meta-analysis. *Obes Rev*. 2017;18(7):742-754. doi:10.1111/obr.12535
4. Park MH, Falconer C, Viner RM, Kinra S. The impact of childhood obesity on morbidity and mortality in adulthood: a systematic review. *Obes Rev*. 2012;13(11):985-1000. doi:10.1111/j.1467-789X.2012.01015.x
5. Inchley J, Currie D, Budisavljevic S, et al. (Eds). Spotlight on adolescent health and well-being. Findings from the 2017/2018 Health Behaviour in School-aged Children (HBSC) survey in Europe and Canada. 2020. International report; No. 1. Key findings.
6. Jackson-Leach R, Montague F, Lobstein T. *Obesity Atlas for the European Union: 2017*. London, UK: World Obesity Federation; 2016.
7. Inchley J, Currie D, Jewell J, Breda Jo, Barnekow V. (Eds). *Adolescent obesity and related behaviours: Trends and inequalities in the WHO European Region, 2002–2014: Observations from the Health Behaviour in School-aged Children (HBSC) WHO collaborative cross-national study*. Copenhagen: WHO Regional Office for Europe; 2017.
8. Wijnhoven T, van Raaij J, Spinelli A, et al. WHO European Childhood Obesity Surveillance Initiative: body mass index and level of overweight among 6–9-year-old children from school year 2007/2008 to school year 2009/2010. *BMC Public Health*. 2014;14(1):1, 806-16. doi:10.1186/1471-2458-14-806
9. European Commission. *EU Action Plan on Childhood Obesity 2014–2020*. Belgium: European Commission Brussels; 2014.
10. World Health Organization. *Global action plan for the prevention and control of noncommunicable diseases 2013–2020*. World Health Organization; 2013.
11. World Health Organization. *Report of the commission on ending childhood obesity*. World Health Organization; 2016.
12. World Health Organization. European food and nutrition action plan 2015–2020. 2015.
13. United Nations. The Sustainable Development Goals Report 2017. 2017.
14. DiClemente RJ, Salazar LF, Crosby RA. *Health behavior theory for public health: Principles, foundations, and applications*. Jones & Bartlett Publishers; 2013.
15. Waters E, de Silva-Sanigorski A, Burford BJ, et al. Interventions for preventing obesity in children. *Cochrane Database Syst Rev*. 2011;12:CD001871.
16. http://www.nhs.uk/change4life. Accessed April 11, 2022.
17. AMEA. http://ameaprogram.com/amea-teens/. Accessed April 11, 2022.
18. Salas XR. The ineffectiveness and unintended consequences of the public health war on obesity. *Can J Public Health*. 2015;106(2):e79-e81. doi:10.17269/cjph.106.4757
19. Rutter H, Bes-Rastrollo M, De Henauw S, et al. Balancing upstream and downstream measures to tackle the obesity epidemic: a position statement from the European Association for the Study of Obesity. *Obes Facts*. 2017;10(1):61-63. doi:10.1159/000455960
20. Finegood DT, Merth TD, Rutter H. Implications of the foresight obesity system map for solutions to childhood obesity. *Obesity (Silver Spring)*. 2010;18(n1s):S13-S16. doi:10.1038/oby.2009.426
21. Koplan JP, Liverman CT, Kraak VA. (Eds). *Preventing childhood obesity: Health in the balance. Institute of Medicine, Committee on Prevention of Obesity in Children and Youth*. Washington DC: National Academies Press; 2005.
22. www.co-create.eu. Accessed June 2, 2022.
23. Norwegian Institute of Public Health. The CO-CREATE Youth Declaration: Time to Act and Ensure Good Health for All. https://www.fhi.no/contentassets/0a74196d35c64da89d337e25af982f5f/co-create-youth-declaration-on-ending-childhood-and-adolescent-obesity.pdf. Published 2020. Accessed March 8, 2022.
24. Sterman J. *Business dynamics*. McGraw-Hill, Inc; 2000.
25. Savona N, Macauley T, Aguiar A, et al. Identifying the views of adolescents in five European countries on the drivers of obesity using group model building. *Eur J Public Health*. 2021;31(2):391-396. doi:10.1093/eurpub/ckaa251
26. Barlas Y. Formal aspects of model validity and validation in system dynamics. *Syst Dyn Rev*. 1996;12(3):183-210. doi:10.1002/(SICI)1099-1727(199623)12:3<183::AID-SDR103>3.0.CO;2-4
27. Aguiar A, Gebremariam M, Kopainsky B, Savona N, Allender S, Lien N. *Review of existing system dynamics models on overweight/obesity in children and adolescents*. University of Oslo 2019. https://ec.europa.eu/research/participants/documents/

