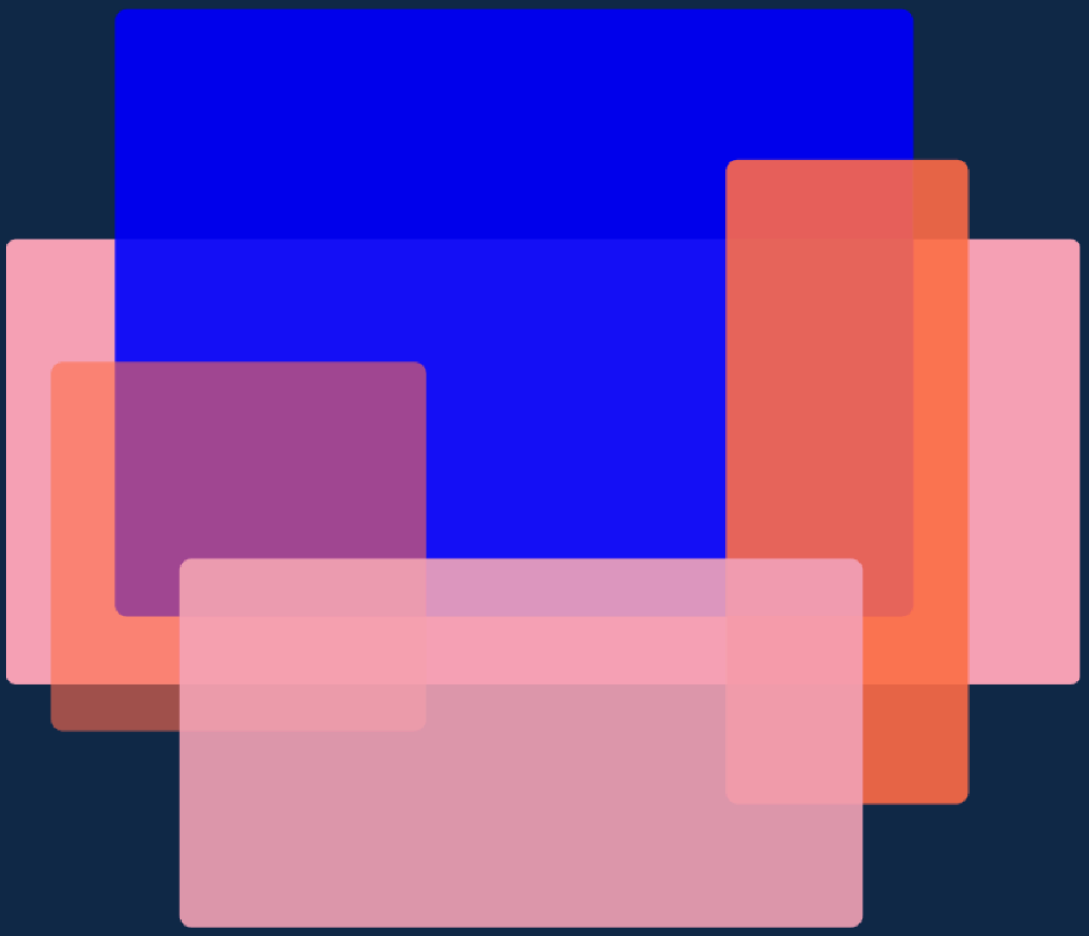
Blueprint
by nesta

Blueprint for halving obesity: technical appendix



Authors: Anish Chacko, Kate Tudor, Katherine Parkin, Partricia Beloe and Hugo Harper

Blueprint
by nesta

Contents

Background

About the project

This technical appendix accompanies the blueprint for halving obesity toolkit, with a focus on the analytical approach used to model the impact and cost effectiveness of the policies that could be implemented to reduce national obesity rates. The analysis was conducted by Nesta and HealthLumen, and incorporates findings previously published by Frontier Economics. The programme has been supported by an Expert Advisory Group (EAG) of leading experts in the field of obesity and health. Each policy is scored across three metrics: **impact, cost to implement**, and **strength of evidence (SoE).** Presentation of a score for each of these metrics is what distinguishes the toolkit from existing reports – it provides an opportunity to directly compare each policy. We intend to add to this list in future updates of this toolkit and so encourage input from policymakers and academics to build upon this list as the evidence base continues to grow.

Summary of the project methodology

The toolkit presents findings against three metrics. This section provides a high-level summary of what each metric represents, why we chose this metric, and how we calculated the score for each metric. The subsequent sections go into more methodological detail.

Impact scores

The impact scores reflect how much the policy reduces the percentage of people in Great Britain living with a BMI score of 30 and above after five years. This is a relative reduction, not an absolute reduction. We chose this metric because obesity status is a strong predictor of negative health outcomes and so it is a good measure of whether a policy will be beneficial for public health.

Figure 1. Screenshot of policy impact represented in the toolkit

To calculate this score, for each policy we conducted rapid evidence reviews to establish the effect of the policy on obesity-related outcomes (eg, change in weight, BMI, self-reported physical activity, calories purchased, calories consumed etc.) In order to compare the impact of each policy against the same benchmark, we used Health Survey for England (2019) and Scottish Health Survey (2019) data, to model the extent to which the policy would reduce the population prevalence of obesity in the UK. This resulted in a ranked list of the policies, which we tested with our EAG and a further group of 40 experts via a consensus building survey (see Building consensus for policy impacts section below for more detail).

See table 1 for a breakdown of how impact score relates to changes in the prevalence of obesity.

Table 1. Breakdown of how impact rating relates to the impact of each policy on the relative reduction of population obesity prevalence

Relative reduction in the prevalence of people living with obesity	How it appears on the site
15%+	Very high
5-14.9%	High
0.5-4.9%	Moderate
0.1-0.49%	Low
0%	Very low

Cost scores

The cost scores reflect the absolute amount of money required by governments to deliver the policy over a five-year period. Cost is a key driver of feasibility of a policy being enacted and so is important for policymakers to know.

