C. PCT methodology: commuting layer

0. What's new?

This Section of the Manual summarises the methods used to create the Propensity to Cycle (PCT) commuting layer. Further details can be found in the technical appendix of our publication Lovelace et al [1]. In most respects the methods described in Lovelace et al. [1] are still being used in the PCT, with the following improvements:

1. The PCT now exists at the Lower-layer Super Output Area (LSOA) as well as at the Middle-layer Super Output Area level (MSOA). Results presented in the PCT at the MSOA layer (and above) are aggregated from those at the LSOA layer.
2. The PCT is now calculated with reference to an individual-level synthetic population, rather than the aggregate data in previous versions. Full details of the creation of this synthetic population can be found in Appendix 1. This allows the health and carbon impacts to be calculated more accurately, by allowing variables such as mortality to vary according to an individual's age.
3. A fifth 'Government Target (Near Market)' scenario has been added. This is described in the main text of this document, with additional details in Appendix 2. The previous four PCT scenarios are not changed, although the previous "Government Target" scenario has been renamed "Government Target (Equality).
4. Input parameters from the British and Dutch National Travel Surveys have been updated to both now use data from 2010-2016, rather than 2008-2014 (England) and 2010-2014 (Netherlands). This has included updating parameters used in the Go Dutch and Ebikes scenarios. We have likewise updated our input mortality data to come from 2016, and have updated to 2017 the input parameters published by the Department for Transport (DfT) and the Department for the Environment, Food and Rural Affairs (DEFRA). An updated table of inputs for the health and carbon calculations is in Appendix 5.
5. We now estimate cycling and walking speed, and physical activity energy expenditure while cycling and walking, as a function of route hilliness.
6. We now estimate impacts on mortality in terms of the reduction in Years of Life Lost (YLLs) as well as the impact on deaths. We also now additionally estimate impacts on sickness absence.
7. Both the clickable and the image route networks are now calculated at the more detailed LSOA level, providing higher geographic resolution and improved route network results, especially in rural areas.

1. PCT input datasets

i. Core input dataset: 'travel to work' origin-destination dataset from the Census 2011

Using Census 2011 data to build an individual-level synthetic population

To estimate cycling potential, the PCT was designed to use the best available geographically disaggregated data sources on travel patterns. Currently for England and Wales this is the 2011 Census data on main mode of travel to work. For this reason, the commuting layer was the first layer added to the PCT. The 2011 Census was conducted on 27th March 2011 and covered an estimated 94% of the population. All individuals aged 16 or over with a current job were asked "How do you usually travel to work? (Tick one box only, for the longest part, by distance, of your usual journey to work)". The commuting layer of the PCT is based on the 23,903,549 commuters living in England and Wales in 2011, with adults who reported that their home address was also their place of work being treated as non-commuters.

The core input dataset for the synthetic population was Census 2011 origin-destination (OD) pair data that linked each commuter's usual place of residence to the workplace location of their main job (safeguarded dataset 'WM12EW[CT0489]_lsoa' from https://wicid.ukdataservice.ac.uk/).[1] The data are disaggregated by sex; age (categories: 16-24; 25-34; 35-49; 50-64; 65-74; 75+) and mode of travel to work (categories: bicycle; walking; car driver; car passenger; motorcycle; train; underground or light rail; bus; taxi; or other). Usual place of residence and place of work are identified at the level of the Lower-layer Super Output Area (LSOA), although we subsequently aggregated these up into OD pairs between Middle-layer Super Output Areas (MSOAs) for the MSOA layer of the PCT. LSOAs are administrative regions designed to contain a population of around 1560 individuals (average 690 commuters). MSOAs are administrative regions designed to contain a population of around 7500 individuals (average 3330 commuters).

We enhanced this initial OD dataset by merging in information on:

- **Income Deprivation of the home LSOA**. This came from the Index of Multiple Deprivation data from England (IMD2015[2]) and Wales (IMD2014[3]). We ranked LSOAs into fifths for income deprivation, with the fifths defined relative to the country in question.

[1] Note that this dataset was created using the same method that was used in the 2001 census, namely assigning no travel mode to people who work 'at or from home'. Some 2011 ONS travel to work datasets instead use an alternative method in which people who work 'from home' but who do travel in the course of their work are given the mode they usually use for that travel. This does not make much difference, particularly for cycling, but does lead to small discrepancies between our Census 2011 numbers and some other published census data sets. Another potential reason for discrepancy is that some census data sets only include commuters aged 16-65, whereas our dataset includes all ages 16+.