downloadPublic?documentIds=080166e5c8d1c9d0&appId=PPGMS. Accessed June 1, 2022.
28. Morshed AB, Kasman M, Heuberger B, Hammond RA, Hovmand PS. A systematic review of system dynamics and agent-based obesity models: evaluating obesity as part of the global syndemic. *Obes Rev*. 2019;20(S2):161-178. doi:10.1111/obr.12877
29. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006;27(1):297-322. doi:10.1146/annurev.publhealth.27.021405.102100
30. Pereira M, Padez C, Nogueira H. Describing studies on childhood obesity determinants by Socio-Ecological Model level: a scoping review to identify gaps and provide guidance for future research. *Int J Obes (Lond)*. 2019;43(10):1883-1890. doi:10.1038/s41366-019-0411-3
31. Homer J. Best practices in system dynamics modeling, revisited: a practitioner's view. *Syst Dyn Rev*. 2019;35(2):177-181. doi:10.1002/sdr.1630
32. Mahamoud A, Roche B, Homer J. Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. *Soc Sci Med*. 2013;93:247-255. doi:10.1016/j.socscimed.2012.06.036
33. Milstein B, Homer J. Which priorities for health and well-being stand out after accounting for tangled threats and costs? Simulating potential intervention portfolios in large urban counties. *Milbank Q*. 2020;98(2):372-398. doi:10.1111/1468-0009.12448
34. Jebeile H, Kelly AS, O'Malley G, Baur LA. Obesity in children and adolescents: epidemiology, causes, assessment, and management. *The Lancet Diabetes & Endocrinology*. 2022;10(5):351-365. doi:10.1016/S2213-8587(22)00047-X
35. Malik VS, Pan A, Willett WC, Hu FB. Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. *Am J Clin Nutr*. 2013;98(4):1084-1102. doi:10.3945/ajcn.113.058362
36. Swinburn BA, Caterson I, Seidell JC, James WPT. Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutr*. 2004;7(1a):123-146.
37. van Ekris E, Altenburg T, Singh AS, Proper KI, Heymans MW, Chinapaw MJ. An evidence-update on the prospective relationship between childhood sedentary behaviour and biomedical health indicators: a systematic review and meta-analysis. *Obes Rev*. 2016;17(9):833-849. doi:10.1111/obr.12426
38. Jiménez-Pavón D, Kelly J, Reilly JJ. Associations between objectively measured habitual physical activity and adiposity in children and adolescents: systematic review. *Int J Pediatr Obes*. 2010;5(1):3-18. doi:10.3109/17477160903067601
39. Monzani A, Ricotti R, Caputo M, et al. A systematic review of the association of skipping breakfast with weight and cardiometabolic risk factors in children and adolescents. What should we better investigate in the future? *Nutrients*. 2019;11(2):387.
40. HBSC Data Management Centre. https://www.uib.no/en/hbscdata/113290/open-access. Updated January 20, 2022. Accessed September 10, 2021.
41. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*. 2000;320(7244):1240-1243. doi:10.1136/bmj.320.7244.1240
42. Homer J, Milstein B, Dietz W, Buchner D, Majestic E. Obesity population dynamics: exploring historical growth and plausible futures in the US. Paper presented at: 24th International System Dynamics Conference, Nijmegen, The Netherlands; 2006.
43. Homer J. Child Stunting and Adult Productivity Loss: A Country-Level Model Applied to India 1980–2080 (2016). In: *More models that matter: System dynamics writings 2011–2017*. Barrytown, NY: Grapeseed Press: 2017.
44. Kobes A, Kretschmer T, Timmerman G, Schreuder P. Interventions aimed at preventing and reducing overweight/obesity among children and adolescents: a meta-synthesis. *Obes Rev*. 2018;19(8):1065-1079. doi:10.1111/obr.12688
45. Doak C, Visscher T, Renders C, Seidell J. The prevention of overweight and obesity in children and adolescents: a review of interventions and programmes. *Obes Rev*. 2006;7(1):111-136. doi:10.1111/j.1467-789X.2006.00234.x
46. Harris JA, Carins JE, Rundle-Thiele S. A systematic review of interventions to increase breakfast consumption: a socio-cognitive perspective. *Public Health Nutr*. 2021;24(11):3253-3268. doi:10.1017/S1368980021000070
47. Loon AWG, Creemers HE, Beumer WY, et al. Can schools reduce adolescent psychological stress? A multilevel meta-analysis of the effectiveness of school-based intervention programs. *J Youth Adolesc*. 2020;49(6):1127-1145. doi:10.1007/s10964-020-01201-5
48. Feiner RD, Brand S, Adan AM, et al. Restructuring the ecology of the school as an approach to prevention during school transitions: longitudinal follow-ups and extensions of the school transitional environment project (STEP). *Prev Hum Serv*. 1994;10(2):103-136. doi:10.1300/J293v10n02_07
49. Hu FB. Resolved: there is sufficient scientific evidence that decreasing sugar-sweetened beverage consumption will reduce the prevalence of obesity and obesity-related diseases. *Obes Rev*. 2013;14(8):606-619. doi:10.1111/obr.12040
50. Bannink R, Pearce A, Hope S. Family income and young adolescents' perceived social position: associations with self-esteem and life satisfaction in the UK Millennium Cohort Study. *Arch Dis Child*. 2016;101(10):917-921. doi:10.1136/archdischild-2015-309651
51. Rutter H, Savona N, Glonti K, et al. The need for a complex systems model of evidence for public health. *Lancet*. 2017;390(10112):2602-2604. doi:10.1016/S0140-6736(17)31267-9
52. Branca F, Chambers T, Sassi F. How to tackle childhood obesity? Evidence and policy implications from a STOP series of systematic reviews. *Obes Rev*. 2021;22(2):e13181.
53. Proctor CL, Linley PA, Maltby J. Youth life satisfaction: a review of the literature. *J Happiness Stud*. 2008;10(5):583-630. doi:10.1007/s10902-008-9110-9
54. Huebner ES, Suldo SM, Smith LC, McKnight CG. Life satisfaction in children and youth: empirical foundations and implications for school psychologists. *Psychol Sch*. 2004;41(1):81-93. doi:10.1002/pits.10140
55. Tomiyama AJ. Stress and obesity. *Annu Rev Psychol*. 2019;70(1):703-718. doi:10.1146/annurev-psych-010418-102936
56. Fismen A-S, Galler M, Klepp K-I, et al. Weight status and mental well-being among adolescents: the mediating role of self-perceived body weight. A cross-national survey. *J Adolesc Health*. 2022;71(2):187-195. doi:10.1016/j.jadohealth.2022.02.010
57. Hirsch G, Homer J, Trogdon J, Wile K, Orenstein D. Using simulation to compare 4 categories of intervention for reducing cardiovascular disease risks. *Am J Public Health*. 2014;104(7):1187-1195. doi:10.2105/AJPH.2013.301816
58. Homer J, Wile L, Yarnoff B, et al. Using simulation to compare established and emerging interventions to reduce cardiovascular disease risk in the United States. *Prev Chronic Dis*. 2014;11:E195. doi:10.5888/pcd11.140130
59. Wang YC, Gortmaker SL, Sobol AM, Kuntz KM. Estimating the energy gap among US children: a counterfactual approach. *Pediatrics*. 2006;118(6):e1721-e1733. doi:10.1542/peds.2006-0682
60. Fallah-Fini S, Rahmandad H, Huang TT-K, Bures RM, Glass TA. Modeling US adult obesity trends: a system dynamics model for