Figure 2. Screenshot of policy cost represented in the toolkit

To calculate this score we commissioned the cost-modelling consultancy HealthLumen to conduct literature reviews to establish the cost to implement each policy. The final score out of five is based on where the policy ranked in terms of cost. The most expensive policies were allocated a score of 'very high', and the least costly a score of 'very low' (and so on). You can find more information about the data sources used to estimate the costs in HealthLumen's report, on the Methods page of the Blueprint for Halving Obesity toolkit.

Table 2. Breakdown of how cost rating relates to the estimated cost to governments over a five-year period

Cost to governments	How it appears on the site
£0-0.9m	Very low
£1m-£24.9	Low
£25m-£499m	Moderate
£500m-£999m	High
£1bn+	Very high

Strength of Evidence (SoE) scores

The evidence strength scores reflect a rating of how valid and reliable the evidence is to back up the impact of each policy (ie, the extent to which the policy has been tested scientifically). The rating was conducted by our EAG during a workshop hosted by the Blueprint for Halving Obesity team at Nesta. The method of using a panel of experts is commonly used to establish consensus regarding the strength of scientific evidence.

Figure 3. Screenshot of the SoE for each policy represented in the toolkit

To assign scores, an SoE scale was created that assessed reliability and validity of evidence. Existing scales were not used as real-world implementation of interventions was valued on a similar level with trials and lab-based research. Policies were scored on a scale from 1-5 (no evidence, to very strong evidence).

Table 3. Breakdown of how SoE ratings relate to the estimated cost to governments over a five-year period

Strength of evidence	How it appears on the site
Very strong	Very high
Strong	High
Medium	Moderate
Limited	Low
No evidence	Very low

Calculating the impact rating

Synthesising the evidence

Identifying policy levers

We reviewed existing reports that made recommendations for action that could reduce the prevalence of UK obesity [1], [2], [3], [4] which resulted in a list of over 150 potential policy levers. We categorised these into levers that could be enacted across the food, health, and education system.

A list of 150 directionally positive policy levers is not practical for policymakers. Being informed by similar toolkits (eg, Climate Toolkit, Teaching and Learning Toolkit), we estimated about 30 policy levers would be an appropriate amount, striking the balance between being comprehensive and user friendly. To reduce the 150 identified policy levers to approximately 30, we removed duplicates extracted from reports. Next, where reports suggested multiple different versions of a similar policy lever, we chose one or two levers to represent this category of intervention. For example, there were over 10 suggestions for how the governments could use fiscal measures to promote reformulation or reduced portion sizes. We chose two options to represent fiscal measures in the toolkit. We chose these based on (a) the extent of the evidence already available on the impact of the lever and (b) how central the levers are to the debate around solutions for obesity. Similarly, there were over 10 versions of levers that would increase the provision of nutritional information on food and drink products and we chose a single lever to reflect this category of intervention. Hence, our final list of policies reflected the 150 policies that were extracted from existing reports. To select the most appropriate policy lever, we held internal workshops with expert cross-sector policy knowledge. We also consulted stakeholders with policymaking experience to inform the chosen policy levers.

Rapid evidence reviews

To establish the per-person effect of each policy lever the team at Nesta conducted rapid evidence reviews. The aim was to identify the highest quality and up-to-date meta-analysis or narrative synthesis of the impact of the policy in question. Where

such evidence was not identified, we searched for high quality individual studies, particularly for those intervention subcategories that either didn't have a relevant review or had an outdated review. We searched grey literature and policy databases and aimed to include impact assessments and evaluations of similar policies implemented in other high income, and comparable countries.

Our rapid review search specified to identify reviews and studies with obesity-related outcomes. Obesity-related outcomes included:

- change in body composition/weight of an individual (BMI, body fat percentage, waist circumference, weight)
- change in obesity prevalence in a population (eg, change in the proportion of the sample with BMI≥30)
- change in levels of physical activity
- change in dietary intake (ie, kilocalories/kilojoules consumed/purchased over a specific time period).

We refer to the outcomes expressed in one of the above mentioned variables as the **effect size** or **magnitude of effect** of an intervention. We excluded articles that did not report one of these outcomes as this information was required to model the impact against our benchmark of interest: **change in UK obesity prevalence over five years.**

Modelling the impact

About the datasets

The core datasets used to model the impact of a policy are datasets produced as part of the health surveys in England and Scotland in 2019: Health Survey for England 2019 (referred to as HSE 2019, hereafter) and the Scottish Health Survey 2019 (referred to as SHeS 2019, hereafter). In both the datasets, we subset the variables of interest that we use across different policies – age, sex, body weight, height, sex, body mass index (BMI), Quintiles of Index of Multiple Deprivation score 2019 (QIMD), number of children, Jobseeker's Allowance, Income Support, Pension Credit, Child Tax Credit, Universal Credit, ethnicity, diabetes, cardiovascular disease. Subsequently, we filter to drop rows where body weight or height values are less than zero. After both these operations, we do not lose any of the rows in the dataset.

Unlike SHeS 2019, in HSE 2019, individual ages are not available and are recorded in three-, four- or five- year bands. We use the variable with the smallest band width – 'Age35g'. Based on the age, we split the dataset for those over 18 for adult modelling and those under for child modelling. Using this threshold in the HSE 2019 dataset was not direct as all 16-19 year olds were in the same band. And since the majority (16-18 years) of those in this age band are under 18 years, we include all data points in the scope of child modelling. The baseline obesity prevalence after this operation that we measure against is 29% for adults and 20% for children.

Table 4 below shows the respective variables from the two surveys.