[2] https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

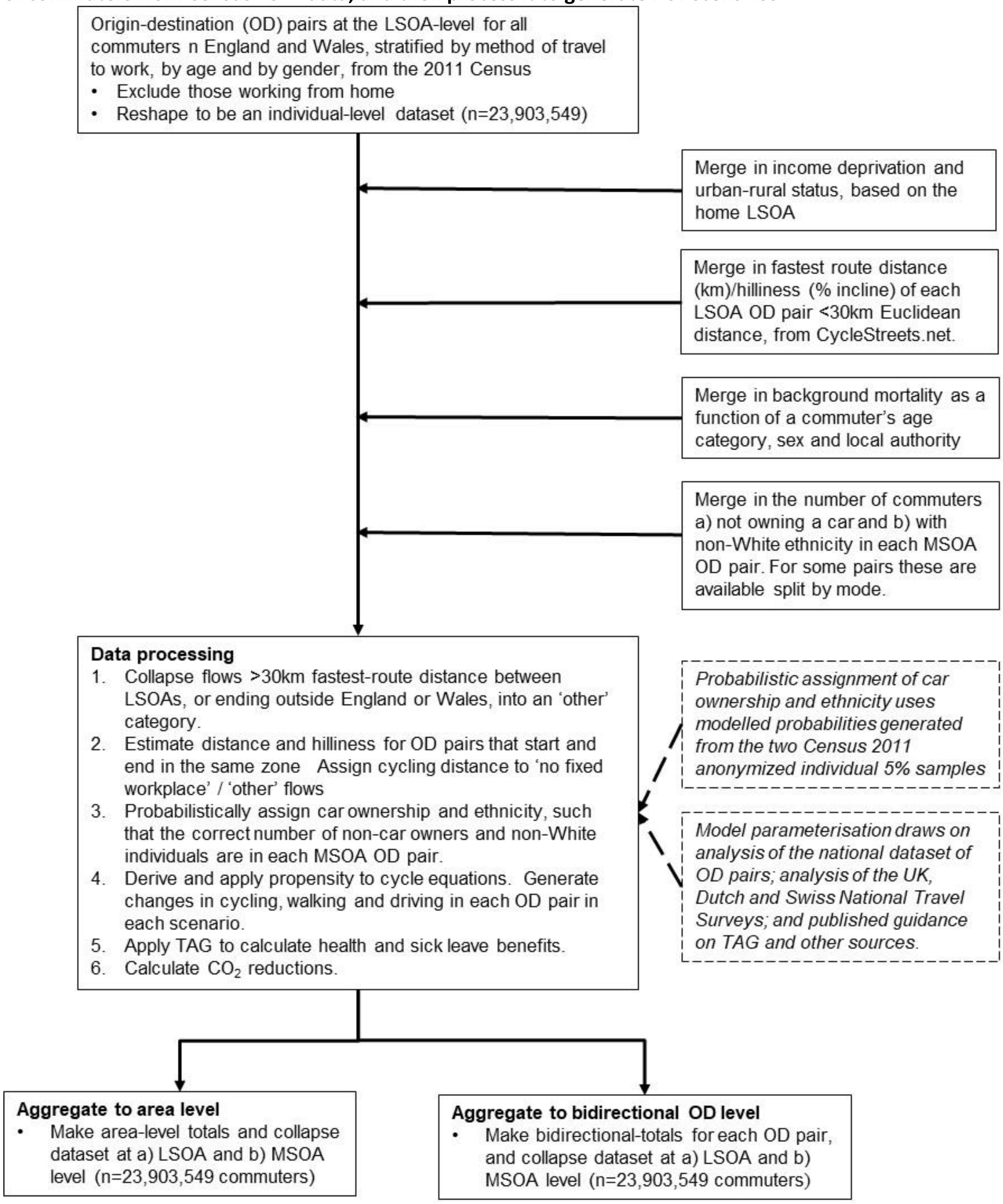
[3] https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/WIMD-2014

- **Urban-rural status and sparsity of the home LSOA**. This came from the Rural Urban Classification (2011) of Lower Layer Super Output Areas in England and Wales.[4] Urban-rural status is categorised into five categories: Urban major conurbation; Urban minor conurbation; Urban city and town; Rural town and fringe; Rural village and dispersed. This dataset also provides a sparsity index, identifying the sparsest 5% of areas in terms of population.
- **Estimated distance and gradient of the 'fastest' routes between the home LSOA and work LSOA.** This was estimated by CycleStreets.net, using the same methods that have been used in previous iterations of the PCT. As in previous versions of the PCT, gradient was measured as a percentage corresponding to the average slope experienced along the course of the route.
- **The background mortality rate**, stratified by age category, sex, and home local authority.
- **Car ownership and ethnicity**. These two variables were probabilistically assigned by drawing on other safeguarded Census 2011 datasets about the number of people a) with no household car and b) of non-white ethnicity in each OD pair. These characteristics were probabilistically assigned, rather than being known for certain for each person in the core input dataset, hence the description of the population created as a "synthetic population". The synthetic population is similar to the true population, in having the correct total number of non-white and non-car-owning individuals in each OD pair, and the correct distribution of these characteristics by age, sex population, region of residence and travel mode. Full details are provided in Appendix 1

A schematic summary of these and other data processing stages described below is presented in Figure 1.

[4] http://ons.maps.arcgis.com/home/item.html?id=9855221596994bde8363a685cb3dd58a

Figure 1: Flow diagram illustrating the input data and processing steps used to create the synthetic population of commuters from Census 2011 data, and then process it to generate PCT scenarios



LSOA = Lower-layer Super Output Area, OD pair = origin-destination pair, MSOA = Middle-layer Super Output Area, TAG

Complementary analyses of national travel surveys, to parameterise scenarios

In addition, some of our analysis decisions and model parameterisation drew on analyses of the National Travel Surveys in England and Wales (2010-2016, although data for Wales was only collected up to 2012, accessed from http://discover.ukdataservice.ac.uk/), the Netherlands (2010-2016, accessed from https://easy.dans.knaw.nl/ui/home) and Switzerland (2010, obtained from the Swiss Federal Statistical Office, Neuchâtel [2], with data processing by Thomas Götschi). All three are nationally-representative surveys that include a travel diary, of duration 1 week in England and Wales, and 1 day in the Netherlands and Switzerland.

ii. Who is included in the propensity to cycle models?