estimating energy imbalance gap. *Am J Public Health*. 2014;104(7):1230-1239. doi:10.2105/AJPH.2014.301882
61. Huang TT, Drewnowski A, Kumanyika SK, Glass TA. A systems-oriented multilevel framework for addressing obesity in the 21st century. *Prev Chronic Dis*. 2009;6(3):A82.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Romanenko E, Homer J, Fismen A-S, Rutter H, Lien N. Assessing policies to reduce adolescent overweight and obesity: Insights from a system dynamics model using data from the Health Behavior in School-Aged Children study. *Obesity Reviews*. 2022;e13519. doi:10.1111/obr.13519

Appendix 2

The manuscript of journal article entitled "As Simple as Possible but not Simpler: structural sensitivity testing of a dynamic model of adolescent overweight and obesity" (submitted to System Dynamics Review on December 13 2022) is attached below.

Title: As Simple as Possible but not Simpler: structural sensitivity testing of a dynamic model of adolescent overweight and obesity

Running title: Structural sensitivity testing of a model of adolescent obesity

Keywords:
Structural sensitivity testing, Evidence-based modeling, Model adequacy and parsimony, Adolescent behavior and psychology, Social determinants of health

AUTHORS
Eduard Romanenko[1]
Jack Homer[2]
Nanna Lien[1]

[1]Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
[2]Homer Consulting and MIT Research Affiliate, Barrytown, NY, USA

CORRESPONDENCE
Eduard Romanenko
Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, PO Box 1046, Blindern, N-0316, Oslo, Norway
E-mail: eduard.romanenko@medisin.uio.no

FUNDING INFORMATION
The CO-CREATE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 774210. The content of this article reflects only the authors' views and the European Commission is not liable for any use that may be made of the information it contains.

Abstract

We recently published results from an SD model of adolescent overweight and obesity using data from 31 European countries that participate in the Health Behavior in School-aged Children (HBSC) study. During model development, we sought to identify a feedback structure with high explanatory power but avoiding speculative relationships. Expert reviewers generally agreed with our modeling decisions, but two decisions did raise questions: (1) excluding the influences of food environment and built environment, for which HBSC provided no data; and (2) including five causal links that were supported statistically but might be considered disputable. To address the reviewers' questions, we created four possible model structures and performed automated calibration followed by intervention testing and ranking. We then compared the goodness of fit and intervention results. We discuss implications for how to move forward with the model, including through additional data gathering.

Introduction

We may assume the superiority, all other things being equal, of the demonstration which derives from fewer postulates or hypotheses.

– Aristotle (384-322 BC)

Plurality should not be posited without necessity…It is futile to do with more things that which can be done with fewer.

– William of Ockham (c. 1287-1347)

It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience…Everything should be as simple as possible but not simpler.

– Albert Einstein (from "On the Method of Theoretical Physics," the Herbert Spencer Lecture, Oxford, June 10, 1933; and attributed to Einstein, New York Times, January 8, 1950)

Consistent with descriptions of the scientific enterprise going back to Aristotle and culminating with Einstein, system dynamics (SD) seeks to develop models that are adequate in complexity for addressing the problem at hand, but also parsimonious enough to be comprehensible and well supported by the available evidence (Homer 2014). Models often evolve from simpler to more complex in order to produce outputs that are more realistic or speak to particular policy concerns (see, e.g., Randers 1973, Alfeld and Graham 1976, Homer 1996, Sastry 1997). Yet one must also be careful not to clutter a model with excessive detail that undermines its clarity and explanatory power (Forrester 1961).

Sensitivity testing is one of our most important tools, not only for model analysis, but also for model improvement. Parametric sensitivity testing helps us understand the behavioral and policy implications of parameters of uncertain value, while structural sensitivity testing helps us understand the implications of variables or causal links of uncertain importance. Structural analysis includes both "boundary adequacy" and "structure assessment" testing (Sterman 2000) and allows us "to evaluate the impact of controversial or disputable relationships" (Tank-Nielsen 1980).