Table 4. Respective variables from HSE 2019 and SHeS 2019

Description	HSE 2019 variable	SHeS 2019 variable
Age	Age35g	Age
Body weight	WtVal	wtval
Height	HtVal	htval
Sex	Sex	Sex
BMI value	BMIVal	bmival
BMI category	BMIVG5	bmi
IMD quintile	qimd19	SIMD20_SGa
Number of children	Nofch3	-
Income – Jobseeker's Allowance	srcin05d	SrcInc8a
Income – Income Support	srcin07d	SrcInc9a
Income – Pension Credit	srcin08d	SrcInc5a
Income – Child Tax Credit	srcin10d	SrcInc6a, SrcInc10a
Income – Universal Credit	srcin14d	SrcInc7a
BMI category of parent	fath_bmi2 for father moth_bmi2 for mother	-

Ethnicity	origin2	Ethnic05
Diabetes status	diabete2	diabete2
Cardiovascular disease status	CardioTakg2	medtyp1B
Identity	SerialA	CPSerialA
Sampling weight	wt_int	int19wt for adults cint19wt for children
Primary sampling unit	PSU_SCR	PSU
Strata	cluster94	Strata

After this, we create new variables to calculate the basal metabolic rate and the total energy expenditure. The total energy expenditure is calculated as the basal metabolic rate (BMR) times physical activity levels (PAL).

We estimate the baseline basal metabolic rate for individuals in the health survey datasets using the following equations:

- for adults we used Mifflin St Jeor equations
- for children we used Henry (1999) equations.

PAL is a factor determined by the frequency, duration and intensity of physical activity by an individual. It varies between adult and child populations. For our population level calculations for adults, we assume it to be 1.6 as per the Dietary Reference Values for Energy (SACN, 2011) report and it describes an individual doing very light activity at school or work and a moderate physical activity (walking or cycling) once a week.

For children, we take age-based PAL values described by the Scientific Advisory Committee on Nutrition (SACN) in its report – Dietary Reference Values for Energy (SACN, 2011). The median values of PAL adjusted for growth are taken based on the child's age. For ages 1-< 3, we take 1.40; 3-<10, 1.58; 10-18, 1.75 (please see table 7, SACN, 2011).

Then using the following equation we estimate the total energy expenditure (TEE) for an individual:

$$TEE = RMR \times PAL$$

At the baseline, we assume that people are maintaining their body weight, therefore, their energy intake equals the TEE for physical activities and for metabolic homeostasis.

For more details and scripts please see the Github repo.

Note:

1. All data preparation, modelling and analysis has been done in R version 4.4.0 (2024-04-24 ucrt).
2. Both datasets – SHeS 2019 and HSE 2019 are downloaded from the UK Data Service portal.

Modelling the impact on a single outcome

We modelled the impact of implementing each policy across the population that it would reach, in order to estimate overarching contribution to reducing obesity prevalence. The outcome is a quantitative estimate of the population level effect that each intervention has on obesity prevalence in the UK, eg:

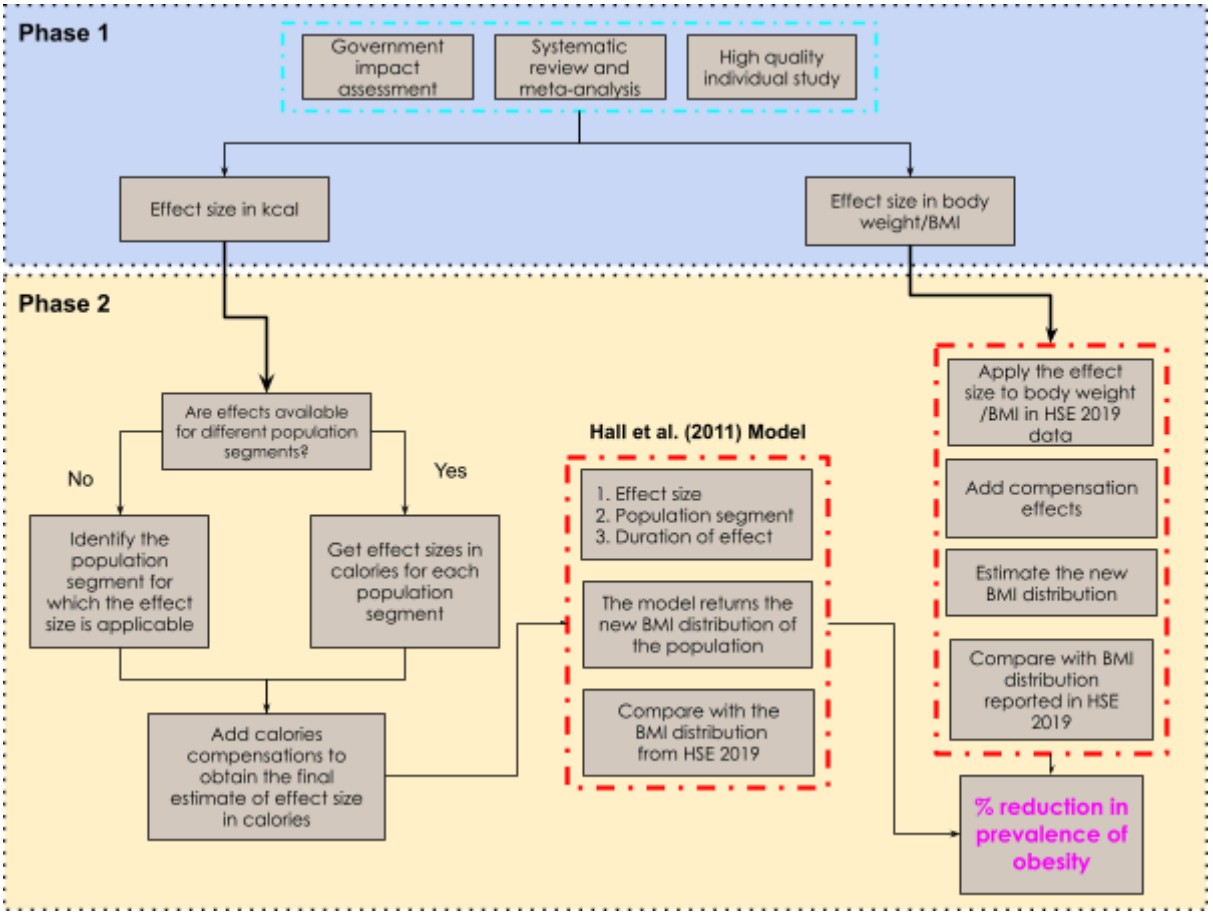
> *"Implementing intervention 'INT' leads to a P% reduction in the prevalence of obesity* in the UK compared to the obesity levels in 2019."*

* defined as people with BMI ≥30

Throughout this technical appendix and the toolkit, we refer to this as the impact of a policy and it is measured in the percentage reduction in prevalence of obesity. Some policies (eg, access to bariatric surgery) target a smaller proportion of the population (lower reach) but have a larger effect size (higher effect). In contrast, other policies (eg, mass media campaigns) target the majority of the population (higher reach) but have a lower effect size for weight loss at a per person level (lower effects).