In our analysis, we distinguish between 4 types of OD pairs as shown in Table 1 with reference to the LSOA layer. As this table shows, all commuters are included in our counts of the number of cyclists at baseline. However, we do not model cycling as increasing for OD pairs that have fast route distance of >30 km, or where the workplaces outside England and Wales. All types of OD pairs are included in our zone-level summaries on the PCT. Only some OD pairs are represented as lines in the PCT interface. Specifically, each region only shows lines that a) have a fast-route distance less than 20km, and b) contain more than a certain number of commuters (usually 10 for the MSOA layer and 5 for the LSOA layer) by any mode, counting commuters in both directions. In addition, the Route Network (MSOA) only includes commuters who start and end in the PCT region. The Region Stats tab gives details of the criteria used in each region.

Table 1: Summary of how different types of OD pairs are modelled and represented in PCT, for the LSOA layer*

Type of OD pair	% of comm-uters	% of cyclists at baseline	Included in count of cyclists at baseline?	Modelled as increasing in scenarios?	Included in zone-level summaries in the PCT interface?	Represented as lines in the PCT interface?	Included in Route Network estimates in the PCT interface?
Type 1: <30km, between LSOAs	75.6%	86.9%	Yes	Yes	Yes	Sometimes, see Region Stats tab	Sometimes, see Region Stats tab
Type 2: within LSOAs	3.4%	4.4%	Yes	Yes	Yes	No, represented as centroids	No
Type 3: No fixed workplace	9.1%	4.9%	Yes	Yes	Yes	No	No
Type 4: >30km within England or Wales, or workplace outside England or Wales	11.9%	3.9%	Yes	No	Yes	No	No

* Results for the MSOA layer are similar except that there are a higher proportion of commuters are in Type 2 as opposed to Type 1 flows

2A. Modelling 'route-based baseline propensity to cycle', using distance and hilliness only

i. Plain language overview

In order to generate 'what if' scenarios regarding possible future levels of cycling, we first sought to model current propensity to cycle – i.e. the current proportion of commuters who cycle to work. We did this using OD data from the 2011 Census, and modelling cycling commuting as a function of route distance and route hilliness. We modelled cycling at baseline using logistic regression applied at the individual level, modelling the relationship between the proportion of commuters cycling (the dependent variable) and the fastest-route distance and route gradient (the two explanatory variables). Our equations included squared and square-root terms for distance to capture the non-linear impact of distance on the likelihood of cycling, and included 'interaction' terms to capture the fact that the impact of trip distance varies according to the level of hilliness. We also developed equations to estimate commuting mode share among groups with no fixed workplace.

This model of baseline propensity to cycle formed the basis of three of the five scenarios (Government Target (Equality), Go Dutch and Ebikes), as described in more detail in the next section. Because this model of propensity to cycle relies only on distance and hilliness, we refer to it as "route-based baseline propensity to cycle"

ii. Why focus on distance and hilliness?

In modelling route-based baseline propensity to cycle, we focused on the two characteristics of distance and hilliness as both are strong predictors of the probability of cycling a trip, and as both are likely to continue to have some effect on cycling propensity in all cycling futures. For example, even in high-cycling places like the Netherlands, people are much more likely to cycle a 2 km trip than a 10 km trip. By contrast, other possible predictors of current propensity to cycle, such as sex or age, may be more amenable to change. For example, although cycling in England and Wales is concentrated among younger males, in the Netherlands cycling is more common among women than among men, and is common across all age groups (see Manual C3ii). We did not include such individual-level characteristics in this model as we wanted to generate some scenarios that did not assume that future cyclists in England and Wales would have the same characteristics as current cyclists.

iii. Why focus on more direct 'fast' routes?

In measuring trip distance and hilliness, we focused on the 'fastest' (i.e. more direct) routes presented by CycleStreets. We did this despite the fact that many cyclists currently choose to take a quieter route at the cost of extra time because often the fast route involves sharing with motor traffic on busy roads. However, the aim of the PCT is not to predict exactly where people are currently cycling, rather we are trying to prioritise where to put new infrastructure.

We believe that in general the fastest route should be considered as the first choice for creating good cycling routes. This is particularly the case if one is seeking to encourage cycling among groups currently underrepresented, such as women and older people. This is important for 2 reasons. First, these groups are more likely to be put off cycling on direct routes in the absence of high quality infrastructure. A systematic review found that most people find cycling with busy traffic is hugely off-putting, and this is particularly true of women and probably also older people and those riding with children [3]. Second, these groups are also more likely to be put off by cycling longer distances, which alternative 'quieter' routes may involve. For example, analysis of the National Travel Survey indicates that if a quieter route creates a detour such that a 2-mile trip becomes effectively a 3-mile trip, younger men's propensity to cycle the route will decrease by 11%. But for younger women, the decline is 19%, and for older adults (60+) the propensity would decrease by 35%.