Such testing can help us decide what is essential to include in a model and how to proceed in gathering more evidence. If a variable or causal link lacks strong evidence (that is, weaker than the rest of the model) and does not affect policy findings, then one may consider excluding it (see, e.g., Mahamoud et al 2012). However, if such an uncertain variable or link does affect policy findings, one may include it conditionally; namely, on the condition that more evidence on it will be sought.

We recently completed the first phase of an SD study of adolescent overweight and obesity (AdOWOB) in Europe, based on survey data from the Health Behavior in School-Aged Children study (HBSC) from 31 countries and with particular emphasis on the five countries involved in the EU-funded CO-CREATE project (Romanenko et al 2022). In this work, we utilized a combination of literature review, statistical screening procedures, and SD modeling to build a strongly evidence-based model with only 12 major variables (8 of them endogenous and 4 exogenous) and 30 causal links (with corresponding strengths known as hazard ratios). Automated calibration showed that the model could nicely reproduce HBSC data patterns from 24 different cases (differing by country, gender, and perceived wealth status) over the period 2002-2014. For each case, we tested 10 potential points of intervention (starting in 2018) and ranked them by projected reduction of AdOWOB by 2026. We used our model-based findings to support or supplement the policies suggested by the adolescent participants who were part of CO-CREATE.

Our objective in model development was to identify a cluster of interrelated variables that demonstrated high explanatory power but was parsimonious with respect to available data; that is, a model that avoided speculative relationships. This approach had implications for which variables and causal links we did or did not include in the model. Public health experts involved in internal review of the model during the project generally agreed with our decisions, but two decisions did raise some questions among some experts.

The first of those decision was to exclude the food environment (FE, affecting dietary behaviors) and the built environment (BE, affecting physical activity). The literature points to the potential significance of FEBE as a factor affecting adolescent obesity (Elbel et al 2020, Malacarne et al 2022, Gilliland et al 2012), but neither HBSC nor any other multi-country European survey to date includes questions related to FEBE. We excluded FEBE because we had no data, not even proxies or trend data, to estimate it.

The second decision that raised questions was the inclusion of five causal links that were supported by statistical screening but were deemed "indirect", meaning that their support from the literature required assumptions about an unmeasured intermediate variable, specifically high-calorie snacking. For example, the statistical screening suggested a link from nervousness to AdOWOB, which required explanation in two steps: from nervousness to snacking, and from snacking to AdOWOB. Our model includes several other dietary behavior variables, but it does not include snacking. Optional snacking questions were part of the HBSC survey, but only data from the mandatory questions of the survey were available to us through open access. We described snacking to the experts as a hidden variable in the model, kept implicit for lack of data, and we made the point that all models include implicit variables (see Alfeld and Graham 1976); but still some experts questioned this approach.

Reflecting on our first phase of modeling, we realized that we might use structural sensitivity testing to address the experts' questions. First, perhaps we could find a way to infer trends in FEBE despite the lack of direct data on it. Might the inclusion of such trends affect our policy conclusions? Second, what would happen if we eliminated the five “indirect” causal links? Might such elimination affect our policy conclusions?

In this paper we describe this two-fold structural sensitivity analysis of the existing model and its implications for future data needs.

Structural testing procedures

Alternative model structures

We started by making two types of modifications to the original model structure. One was to incorporate the concept of FEBE through three assumed linear trends (switchable on or off) affecting the variables of inadequate exercise, inadequate fruit, and inadequate vegetables. Despite the lack of data on FEBE, we reasoned that if (a) the inclusion of such trends (after optimized calibration) allowed for a better overall goodness of fit, and (b) they ended up altering the policy conclusions, then we could justify the inclusion of these trends in the model. Each linear effect was formulated as a ramp starting in 2002 with two parameters to be optimized: End Year and End Change (that is, the ramp’s percentage change from 2002 to End Year).

Another modification was to allow reduction of the model by selectively (switchable on or off) excluding the five “indirect” causal links that implicitly go through high-calorie snacking. Two of these links (from School Pressure and Feel Nervous) bypass behavioral variables on the way to AdOWOB. The other three links capture the effect of environmental variables (School Pressure and Life Dissatisfaction) on fruit and vegetables consumption. If we found that excluding these “disputable” links (to use the Tank-Nielsen term) did not alter the model's policy conclusions, then, by the logic of parsimony, we might safely eliminate them from the model. If, on the other hand, they did alter policy conclusions, then we would lean toward the original model, but on the condition that we could find more evidence to support the disputable links.

The structural sensitivity analysis was performed through the testing of four possible model configurations: (1) the original model including the five disputable links but excluding FEBE

(“Full_noFEBE”); (2) a model including both the five links and FEBE (“Full_FEBE”); (3) a model excluding the five links as well as FEBE (“Reduced_noFEBE”); and (4) a model excluding the five links but including FEBE (“Reduced_FEBE”).

Figure 1 is an interpretive sector diagram of the original model (with both explicit and implicit links), showing the five disputable links going through the implicit (unmodeled) variable of high-calorie snacking. **Figure 2** extends this diagram to include the possible FEBE influences.