The modelled policy impact is therefore determined by both the intervention effect size and the population reach. This task was complex and involved making informed assumptions in our analytical modelling. We created an analytical model for each policy listed in the toolkit, taking into account population exposed to the policy and the effect size reported in the scientific literature. The flowchart in figure 4 below schematically illustrates our method and the different routes we've taken to convert the magnitude of effect identified in the evidence synthesis phase to the outcome of 'P% reduction in UK obesity prevalence'.

Figure 4. Process flow chart of estimating the % reduction in prevalence of obesity in a population



Handling magnitude of effect expressed in energy units (kcal/kilojoules)

When outcomes were changes in calories purchased or consumed reported in energy units, the first step was to convert this to an equivalent reduction in daily energy intake. We used different data sources and publicly available information to estimate the daily reduction. The way of calculating this varies across policies and has been detailed under the section 'Estimating the per person impact' on the respective policy pages.

Once the change in daily calorie intakes were estimated, we applied these changes to the daily energy intake values calculated for each person in the cleaned dataset prepared from the HSE 2019 and SHeS 2019 datasets. For more details on this see section About the datasets.

The estimated change in daily energy intake was applied to the estimated daily energy intake. We used different approaches to determine the weight change in children and adults.

For adults, we used the dynamic weight change model described in Hall et al. (2011) (which we will refer to as the Hall equations hereafter) to estimate the weight change trajectory of individuals for a period of five years. The Hall equations take change in calories consumed, along with other factors (detailed below), as inputs to estimate the subsequent change in body weight over a specified time period.

The inputs to the Hall equations are:

1. Magnitude of effect as a change in daily energy intake: Effect sizes are reported as a change in calories purchased, change in calories consumed, or a change in daily energy intake. When it is the former, we first estimate the equivalent change in daily energy intake, as described above. In all models we assume that all calories purchased are subsequently consumed as there is no agreement or evidence regarding how food waste varies by nutrient, so there is no reliable estimate of the number of calories that result in food waste. Therefore we follow previous research outputs, including analysis conducted for the National Food Strategy (Griffith et al., 2021), in making the assumption that food waste is zero and all calories purchased are consumed. Where effect sizes were a reduction in daily energy intake, we apply a compensation effect of 23% (Robinson et al., 2019) to account for any

compensatory behaviour that individuals may show as a result of a decreased intake. For example, if a policy reduces daily energy intake by 100 kcals, we adjust the effect size and reduce it by 23kcals so that the net reduction in daily energy intake is 77kcals.

2. Duration of implementation: We model the impact of policies for a period of five years. Given that we are interested in the complete effect of policies, the duration of five years will capture all the body weight changes. Most (95%) of the effect of the change in energy intake is realised within the first three years, with minimal population level effects on weight or BMI in subsequent years (Hall et al., 2011).
3. Anthropometric measures: Body weight in kilograms and height in metres.
4. Age: Age of an individual measured in years.
5. Sex: Sex of an individual as male or female.
6. Change in sodium intake: Change in intake of sodium as a result of the policy. We assume that there is no change in sodium intake of an individual as a result of a policy being implemented.

In addition to the above, we specify the 'population reach' (ie, the specific population for which the magnitude of effect should be applied). This allows us to estimate the population level impacts of the policy. For policies that result in changes in the food environment, we model the effect on people with BMI=/> 25 (ie, people living with excess weight) only, making the assumption that small changes to number of calories available in the environment will not impact people who do not have excess body fat, as advised by the Calorie Reduction Expert Group. This also makes our estimates conservative (ie, we may underestimate the impact of policies) as it is likely that a proportion of people with a healthy BMI would move to an overweight BMI category over our five-year modelling period.

With these inputs, the model returns the weight change trajectory for the change in energy intake, producing body weights for each day of the implementation period. We take the body weight of the individual at each 365th day (end of a year) and calculate the new BMI assuming that the heights of adults don't change in this period. Using the new BMIs we compute the percentage of individuals in each BMI category and compare it to the baseline distribution. If the magnitude of the effect is negative, indicating a reduction in daily energy intake, we expect to see a

downward shift in percentage of people living with obesity in the fifth year compared to the baseline.

We calculate the difference in percentage of people living with obesity at baseline and in the fifth year. Then, we calculate the proportion that this difference is of the baseline prevalence of obesity (ie, 29.1%) which gives us the relative reduction in prevalence of obesity. This is then used to assign an impact score for each policy.

Note: The Hall equations are used for modelling using the bw: Dynamic Body Weight Models for Children and Adults. R package version 1.0.0 developed by Dalia Camacho García Formentí and Rodrigo Zepeda-Tello in 2018. The package can be accessed at INSP-RH/bw: Dynamic Body Weight Models for Children and Adults.

Children:

Our initial approach to modelling the impact for children was to implement the Hall equations for children as described in Hall et al. (2013). These equations have a growth component to account for the fact that children grow as they age (unlike adults who experience a growth plateau). The equations work by apportioning energy intake to growth and increase in body weight over a period of time. This meant that the change in energy intake due to the policy could not be applied directly to the energy intake at baseline, and required adjustment for calories required for growth. In addition, even if we had adjusted for growth calories, the Hall equations would predict the new body weight. While knowing the body weight is very useful, we were unable to recalculate the new BMI given the inability to predict the new height of the child as they age. We aimed to resolve this using different approaches such as using UK 1990 growth charts to predict the height of a child for different age and sex. However, we felt that the process – adjusting for growth calories and predicting height based on UK 1990 growth charts – would introduce a high degree of error in our estimates of BMI.