Thus, for utility trips, improving direct routes will reduce safety and time disincentives to cycling. So, while a good proportion of current cyclists may use the 'quieter' route, a big increase in capacity will likely necessitate substantial improvements to the fastest route, which will then carry many more riders from a wider demographic.

iv. Technical details

For all within-LSOA and between-LSOA OD pairs in England and Wales with a fastest-route distance of <30km, we modelled the relationship between the proportion of commuters cycling (the dependent variable) and the fastest-route distance and route gradient (the two explanatory variables). We did this using an individual-level logit model, with the observations being the ~19 million commuters in our synthetic population with OD pair type 1 or 2 (see Table 1). The effect of distance was modelled using linear, square-root and square terms (Equation 1A[5]). The 'gradient' variable was entered as the original gradient derived from CycleStreet.net minus 0.78%, which is the estimated average route gradient in the Netherlands. By centring our gradient measure on the estimated Dutch average in this way, we facilitated the subsequent addition of 'Go Dutch' parameters to the baseline equation (see Section C3ii). Interaction terms were included to capture the fact that the deterrent effect of a steeper slope appeared to be stronger for individuals travelling intermediate distances. The resulting equation[6] for baseline propensity to cycle was:

Equation 1A:

$$\text{logit (pcycle)} = -4.018 + (-0.6369 * \text{distance}) + (1.988 * \text{distance}_{sqrt}) + (0.008775 * \text{distance}_{sq}) + (-0.2555 * \text{gradient}) + (0.02006 * \text{distance*gradient}) + (-0.1234 * \text{distance}_{sqrt} * \text{gradient})$$

$$\text{pcycle} = \exp([\text{logit (pcycle)}]) / (1 + (\exp([\text{logit(pcycle)}])$$

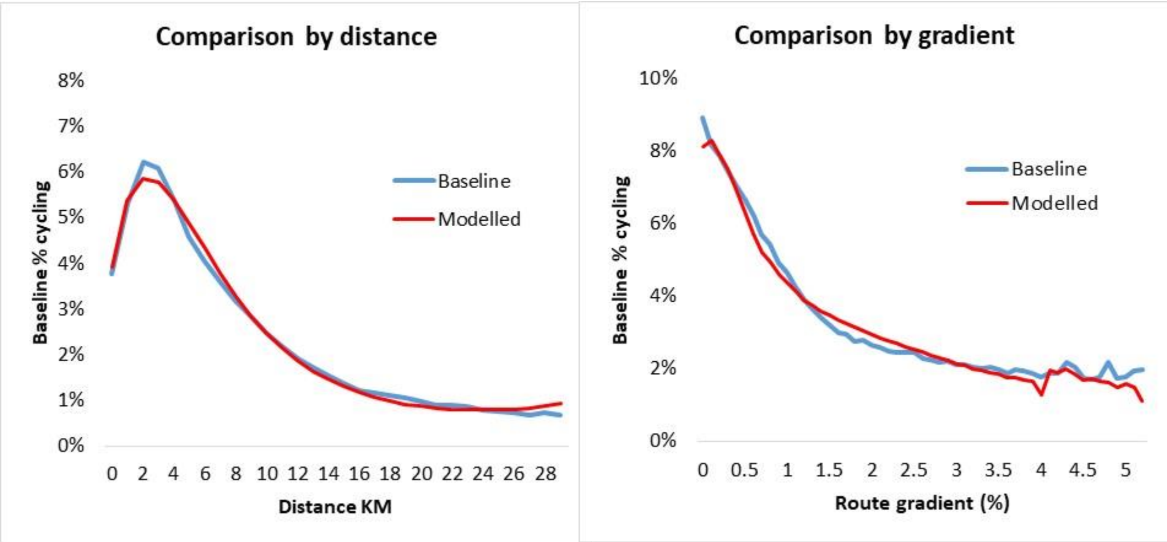
Where 'pcycle' is the proportion of cyclists expected; 'distance' is the fastest-route distance in km, 'distance_{sqrt}' and 'distance_{sq}' are, respectively the square-root and square of distance; and 'gradient' is the fastest-route gradient (centred on 0.78%). Equation 1A showed good fit to the observed data with respect to both distance and hilliness (Figure 2).

---

[5] Equation 1A and Equation 2A were initially generated using MSOA OD pairs.

[6] The equation parameters differ very slightly from those published in Lovelace 2017 because a) they are based on models at the LSOA not MSOA level and 2) they used data from a December 2018 national build that drew on an updated version of CycleStreets with a slightly refined algorithm for estimating hilliness.