Model calibration and testing

For each model configuration, we repeated the analysis we had done in our original study (Romanenko et al 2022). First, we used Powell optimization to calibrate each of the alternative structures to the HBSC data for each of our 24 country-gender-wealth cases. Next, we calculated two types of goodness-of-fit statistics for the cases, for all eight of the model's endogenous variables: (1) the mean absolute percentage error (MAPE) between simulated output and data and (2) a customized R-squared measure (“R2i”, range 0 to 1) of how well the model predicts changes away from the initial data point in 2002.

We recorded these statistics for AdOWOB (the main variable of interest in the model and the ultimate target for intervention testing), as well as averaged across all eight endogenous variables (hereafter “All8”, of which AdOWOB is one). This resulted in four goodness-of-fit measures (AdOWOB MAPE, All8 MAPE, AdOWOB R2i, and All8 R2i) for each model and each case. To facilitate comparison between the four models, we transformed MAPE into a 0-1 index (“MAPE index”) in which 0% MAPE is 1 (best) and ≥20% MAPE is 0 (worst). We averaged the two MAPE indices with the two R2i measures, weighting all four equally, to produce a combined index of model adequacy. We did this for each case and then averaged across all 24 cases for an overall average of model adequacy.

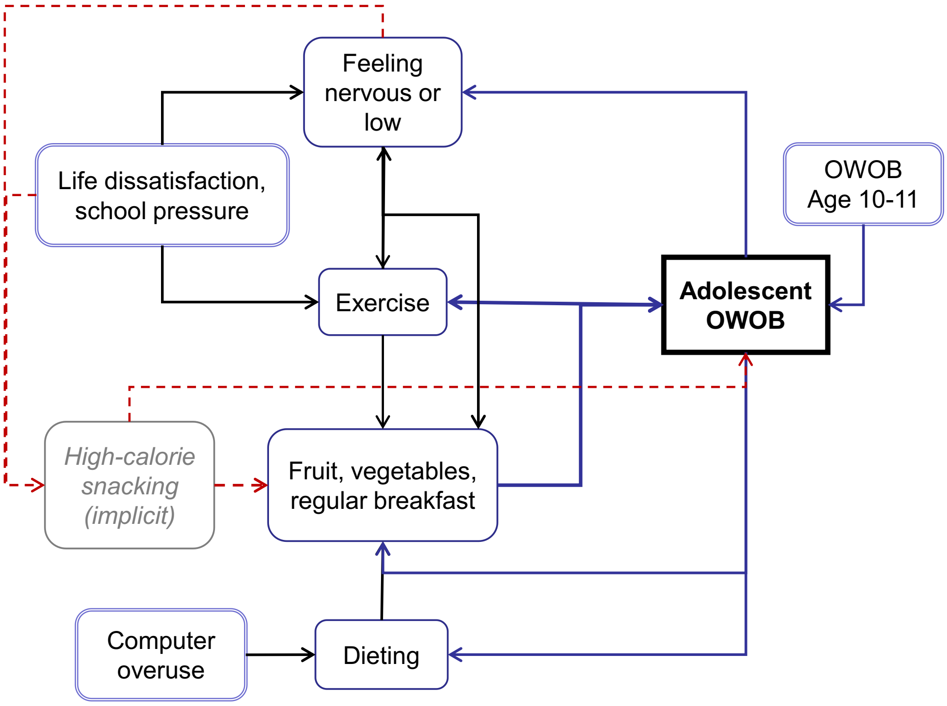

Figure 1. Sector diagram of the original model, showing disputable links (dashed red) going through the implicit (unmodeled) variable of high-calorie snacking. (A standard causal-loop diagram, absent snacking, is presented in Romanenko et al 2022.)