Therefore, we decided to use equations published by Henry et al. (1999), referred to hereafter as Henry equations. These basal metabolic rate prediction equations estimate the change in body weight and BMI as a result of a change in daily energy intake. These equations have the following assumptions:

1. The distribution of the number of children in any age, sex, ethnic group doesn't change massively over years and has a similar trend.
2. Children of the same age have similar heights and that this stays consistent across years. This means that when we estimate how changes in body weight affect BMI, we are focusing just on weight change assuming height differences won't significantly affect our results.

We picked Henry equations on the basis of their relatively better scores in being able to accurately and precisely estimate the predictions (for more details, please see Chima et al. (2020).

The Henry equations require the following inputs:

1. Magnitude of effect as a change in daily energy intake: Effect size of an intervention is often reported as a change in calories purchased/consumed or as a change in daily energy intake. For changes in calories purchased or consumed, we estimated the equivalent change in daily energy intake (details of which are reported for each policy in the toolkit). Where the effect size is a change in daily energy intake, we applied a compensation effect of 23% (Robinson et al., 2019) to account for any compensatory behaviour that children may show as a result of a decreased intake. For example, if a policy reduces daily energy intake by 100kcals, we adjust the effect size and reduce it by 23kcals so that the net reduction in daily energy intake is 77kcals. It is possible that compensatory behaviours manifest differently among children compared to adults but we used compensatory data from adult literature due to a scarcity of available evidence for children. The effect of a given policy on daily energy intake is unlikely to be the same for all children and the effect is calculated separately for each child based on the recommended dietary intake values prescribed in the SACN Dietary Reference Values for Energy (2011) report. By proportioning by the recommended daily energy intake we are less likely to overestimate the impact of a policy. We assume that the full magnitude of effect will be experienced by a 17- or 18-year-old female (and an 18-year-old male) while a proportioned effect commensurate with the recommended dietary intake will be experienced by a younger child. For example, if a policy leads to a 20 kcal reduction in daily energy intake, in our model, we assume that this reduction will occur in a

17-year-old female child, while a seven-year-old female child's reduction in daily energy intake is weighted by 0.62. This is because the energy intake of a seven-year-old female is approximately 62% of a 17-year-old's daily energy intake.

2. PAL: This is a multiplier determined by the frequency, duration and intensity of activity of a child. We have taken the PAL values for different child age groups as identified by the SACN in its report Dietary Reference Values for Energy (2011). The median values of PAL adjusted for growth are taken based on the child's age. For ages -< 3, we take 1.40; 3-<10, 1.58; 10-18, 1.75 (please see table 7, SACN 2011).
3. Sex: Sex of the child as male or female as published in the health surveys.

Using these inputs, the new body weight of the child is estimated. We first apply the proportioned magnitude of effect on the baseline daily energy intake to calculate the new daily energy intake once the policy is in effect. Then we divide by PAL to calculate the basal metabolic rate from which we estimate the new body weight of the child. We subtract the new body weight of the child from the baseline body weight to calculate the new BMI of the child. We use the UK-WHO Growth Charts to estimate the BMI percentile of a child to determine the BMI category. The percentage of children in different BMI categories at baseline and post implementation of the policy is calculated and compared. If the magnitude of the effect is negative, indicating a reduction in daily energy intake, we expect to see a downward shift in percentage of people living with obesity in the fifth year compared to the baseline.

By calculating the proportion of the difference in percentage of children living with obesity at baseline and post-implementation period we get the percentage change in obesity prevalence which we report for each policy. This is then used to assign an impact score for each policy.

Handling magnitude of effect expressed in weight units (kilograms/BMI)

In cases of policies where the evidence source expresses outcomes in weight-related units such as kilograms or BMI, we model the effect directly on body weight or BMI variable in the health survey datasets.

The inputs here are as follows:

1. Magnitude of effect: Effect size of an intervention as a change in body weight or weight loss expressed in units such as kilograms, grams or change in BMI etc. Depending on the specific intervention in a policy, we also account for weight regain the following years. These weight regain estimations vary across policies and are described for each policy in their respective pages on the Blueprint toolkit under the section 'Estimating the per person impact'.
2. Population of interest: The population of interest within the sample to which the intervention should be applied. This helps us incorporate the variation of effect size within various segments of the population, if applicable. In addition, it allows us to estimate the population level impact of interventions that are targeted at a specific segment of the population. For example, a behavioural weight management programme aimed at those above a BMI threshold.

For the specific population subgroup, we apply the change in body weight or BMI, adjusted for weight regain on the respective variables in the dataset to obtain the new body weight or BMI. In case of weight loss, we recalculate the BMI and estimate the percentage of people in each BMI category for each of the five years for which we have modelled. Using the new BMIs we compute the percentage of individuals in each BMI category and compare it to the baseline distribution computed using HSE 2019/SHeS 2019. If the magnitude of the effect is negative, indicating a reduction in BMI or weight loss, we expect to see a downward shift in percentage of people living with obesity in the fifth year compared to the baseline.

By calculating the proportion of the difference in percentage of people living with obesity at baseline and in the fifth year, we get the percentage change in obesity prevalence which we report for each policy. This is then used to assign an impact score for each policy.