Figure 2: Observed versus predicted prevalence of cycling to work among 18,882,504 English and Welsh commuters travelling <30km to work (OD pair types 1 and 2), according to a) route distance and b) route gradient



For commuters with no fixed workplace, we modelled propensity to cycle as a function of the average propensity to cycle among commuters living in the same LSOA and commuting <30km. The resulting equation for route-based baseline propensity to cycle among those with no fixed workplace was:

Equation 2A: logit (pcycle) $= -6.530 + (132.2 * \text{meanpropensity}_{sq}) + (11.47 * \text{meanpropensity}_{sqrt})$

pcycle $= \exp(\text{logit}(\text{pcycle})) / (1 + (\exp(\text{logit}(\text{pcycle})))$

where 'meanpropensity$_{sq}$' is the square of the mean propensity to cycle among commuters in type 1 and type 2 OD pairs in the home LSOA in question, and 'meanpropensity$_{sqrt}$' is the square root term.

Finally, we did not model baseline propensity to cycle among individuals living more than 30km from their place of work or commuting outside England or Wales. Instead, given the considerable uncertainties about where the cycling reported by these individuals was taking place, we assumed no increase in cycling levels among these commuters in our scenarios.

2B. Modelling "multi-characteristic baseline propensity to cycle", using distance, hilliness, and individual characteristics

To generate the Government Target (Near Market) scenario, we again first sought to model baseline (i.e. current) propensity to cycle. As in the previous section, we estimated propensity to cycle among the 19 million commuters with OD pair type 1 or 2 (see Table 1) by fitting logit regression models with cycling as the outcome. We again included the same predictor variables to capture the effect of distance and gradient, and used similar methods to estimate commuting mode share among groups with no fixed workplace.

The difference was that in this model we took account of a wider range of variables, such that we refer to this measure of cycling potential as "multi-characteristic baseline propensity to cycle". Specifically, as well as trip distance and hilliness we additionally took account of:

1. Region (11 regions: the 10 standard regions of England and Wales, subdividing London into Inner and Outer London)
2. Sex (binary)
3. Age category (16 to 24; 25 to 34; 35 to 49; 50 to 64; 65 to 74; 75+)
4. Non-White ethnicity (binary)
5. Having a household car (binary)
6. Fifth of income deprivation
7. Urban-rural status (Urban major conurbation; Urban minor conurbation; Urban city and town; Rural town and fringe; Rural village and dispersed)
8. A sparsity index, identifying the sparsest 5% of areas in terms of population (binary).

We took account of these variables by (i) stratifying by region, sex, and broad age band (16 to 49, and 50+) and then (ii) entering the other variables into the model as predictors. In total, therefore, we modelled baseline propensity to cycle through 44 regression models (11 regions * male/female * 2 age categories). Further details and the coefficients for all the regression equations in all the 44 strata are shown in Appendix 2 in Table 5 - Table 8.

This process of stratification allowed us to take account of the fact the importance of some predictor variables vary according to age, sex, or region. For example, the deterrent effect of longer distance is greater in women and in older people than in young men; and car ownership is less strongly associated with cycling in London than in other regions of England and Wales (further details in Appendix 2)

3. Modelling cycling across scenarios

Five scenarios were developed to explore possible cycling futures in England and Wales. These can be framed in terms of the removal of different infrastructural, cultural, and technological barriers that currently prevent cycling being the natural mode of choice for trips of short to medium distances.

The scenarios are not predictions of the future. They are snapshots indicating how the spatial distribution of cycling may shift as cycling grows based on current travel patterns. At a national level, the Government Target (Equality), Government Target (Near Market) and Gender Equality scenarios could be seen as shorter-term and the Go Dutch and Ebikes scenarios as longer term more ambitious. The choice of scenarios was informed by an Government target to double the number of cycle trips in England and evidence from overseas about which trips *could* be made by cycling.

Each scenario is described below, with both a plain language overview and an account of the technical details. The accounts of the technical details can be complemented by the summary of the scenario generation rules presented in Table 2.

Table 2: Summary of scenario generation rules[7]

Scenario	**Baseline no. cyclists (A)**	**Initial estimation of scenario no. cyclists (B1)**	**Additional processing of scenario no. cyclists (B2)**	**Scenario increase in no. cyclists (C)**
Government Target (Equality)	Recorded no. in Census 2011, OD pair types 1-4.	Column A + (Route-based baseline propensity to cycle [Equations 1A+2A] in OD pair types 1-3 * no. commuters)	• Cap Column B1 at 100%.	Column B2 – Column A
Government Target (Near Market)	Recorded no. in Census 2011, OD pair types 1-4.	Column A + (Multi-characteristic baseline propensity to cycle [Section 2B/Appendix 2] in OD pair types 1-3 * no. commuters)	• Cap Column B1 at 100%.	Column B2 – Column A
Go Dutch	Recorded no. in Census 2011, OD pair types 1-4.	'Go Dutch' propensity to cycle [Equations 1B+2B, with 'dutch'=1 and 'ebike'=0] in OD pair types 1-3 * no. commuters.	• Set Column B1 as equal to Column A if B1 is less than A.	Column B2 – Column A
Ebikes	Recorded no. in Census 2011, OD pair types 1-4.	'Ebikes' propensity to cycle [Equations 1B+2B, with 'dutch'=1 and 'ebike'=1] in OD pair types 1-3 * no. commuters.	• Set Column B1 as equal to Column A if B1 is less than A.	Column B2 – Column A
Gender Equality	Recorded no. in Census 2011, OD pair types 1-4.	Apply Equation 3 in OD pair types 1-3.	• Set Column B1 as equal to Column A if number of males in the OD pair is zero, or if B1 is less than A.	Column B2 – Column A