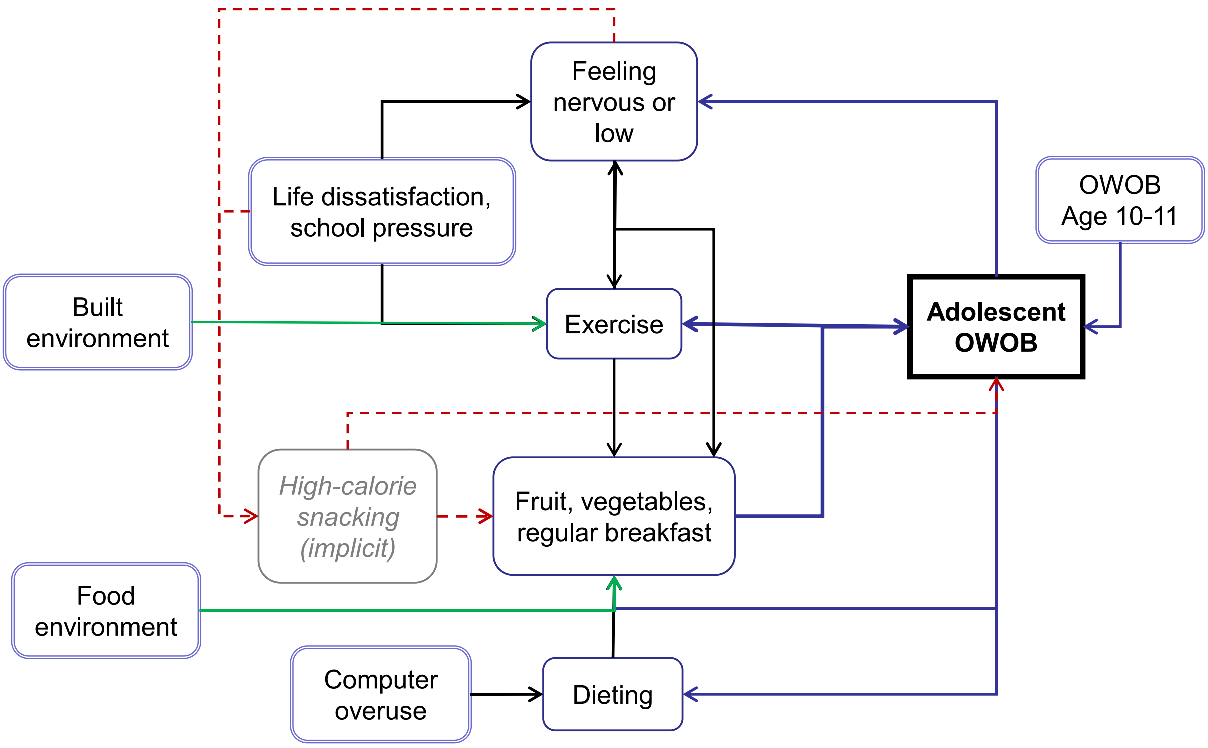

Figure 2. Extended sector diagram, including possible Food Environment and Built Environment influences (solid green)

Next came intervention testing, again done exactly as before (Romanenko et al 2022). We tested the 10 potential intervention points using identical 25% effect sizes starting in 2018, and we ranked the interventions in terms of their ability to reduce projected AdOWOB in 2026. To facilitate comparison of the models, we counted the number of times each intervention point appeared in the top 4 ranking of 10 interventions, across all 24 cases.

Results

Optimized ramp change parameters for FEBE

The optimization of the *FEBE* models (see **Supporting Information** for optimization specifications) resulted in all ramps (for both *Full_FEBE* and *Reduced_FEBE* and for all 24 cases) having an End Year close to 2010. **Table 1** reports the summary statistics for the estimates of the ramp end-change percentages across the 24 cases.

Table 1: Summary statistics for the end-change percentages for the two models with food and built environment (FEBE) influences optimized for the 24 cases. These models allow for three exogenous ramps (each with its own optimized end time and end-change percentage) affecting the prevalence fractions of Inadequate Exercise, Inadequate Fruit, and Inadequate Vegetables, respectively. A positive end-change percentage indicates a worsening trend (more inadequacy), while a negative end-change percentage indicates an improving trend (less inadequacy).

	Full_FEBE			**Reduced_FEBE**		
	Inad Ex	**Inad Fruit**	**Inad Veg**	**Inad Ex**	**Inad Fruit**	**Inad Veg**
Mean	0.2%	1.0%	-1.4%	-0.3%	0.9%	-3.3%
Std. Dev.	2.2%	5.0%	3.2%	2.7%	6.1%	5.2%
Min	-5.4%	-11.9%	-14.5%	-7.7%	-14.1%	-19.2%
Max	5.5%	12.9%	2.7%	4.1%	13.9%	4.5%

The table shows that the optimized end-change percentages are generally of modest size. The parameters for Exercise are always less than 8% in either direction, with means of about zero for both *Full_FEBE* and *Reduced_FEBE*. The parameters for Fruit are always less than 15% in either direction, with means of about +1% (worsening trend) for both models. The parameters for Vegetables are always less than 19% in either direction, with a mean of -1% for *Full_FEBE* and -3% for *Reduced_FEBE* (both improving trends).

Goodness of fit

Table 2 summarizes goodness-of-fit statistics for the four tested model configurations, with each measure averaged across the 24 cases. The individual fit statistics (the first four rows) do not vary by much from one model to another. The combined adequacy measure (the last row) is similarly tight, with the largest model, *Full_FEBE* (at 61.2%) providing only a slightly better fit than the smallest model, *Reduced_noFEBE* (at 59.0%).

Table 2: Goodness-of-fit statistics for the four tested model versions, averaged across the 24 cases. MAPE is Mean Absolute Percentage Error; R2i is a novel R-squared measure (see Romanenko et al 2022), All8 refers to all eight endogenous variables.

	Model			
Fit Statistic	**Full_noFEBE**	**Full_FEBE**	**Reduced_noFEBE**	**Reduced_FEBE**
MAPE index - AdOWOB	53.7%	54.3%	52.0%	51.7%
MAPE index - All8	62.3%	63.3%	60.7%	62.3%
R2i - AdOWOB	66.6%	66.9%	66.3%	66.3%
R2i - All8	58.0%	60.5%	57.0%	59.7%
Combined Adequacy (equal weighting)	*60.1%*	*61.2%*	*59.0%*	*60.0%*

Although the differences are not great, this table does provide some information about the types of contribution coming from (a) the five disputable links in the *Full* models, and (b) the three new ramp effects in the *FEBE* models.

The clearest benefit of *Full* is in the two MAPE indices (rows 1 and 2), where *Full* beats *Reduced* by 1.0% to 2.6%. This fact suggests that the five disputable links (four of which come from the exogenous variables of life dissatisfaction and school pressure) give the model a greater ability to follow turning points in the data. It must be that some of the ups and downs in these exogenous variables help to explain corresponding ups and downs in AdOWOB and the other endogenous variables.

The clearest benefit of *FEBE* is in the “All8” fit statistics (rows 2 and 4), where *FEBE* beats *NoFEBE* by 1.0% to 2.7%. The addition of the exogenous ramps for exercise, fruit, and vegetables improves the model’s fit to those variables, but it does not improve the fit to AdOWOB.