Handling magnitude of effect for physical activity outcomes

Where the outcome variable reported is objective or self-reported physical activity, we establish estimates of what this equates to in energy consumed (in kcal) before applying the Hall equations. Particularly, when outcomes are in increased time spent

in physical activity measured in minutes, we use the following equation to estimate the energy expenditure:

$$Energy\ Expenditure\ (in\ kcals)\ = \left[\frac{(Metabolic\ Equivalent\ x\ 3.5\ x\ body\ weight\ kg)}{200}\right]\ x\ time\ spent\ doing\ activity\ (in\ minutes)$$

We pick the metabolic equivalent of physical activity. In case evidence indicated multiple physical activities we take an average of the metabolic equivalents of all the activities. And using the body weight of the individual and time spent engaging in physical activity, we calculate the energy expenditure in kilocalories. This is assumed to be the equivalent of a reduction in energy intake which is input to the Hall equations and we follow the approach described in the section Modelling the impact on a single outcome.

Handling disparate population effects

We checked if the evidence source provided a different effect size for various segments of the population, split by population characteristics such as age-group, sex etc. If this was not available, we identified the specific population of interest for which the estimated effect was applicable.

For policies that affect the food environment the majority of the population is impacted, therefore, we modelled the effect on people with BMI=/>25 (ie, people living with overweight or obesity) only, making the assumption that small changes to the environment will not impact people who do not have excess body fat. If the evidence source detailed the effect sizes for different population groups, then the effect sizes and the population segment for which it is applicable was obtained.

Handling policies that applied to specific populations

For some policies, the population reach is determined by eligibility criteria, for example policies that involve a provision of ring-fenced funding or policies that focus on treatment (eg, weight management programmes, bariatric surgery and pharmacological treatments). For such policies, we randomly select eligible individuals in the HSE 2019 and SHeS 2019 datasets and apply the effects to those individuals only.

Getting number estimates from survey data

Based on appendix C: about the population estimates (HSE 2019 Methods)

1. HSE 2019 has survey weights and the appropriate one for use for the variables used for modelling is '*wt_int*'. The sum of '*wt_int*' doesn't equal the population of England.
2. Ratio of sum of sampling weights of an observation in the health survey dataset meeting a criteria to the sum of weights of the entire dataset gives us an estimate for the prevalence of that criteria, we will call this prevalence ratio. For example, assume the variable 'smkng_true' is a variable indicating the smoking status of an observation in the dataset with the value 'Yes' for all people who smoke in the population. If the sum of weights for smkng_true = 'Yes' is 80 and if sum of weights of the survey is 1,600, the ratio of 80/1,600 = 0.05, tells us that the prevalence of smokers in the population is 5%.
3. To get a number estimate of the smokers in the population, we multiply prevalence ratio with the ONS mid year population estimates of the respective survey year.
4. We use HSE 2019 and SHeS 2019 dataset to estimate the prevalence and the ONS mid year population estimates for 2019 to estimate the number of people meeting the eligibility criteria.

How individuals are chosen to receive the intervention

We use the policy title/specification to select individuals who would be impacted by the policy. The policy specification clarifies two things:

- Eligibility criteria: criteria for eligibility for exposure to a policy
    - Helps estimate the number of individuals who are eligible to receive an intervention in the population
- Budget allocation and approximate unit cost of intervention
    - Helps estimate the number of individuals that can receive the effects of a policy

Using the eligibility criteria, we filter the HSE 2019 data to get a subset of all individuals who are eligible for treatment based on the specified criteria. We then get the sum of the survey weights of all individuals meeting the criteria.

$$weight_{eligible} = sum\ of\ survey\ weights\ of\ those\ meeting\ eligibility\ criteria$$

$$weight_{survey} = sum\ of\ survey\ weights\ of\ all\ respondent\ individuals\ in\ HSE\ 2019$$

$$Proportion_{eligible} = \frac{weight_{eligible}}{weight_{survey}}$$

To get the number estimate of the eligible population, we multiply $Proportion_{eligible}$ with the ONS mid year population estimate (see section Getting number estimates from survey data):

$$Population_{eligible} = \frac{weight_{eligible}}{weight_{total}} \times Population_{ONS\ Mid\ Year\ Estimate}$$

Estimating the number of individuals exposed using the allocated budget

From the eligible population, **X** individuals are chosen, hereafter referred to as the 'intervention sample'. The value of **X** is calculated using the budget allocation from policy specification and per unit cost of implementing the policy.

For example, as in policy on GLP-1s, if £500 million is the allocation and assuming £3,360 is the cost of the drugs for a two-year period, the the number of people who can receive the drugs (referred to as the intervention sample) is:

$$Sample_{intervention} = \frac{£\ 500\ million}{£3{,}360} = 148{,}810\ individuals$$

Then we check that:

$$Sample_{intervention} < Population_{eligible}$$

so that there are enough people to choose from within the eligible population. It also is a check to ensure that there isn't any over-allocation of funds.

Now, from the eligible population, we want to choose the intervention sample. Based on the way HSE 2019 survey weights are created, we'd expect to see that the ratio of sum of weights would be equal to ratio of the number estimates of intervention sample and eligible population (see section Getting number estimates from survey data)

$$\frac{weight_{intervention}}{weight_{eligible}} = \frac{Sample_{intervention}}{Population_{eligible}}$$

Rearranging this:

$$weight_{intervention} = \frac{Sample_{intervention}}{Population_{eligible}} \times weight_{eligible}$$

Then, to choose individuals from the survey, we loop through the eligible population to select rows such that the following condition is satisfied:

$$weight_{selected\ rows} \geq weight_{intervention}$$

In this sampling approach it is also ensured that once a row/index is selected, then it is not replaced into the sample. Variables are initiated to ensure that a history of rows selected into the intervention sample is maintained, so that the same observation is not chosen more than once to receive treatment in that year or next, unless the policy specifies otherwise.

Building consensus for policy impacts

Expert survey methodology

Our analytical modelling of the impact of each policy resulted in a list of policies that was ranked in order of how much they would reduce national obesity rates. Before testing our findings with our EAG, we developed a short survey to send to as many experts in the field of obesity, health and food as possible. These experts were primarily working in academia and some were obesity experts working in the civil service and charitable organisations. The survey consisted of a single round and took

approximately 10-15 minutes to complete. Its aim was to gain wider views on the rank order of the different policies in terms of their impact on national obesity prevalence. Respondents were presented with Nesta's suggested order based on the modelling work carried out internally, as well as the model assumptions, and then asked to re-rank any policies if they disagreed with the presented order. There was also an optional question to provide feedback or comments.