[7] We considered two different approaches for implementing our scenarios: (1) Switch a fraction of every non-cycling commuter to cycling in a deterministic manner (comparable to the previous implementation of the PCT); or (2) Switch some whole individuals from not cycling to cycling in a probabilistic manner (more similar to the Impacts of Cycling Tool). We decided to adopt the first to facilitate comparisons with the previous implementation of the PCT, and in order to reduce the role of random variation when examining impacts at the small-area or route level. See Appendix 3 for a further discussion of this point.

i. Government Target (Equality) and Government Target (Near Market) scenarios

Plain language overview

The Government Target (Equality) and Government Target (Near Market) scenarios both model a doubling of cycling nationally, corresponding to the proposed target in the English Department for Transport's draft Cycling Delivery Plan to double cycling in England between 2013 to 2025 [4]. They differ in that the Government Target (Equality) scenario models the increase as occurring **solely as a function of trip distance and hilliness**, i.e. equitably across age, sex, and other socio-demographic groups. By contrast the Government Target (Near Market) scenario models the increase as occurring as a function of trip distance and hilliness, **plus a number of sociodemographic and geographical characteristics** (including age, sex, ethnicity, car ownership, income deprivation).

The result in both scenarios is that cycling overall doubles at the national level, but at the local level this growth is not uniform, in absolute or relative terms. Areas with many short, flat trips and a below-average current rate of cycling are projected to more than double in both scenarios. Similarly, the Government Target (Near Market) scenario, areas with many younger men but a below-average current rate of cycling are projected to more than double.

Although the doubling in the scenarios is substantial in relative terms, the rate of cycling under these two scenarios (rising from 3% to 6% of commuters) remains low compared with countries such as the Netherlands and Denmark.

Technical details

The Government Target (Equality) scenario was generated by adding together a) the observed number of cyclists in the 2011 Census, and b) the modelled number of cyclists, as estimated using the route-based baseline propensity to cycle equations described in Section 2A. The Government Target (Near Market) scenario was generated by adding together a) the observed number of cyclists in the 2011 Census, and b) the modelled number of cyclists, as estimated using the multi-characteristic baseline propensity to cycle equations described in Section 2B and Appendix 2.

As only non-cyclists were switched to cycling, we set commuter cyclists to have a scenario increase in cycling of zero. To compensate for this, we scaled up the propensities among non-cycling commuters, such that the total scenario increase in cycling in an LSOA OD pair was equal to the sum of the scenario propensities. For each non-cycling commuter, the scenario increase in cycling was therefore calculated as equal to:

Scenario propensity to cycle * (sum of scenario propensities to cycle in the LSOA OD pair /
sum of scenario propensities to cycle among non-cycling commuters in the LSOA OD pair)

This is equivalent to what we did in the previous aggregate implementation of the PCT, although the calculation could be presented more simply in that previous implementation because it was applied at the aggregate (OD pair) level.

The scenario increase in cycling for each OD pair (and higher aggregations) was calculated by summing the scenario increase in cycling across all constituent commuters. The scenario number of cyclists for each OD pair was calculated by adding the scenario increase in cycling to the observed number of cyclists in Census 2011.

This is illustrated by the following example. Take an OD pair of 5 commuters containing 2 cyclists in the 2011 Census. The 3 non-cycling commuters have modelled scenario propensities to cycle of 0.052, 0.014, and 0.018. The two cyclists have modelled scenario propensities to cycle of 0.052 and 0.074. Given these latter two values need to be set to nought, the scaling factor used is (0.052 + 0.014 + 0.018 + 0.052 + 0.074)/(0.052 + 0.014 + 0.018) = 2.5 Thus the model scenario propensities for the 3 non-cycling commuters become 0.052 * 2.5 = 0.13; 0.014 * 2.5 = 0.035; and 0.018 * 2.5 = 0.045. The two cycling commuters have individual scenario increase in cycling of zero. In the OD pair, the scenario increase in the number of cyclists is 0.13 +0.035 + 0.045 + 0 + 0=0.21. The scenario number of cyclists is 2+ 0.21 = 2.21.

This illustrates how the Government Target (Equality) and Government Target (Near Market) scenarios lead to a doubling of cyclists in England and Wales as a whole, but not necessarily of each OD pair (e.g. In the above example the increase in the number of cyclists was only from 2 to 2.21). Note the reported ‘baseline’ number of cyclists directly influences the total number of cyclists in the scenario (column B2 in Table 2), but does not influence the scenario increase in the number of cyclists (Column C).