Intervention testing

Table 3 reports, for each of the model configurations, the percentage reductions in AdOWOB (in 2026 relative to no intervention) averaged across the 24 cases for each of the 10 interventions separately and for all 10 combined. The interventions vary greatly in terms of their impact on AdOWOB, even when averaged across the cases. For all model configurations, the most impactful interventions include Fruit, Exercise, Breakfast, Life Dissatisfaction, and Vegetables (in roughly that order). Yet, there are also differences between the models. First, for the *Full* models (unlike the *Reduced* models), Feel Nervous and School Pressure are additional interventions with good impact. These two variables account for three of the five disputable links included in the *Full models.* Second, the inclusion of the three *FEBE* ramps tends to boost the impact of the leading interventions in combination with *Full* (i.e., *Full_FEBE*), but it does not do so in combination with *Reduced* (i.e., *Reduced_FEBE*).

Table 4 reports the overall counts of “top 4” ranking for each of the 10 interventions for the four model configurations, summing across the 24 cases. It is evident that including *FEBE* has no real effect on intervention priorities, whether the starting point is the *Full* or *Reduced* models. In contrast, the inclusion of the five disputable links in the *Full* models does clearly affect the intervention rankings. The *Full* models elevate the rankings of School Pressure and Life Dissatisfaction (and also, somewhat, Feel Nervous), and they demote the ranking of Vegetables (and also, somewhat, Breakfast).

Table 3: Percentage reduction of OWOB in 2026 for the 10 interventions, for the four tested model versions, averaged across the 24 cases

	Model			
Intervention	**Full_noFEBE**	**Full_FEBE**	**Reduced_noFEBE**	**Reduced_FEBE**
Inad Exercise	3.6%	5.1%	4.4%	3.9%
Inad Fruit	5.0%	5.2%	4.7%	4.6%
Inad Veg	1.6%	2.0%	1.9%	1.6%
Inad Breakfast	1.8%	2.5%	2.1%	2.1%
Dieting	0.3%	0.5%	0.5%	0.5%
Feel Nervous	2.5%	3.2%	0.9%	1.0%
Feel Low	1.7%	1.6%	0.9%	0.9%
Life Dissatisfaction	2.4%	3.8%	2.0%	1.6%
School Pressure	3.7%	3.7%	0.8%	0.6%
Computer Overuse	0.2%	0.2%	0.2%	0.1%
All 10 Combined	*13.5%*	*16.4%*	*12.6%*	*12.3%*

Table 4: Counts of Top 4 ranking for the 10 interventions, for the four tested model versions, across the 24 cases (maximum count of 24). Note that, for a given model, some cases may have fewer than four ranked interventions (i.e., interventions with any simulated impact on AdOWOB), and some cases may have more than four "Top 4" ranked interventions (due to ties in percentage impact to the first decimal point). As a result, columns in this table do not sum to 96 (=24 x 4).

	Model			
Intervention	**Full_noFEBE**	**Full_FEBE**	**Reduced_noFEBE**	**Reduced_FEBE**
Inad Exercise	18	19	20	21
Inad Fruit	17	17	18	18
Inad Veg	8	9	12	13
Inad Breakfast	10	12	12	15
Dieting	1	0	1	0
Feel Nervous	6	9	4	5
Feel Low	3	4	3	6
Life Dissatisfaction	16	19	11	10
School Pressure	11	13	3	4
Computer Overuse	0	0	0	0

Discussion

We used structural sensitivity testing to evaluate two decisions made in the development of the AdOWOB model: (1) excluding the influences of FEBE, for which we had no data; and (2) including five causal links that were supported statistically but which some public health experts considered indirect and disputable.

Our analysis showed, first, that exogenous linear trends representing FEBE and affecting exercise and fruit and vegetable consumption could improve the model's fit to those variables but did not improve the fit to AdOWOB itself and had no effect on intervention priorities. This is not to say that we do not recognize the importance of FEBE as a type of intervention in its own right. On the contrary, exercise, fruit, and vegetables are all important intervention points in our model and, in the real world, may be influenced by FEBE interventions. However, our results showed that allowing for possible historical trends in FEBE did not alter optimized hazard ratios enough to change intervention rankings. Therefore, we concluded that FEBE trends did not add value to the original model and, according to the logic of evidence and parsimony, could be safely excluded.

Our testing also showed that the inclusion of the five disputable causal links in the *Full* model configurations provided a somewhat better fit to all variables including AdOWOB (by the MAPE criterion) and affected the intervention rankings, elevating the priority of school pressure and life dissatisfaction. This policy sensitivity suggests that one should be cautious about eliminating the links in question and rather lean towards the original model, on the condition that more evidence to support these links could be found.

The five links were identified as disputable because they all go through high-calorie snacking, an intermediate variable which is not included in the mandatory questions of the HBSC study. A logical direction for gathering further supporting evidence is to collect data on individual snacking or daily caloric intake. Obtaining the data on snacking from optional HBSC questions (asked by a subset of countries) could be useful. However, the HBSC measures only frequency of snacking and not the type or amount. Better data could be obtained, for example, by applying diet checklists for snacking behavior over a sufficiently long time period, probably two weeks or more as described by the DAPA Measurement Toolkit (https://www.dapa-toolkit.mrc.ac.uk/diet/subjective-methods/diet-checklist). Data from even just a few countries, perhaps some of those in the five-country CO-CREATE project, could help determine whether the disputable links can be supported by direct evidence.