In total, 107 people were invited to complete the survey, and of these 42 completed the survey. The mean rank and standard deviations of each policy were then calculated, and we explored how the new ranks compared to Nesta's suggested ranks (eg, policy X moved up the rank two places). After analysis, we shared the survey results with our EAG and discussed if and how to adjust rankings based on the survey responses, with a focus on the policies with greatest uncertainty or disagreement.

The rationale of taking this four-step approach (internal modelling, wider survey, small group discussion with EAG, final rank agreed by Nesta) was a pragmatic choice to balance breadth of oversight with time/engagement with the task. Specifically, we wanted to ensure a large number of experts in the field had contributed to the final ranking, but believed the time-ask needed to be low for a survey; this was then supplemented by a deep-dive discussion with a smaller number of experts who were motivated and engaged to consider the task.

In quantitative terms, we found the average ranked score for each policy following the expert survey did not significantly differ from the ranked score resulting from Nesta's internal modelling. There was some pertinent qualitative feedback that warranted further discussion which we took to the EAG too. Specifically, two respondents queried the differing impact of two policies: restricting marketing of HFSS products on public transport (resulting in an approximate 4.5% reduction in national obesity rates) compared with restricting marketing of HFSS products on TV and online broadcast (resulting in a 0% reduction in national obesity rates). After discussion with the EAG, the decision was made to combine these two levers into a single policy. Three respondents of the survey queried whether we had potentially underestimated the impact of policies that restricted the sale or promotion of HFSS products in the out-of-home sector. We took this query to the EAG for further discussion, and it was concluded that as a significant proportion of out-of-home

organisations are small and medium sized enterprises, they would not fall under legislation proposed in our policies. Hence, the impact of such policies would rightly be lower compared to policies targeting retailers or manufacturers. We shared the final policy ranks with our EAG to invite final queries.

Expert Advisory Group (EAG)

We convened an EAG of 17 academics across UK institutions. We sought their advice and feedback at four two-hour workshops across an 18-month period, as well as ad hoc smaller meetings, and written communication. We sought feedback on various aspects of the project and methodology. For example, we tested the findings of our internal rapid reviews to ensure that key scientific papers or impact assessments had not been missed in our search strategy. For more complex policies, we sought feedback on our analytical model inputs to ensure that we were estimating the correct number of individuals affected by a given policy. We held a workshop to gain consensus on the strength of the evidence base for each policy. Finally, we shared our modelled impacts with the EAG to review and sign off. Not every EAG member provided feedback for every query we proposed. We targeted our questions to EAG members we viewed to have the greatest area of expertise to answer specific questions. Whilst no single member has had sight of every decision, all members were able to query any aspect of the approach, and any policy it was applied to.

Calculating the cost rating

Evidence reviews

We commissioned an independent agency, HealthLumen, to conduct literature reviews to establish the costs of each policy, both to governments and to industry. Full details of the methodology and findings can be found in a separate report on the Methods page of the toolkit.

Applying the cost rating

Once HealthLumen shared the estimated costs to governments, we applied a score for each policy based on the extent of the cost to the UK and devolved governments. A breakdown of the costs to governments and associated scores on a scale of 1 (very high cost) to 5 (very low cost) can be found in the Background section of this technical appendix. We assume the costs to implement are at a UK level. For policies that sit under devolved government powers, we assume that the cost to implement is proportional to the population that would be affected.

Cost savings to the governments

For each policy we estimated the annual cost saving to the UK and devolved governments using analysis conducted by Frontier Economics for the Tony Blair Institute (2023). This analysis showed that current rates of obesity incur an annual cost of £74 billion (this excludes the costs attributed to rates of overweight). These costs are broken down into quality-adjusted life years (QALYs), social care costs (both formal and informal), NHS costs, and productivity losses (primarily driven by losses driven by economic inactivity and early exit from the labour market).

Figure 5. Estimates of the costs in £ billions of obesity and overweight, 2021, taken from [Frontier Economics analysis]

	Obesity	***Overweight***	**Total**
Individual costs	**53.6**	**9.5**	**63.1**
Quality-adjusted life years	48.1	8.5	56.6
Informal social care	5.5	1.0	6.5
NHS costs	**11.4**	**7.9**	**19.2**
Obesity-related illnesses	11.3	7.9	19.2
Mental health	0.0	0.0	0.0

Frontier's analysis takes into account adult obesity. Hence, for policies where children were predominantly affected (for example increasing the provision of free school meals) we used estimates of the future cost of current childhood obesity from Ochoa-Mereno et al. (2024). This study estimated that a four percentage point change in the current prevalence of childhood overweight and obesity is associated with an annual cost of approximately £45 million (inclusive of annual healthcare costs, and costs to society). The study estimated a gradual increase in cost for each decade the current cohort children age. We used the £45 million figure to estimate the potential cost savings of each child-focused policy. For example, if we estimate that a policy decreases childhood obesity prevalence by one percentage point, we assume an annual cost saving of approximately £11 million (ie, £45 million/4).

Figure 6. Annual overweight and obesity-related costs in excess of trends.
(Figure reproduced from Table 9 in Ochoa-Mereno et al., 2024)

Year	Age	Annual healthcare costs (£ millions)	Annual cost to society excluding healthcare costs (£ millions)
2031	20	2.5 (2.4 – 2.8)	42.6 (39.5 – 46.7)
2041	30	2.9 (2.6 – 3)	44.6 (37.4 – 45)
2051	40	5.3 (4.4 – 5.2)	83.6 (68.9 – 82.8)
2061	50	14.7 (12 – 14.7)	171.7 (141.2 – 171.5)
2071	60	22.8 (18.1 – 23.2)	253.3 (204.4 – 255.7)
2081	70	17.9 (13.3 – 18.2)	114.6 (85.2 – 116.4)
2091	80	9.3 (5.6 – 8.7)	58.2 (35.8 – 54.6)

Note: Annual costs are associated with a 4 percentage point (4pp) increase in overweight and obesity and 1 kg/m² in BMI.