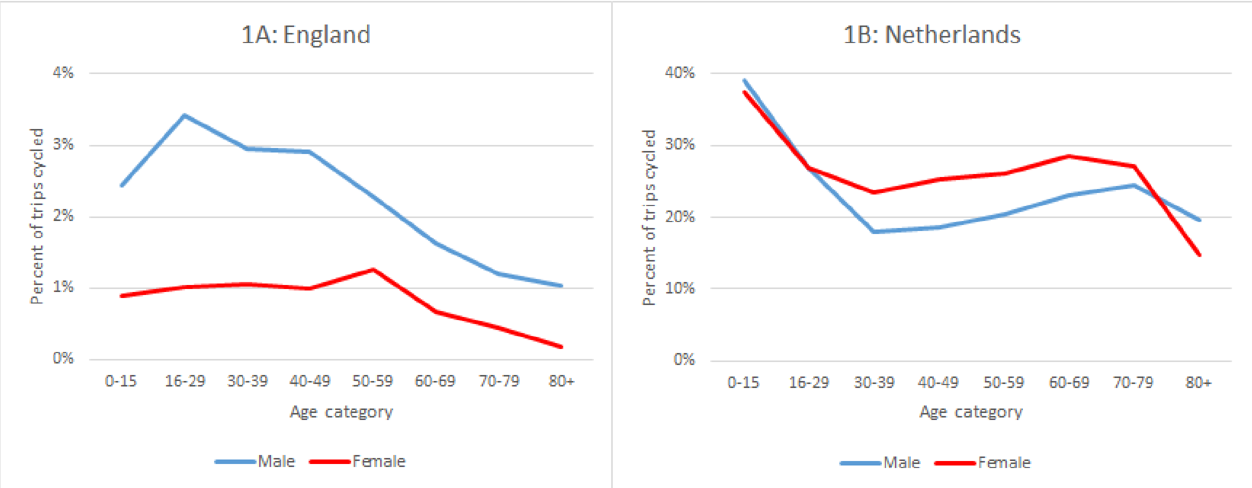
ii. Go Dutch and Ebikes scenarios

Plain language overview

While the Government Target (Equality) and Government Target (Near Market) scenarios model relatively modest increases in cycle commuting, the Go Dutch and Ebikes scenarios are an ambitious vision for what cycling in England and Wales could look like. People in the Netherlands make 28.4% of trips by bicycle, fifteen times higher than the figure of 1.6% in England and Wales. In addition, cycling in England and Wales is skewed towards younger, male cyclists (illustrated in Figure 3 with reference to England). By contrast in the Netherlands cycling remains common into older age, and women are in fact slightly more likely to cycle than men (Figure 3, right-hand side).

This means that the difference between England and the Netherlands is particularly large for women and older people. For example, whereas the cycle mode share is ‘only’ six times higher in the Netherlands than in England for men in their thirties, it is over 20 times higher for women in their thirties or men in their seventies and eighties. For women in their seventies and eighties, the cycle mode share is over 60 times higher in the Netherlands than in England.

Figure 3: Proportion of trips cycled in England versus the Netherlands stratified by age and sex



The Go Dutch scenario represents what would happen if English and Welsh people were as likely as Dutch people to cycle a trip of a given distance and level of hilliness. This scenario thereby captures the proportion of commuters that would be expected to cycle if all areas of England and Wales had the same infrastructure and cycling culture as the Netherlands (but retained their hilliness and commute distance patterns). The scenario was generated by taking the route-based baseline propensity to cycle (see Section 2A) and applying Dutch scaling factors calculated through analysis of the English/Welsh and Dutch National Travel Surveys. The Go Dutch scaling factors comprised two parameters which boost the rate of cycling for each OD pair above the baseline model, with one fixed and one distant-dependent term - the latter takes into account the fact that the "Dutch multiplier" is greater for shorter trips compared to longer trips.

Note that the level of cycling under the Go Dutch scenario is unaffected by the current level of cycling but is instead purely a function of trip distance and hilliness. This means that a few lines or areas show a decrease in cycling under the Go Dutch scenario as compared to baseline; this might happen in a very high-cycling area, where cycle commuting in the 2011 Census is similar to or even higher than the average for the Netherlands. For example, Cambridge, the highest cycling region in England and Wales, shows only a modest overall increase under the Go Dutch scenario for this reason. Planners in Cambridge might therefore want to consider creating a bespoke alternative scenario, e.g. "Go Groningen", using cycling propensity from Groningen, the highest-cycling province in the Netherlands.

The Ebikes scenario models the additional increase in cycling that would be achieved through the widespread uptake of electric cycles ('ebikes'). This scenario is built as an extension of the Go Dutch scenario, making the further assumption that all cyclists in the Go Dutch scenario own an ebike. It builds on the Go Dutch scenario by applying three additional Ebikes scaling factors to account for the increased willingness of ebike users to cycle long distance, hilly and simultaneously long distance and hilly routes. These scaling factors were generated by analysing the impact of ebike ownership based on the Swiss National Household Travel Survey and the Dutch National Travel Survey, weighted to be representative of English and Welsh commuters. This scenario may be particularly suitable for examining cycling potential in hilly areas and/or where trip distances are longer (e.g. in rural areas).

Technical details

For the Go Dutch and Ebikes scenarios, our approach was to start from the Equations estimating baseline propensity to cycle (Equation 1A and 2A) and add additional parameters. Here we provide an overview of the methods and input datasets used: full details can be found in Lovelace et al [1] (but note that the Go Dutch and Ebikes scaling parameters have been updated since publication using more recent data). In calculating the scenario increase in cycling, we deterministically switched fractions of non-cyclist commuters to cycling in a manner comparable to that described for the Government Target (Equality) and Government Target (Near Market) scenarios.