We believe that our work here could contribute to SD modeling practice, in at least two ways.

First, structural sensitivity testing has long been described as an important part of building confidence in SD models, yet the literature gives little guidance on how to do it. Here, we have demonstrated how one may compare alternative models based on their goodness of fit and their effect on policy conclusions[1].

Second, the existing literature gives little guidance on how to balance the competing values of model adequacy and parsimony. Here, we have offered the following approach:
(1) evaluate the strength of the evidence for a causal link in question based on the literature, expert knowledge, and available data;
(2) for a causal link with weaker evidence (an uncertain or disputable link), evaluate whether including the link improves the model's explanatory power or affects policy findings;
(3) eliminate the uncertain link if it does not add value to the model;
(4) if the uncertain link does add value to the model, include it on the condition that more evidence will be sought to confirm or reject the link.

We believe that structural sensitivity testing is an important tool that could allow SD modelers to be more scientific and show that their models are "as simple as possible but not simpler".

BIOGRAPHIES

Eduard Romanenko is a researcher in the Public Health Nutrition group at the University of Oslo, Norway. He has been working as an SD modeler for the CO-CREATE project

---

[1] We recognize there is one aspect of structural sensitivity testing we did not demonstrate; namely, determining whether a model can produce all relevant or problematic modes of behavior, such as oscillation or overshoot-and-decline. The only behavioral pattern we saw clearly in the HBSC data, across the 24 cases, was AdOWOB adjusting in a goal-seeking (decelerating) fashion to perturbations in other variables. This behavior pattern was produced by all four of the alternative model configurations we considered.

described here since 2020. He graduated from the European Master program in SD in 2014 and received a PhD in SD from the University of Bergen, Norway in 2022.

Jack Homer is a system dynamics modeling consultant and has directed Homer Consulting since 1989. Before that, he taught at the University of Southern California and received a PhD from MIT, and BS and MS degrees from Stanford University. He received the SD Society's Forrester Award in 1997 for his work studying cocaine use in the US, and was lead modeler for the team that won the Applications Award in 2011 for the "PRISM" model of cardiovascular disease for the US Centers for Disease Control and Prevention. He is author of the books "Models That Matter" (2012) and "More Models That Matter" (2017).

Nanna Lien is a professor at the Department of Nutrition, University of Oslo, Norway and leads the research group in Public Health Nutrition. She has more than 15 years of experience in school-based intervention research in nutrition and obesity prevention in Norway and on European funded projects. In the CO-CREATE project she leads the work package on "Evaluation of Co-Created policy interventions and the methodology". She is a fellow of the International Society of Behavioral Nutrition and Physical Activity (ISBNPA) and a deputy editor of the International Journal of Behavioral Nutrition and Physical Activity (IJBNPA).

References

Alfeld LE, Graham AK (1976). *Introduction to Urban Dynamics*. Cambridge, MA: MIT Press. (Preface, pp. xiii-xv.)

Elbel B, Tamura K, McDermott ZT, Wu E, Schwartz AE (2020). Childhood obesity and the food environment: a population-based sample of public school children in New York City. *Obesity*; 28(1): 65–72.

Forrester JW (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press. (Appendix O, "Beginners' Difficulties", pp. 449-456.)

Gilliland JA, Rangel CY, Healy MA, Tucker P, Loebach JE, Hess PM, He M, Irwin JD, Wilk P (2012). Linking childhood obesity to the built environment: a multi-level analysis of home and school neighbourhood factors associated with body mass index. *Canadian Journal of Public Health*; 103(9 Suppl 3):eS15-21.

Homer JB (1996). Why we iterate: scientific modeling in theory and practice. *System Dynamics Review* 12(1):1-19.

Homer J (2014). Levels of evidence in system dynamics modeling. *System Dynamics Review* 30:75-80.

Mahamoud A, Roche B, Homer J (2012). Modelling the social determinants of health and simulating short-term and long-term intervention impacts for the city of Toronto, Canada. *Social Science & Medicine* 93:247-255.

Malacarne D, Handakas E, Robinson O, Pineda E, Saez M, Chatzi L, Fecht D (2022). The built environment as determinant of childhood obesity: A systematic literature review. *Obesity Reviews*; 23(S1):e13385.

Randers J (1973). Conceptualizing dynamic models of social systems: lessons from a study of social change. PhD dissertation, MIT Sloan School of Management, Cambridge, MA.

Romanenko E, Homer J, Fismen A-S, Rutter H, Lien N (2022). Assessing policies to reduce adolescent overweight and obesity: insights from a system dynamics model using data from the Health Behavior in School-Aged Children study. *Obesity Reviews* e13519.

Sastry MA (1997). Problems and paradoxes in a model of punctuated organizational change. *Admin Science Quarterly* 42:237-275.

Sterman JD (2000). *Business dynamics: systems thinking and modeling for a complex world*. Boston, MA: Irwin McGraw-Hill. (Ch. 21, "Truth and beauty: validation and model testing", pp. 845-891.)

Tank-Nielsen C (1980). Sensitivity analysis in system dynamics. Chapter 9 (pp. 185-202) in *Elements of the System Dynamics Method*, ed. J Randers. Cambridge, MA: MIT Press.

The CO-CREATE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 774210. The products of the research are the responsibility of the authors: the European Commission is not responsible for any use that may be made of them.