In parentheses: Sensitivity analysis of annual costs associated to 4pp increase in overweight and obesity with a BMI increase of 0.75 – 1.25 kg/m².

Rating the strength of evidence (SoE)

Developing the scale

In order to illustrate the certainty or confidence of evidence for each policy, we devised an SoE score. It was essential to balance both reliability and validity in the scale. This was because the final toolkit needed to be useful to policymakers and other decision makers, thus it needed to bridge the gap between academic evidence and practice/implementation. As such, we wanted to value real-world implementations of interventions on a similar level with trials and lab-based research, where the former is more externally-valid/generalisable but typically less controlled, and the latter is typically well-controlled but may lack generalisability beyond the controlled environment.

We carried out rapid desk research into existing SoE scales to understand whether any existed that met our needs. We explored the following scales:

- GRADE
- Education Endowment Foundation
- Department for International Development
- What Works Centre for Local Economic Growth
- What Works Centre for Children's Social Care
- Agency for Healthcare Research and Quality

None of the existing measures met our requirements. In particular, most favoured reliability without consideration for validity. As noted above, this was important as policies needed to be methodologically robust, and also implementable in the real-world. Others measured both SoE and impact, conflating the two things. For our toolkit, we wanted to present SoE separately from impact. Our desk research led us to conclude that an existing measure did not exist which fit our requirements, and that it was common practice to develop tools from scratch or build on existing tools. As such, we developed our own SoE scale. We then refined the scale slightly after discussion with our EAG, making the distinction between more and less robust or

evidence-based simulations (ie, those that are purely theoretical and those that use existing data, for example).

In order to develop our SoE scale we predominantly built on the Department for International Development scale. We balanced both reliability and validity and conceptualised these as below.

When judging **reliability** of the evidence base, we considered:

- The **size** of the body of evidence.
- The **consistency** of the findings produced by studies constituting the body of evidence.

When judging **validity** of the evidence base, we considered:

- The (technical) **quality** of the studies constituting the body of evidence (or the degree to which risk of bias has been addressed).
- The **context** in which the evidence is set.

Our scale was a five-point scale, from 1-5, where: 1 = 'no evidence'; 2 = 'limited evidence'; 3 = 'medium evidence'; 4 = 'strong evidence'; and 5 = 'very strong evidence'. We operationalised the SoE scale by giving examples of what sorts of research would be given each score (see table 5).

Assigning the evidence scores to policies

To assign each policy with an SoE rating we took an iterative approach, seeking to build an expert consensus about the scores. We held a workshop with our EAG to rate the policies. In this workshop, we gave an overview of how we had developed the SoE scale and gave a worked example of the scale. Holding in mind all four factors above (size, consistency, quality, and context), and being informed by the examples provided, the EAG gave SoE ratings. Initially they completed this rating exercise as an individual task during the workshop (hosted on an online survey); we then analysed the findings during the workshop break and presented the ordered list back to the EAG in the second half of the workshop. As a group exercise, we then discussed the evidence ratings and noted down any particular areas of disagreement. Nesta took this information away and tweaked the ratings with

consideration of: the group's feedback; Nesta's desk research; ensuring full utilisation of the SoE scale (ie, including the top end of the scale to be less harsh in ratings). We presented the tweaked ratings back to the EAG via email with a request for comments if anyone disagreed with these final ratings. Finally, we made minor adjustments based on comments.

By the end of this iterative process, we had reached a consensus on the final SoE ratings for each policy. Our approach took elements of a Delphi method (whereby you seek expert consensus on a topic over a series of rounds) but in a lighter-touch and speedier approach; experts were consulted and allowed to think independently initially, and then consider the opinion of other experts to come to a collective view, and then consider again independently with more time.

Table 5. Examples of SoE scores

SoE score	Example of the evidence base
1 = No evidence	• No trials or studies • Thought experiment • Opinion piece
2 = Limited evidence	• Simulation/modelling without validated assumptions or strong underpinning data • Single high quality lab-based study OR multiple low quality lab-based studies • Multiple high quality lab-based studies showing inconsistent results • Findings from implementation study from another highly related topic (not obesity) that ought to share similar mechanism • Qualitative case studies, interviews or process evaluations
3 = Medium evidence	• Simulation/modelling with assumptions validated in the real-world or based on strong data • Pre/post studies or service evaluations • Multiple high quality lab-based studies showing consistent results • Policy implementation (eg, natural experiment) in a local area or areas

4 = Strong evidence	• Multiple field trials with consistent results (not yet made into systematic review)
	• Systematic review or meta-analysis
	• Policy implementation (eg, natural experiment) in the UK, or a country very similar in context to the UK, or in a few countries excl. the UK
	  ○ eg, systematic review or meta-analysis of RCTs showing consistent results
	  ○ eg, policy implementation (natural experiment) in multiple countries similar to UK

5 = Very strong evidence	• Consistent findings from both real-world AND highly controlled studies
	  ○ eg, consistent results from a meta-analysis of RCTs AND real-world policy implementation in the UK
	  ○ eg, consistent results from a systematic review AND multiple countries have implemented it

Acknowledgements

We thank and acknowledge the work of Dalia Camacho García Formentí (PROSPERiA) and Rodrigo Zepeda-Tello (Columbia University) in developing the open source R package – 'bw' which implements the dynamic weight change equations for adults and children (Hall Equations) which we have used extensively in our population level impact modelling. We particularly thank Dalia for her time and patience in answering numerous questions and helping us understand the implementation of the equations for children in the R package.

We thank Oliver Mytton (University College London) for useful insights on population level child obesity modelling which helped us identify a clear way for our modelling approach.

Please note that these individuals were not directly involved with the Blueprint in any capacity.