The Go Dutch scenario required us to model the increase in propensity to cycle that would be observed if English and Welsh commuters became as likely to cycle a given trip as Dutch commuters. We estimated this using trip-level analysis of the English/Welsh and Dutch National Travel Surveys, restricting the analysis to commute trips of less than 30km. In estimating the increased propensity to cycle among Dutch people, we included both a main effect term and an interaction term with distance (as a linear term). We introduced the interaction term to reflect the fact that Dutch propensities to cycle exceed English and Welsh propensities by a greater amount for short distances (e.g. Dutch people are 5.6 times more likely to cycle a trip of 0-4.9km versus 3.6 times more likely to cycle a trip 10-14.9km). As hilliness data was not available in the Dutch survey, we weighted the data so that the English and Welsh sample of commuters lived in areas with the same hilliness profile as the Dutch commuters.

The Ebikes scenario builds on the Go Dutch scenario and models the further increase in propensity to cycle that would be observed if all Dutch cyclists acquired an ebike. To generate the relevant parameters, we restricted our analysis to the Dutch National Travel Survey 2013-2016, the only years that measured ebikes as a separate mode. We further restricted our analysis to commute trips made by adults who owned a bicycle. We then compared propensity to cycle between the population of ebike owner trips (N = 4838) with the full population of all bicycle-owner trips (N = 50,990). This analysis therefore takes into account the fact that some ebike owners are already present in the 'Go Dutch' scenario, and captures only the extra cycling that would occur if *everyone* with a traditional bicycle acquired an ebike.

In estimating the extent to which this would increase propensity to cycle in the Ebikes scenario, we included interaction terms with distance (as a linear and squared term). We did this to capture the fact that owning an ebike increases propensity to cycle more for long trips than for short trips (e.g. Dutch ebike owners are 1.1 times more likely than all Dutch bicycle owners to cycle a trip 0-4.9km versus 2.3 times more likely to cycle a trip 10-14.9km). Because we did not have data on hilliness in the Dutch National Travel Survey we could not estimate the magnitude of any interaction between ebike ownership and hilliness in this dataset. We therefore instead estimated the interaction term between ebike use and average route gradient using data from the Swiss National Household Travel Survey 2010.

Adding these 'Go Dutch' and 'Ebikes' parameters together, we derived the following propensity to cycle equation:

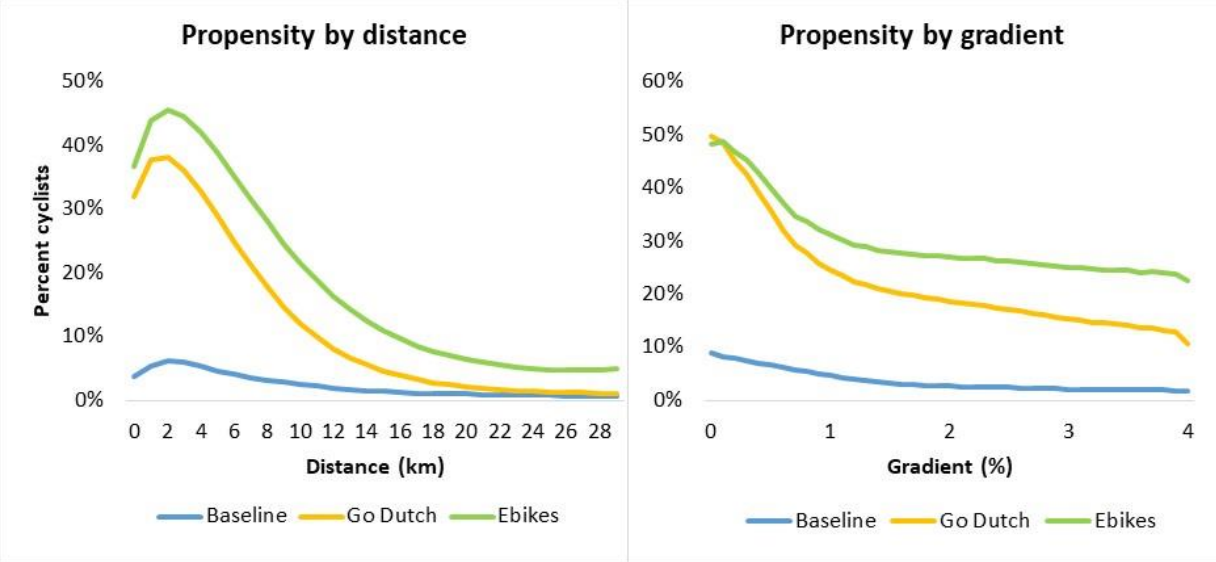
Equation 1B: logit(pcycle) = Equation 1A + Dutch parameters + Ebikes parameters

logit(pcycle) = = -4.018 + (-0.6369 * distance) + (1.988 * $distance_{sqrt}$) + (0.008775 * $distance_{sq}$) + (-0.2555 * gradient) + (0.02006 * distance*gradient) + (-0.1234 * $distance_{sqrt}$ *gradient) + (2.550 * dutch) + (-0.08036 * dutch * distance) + (0.05509 * ebike * distance) + (-0.0002950 * ebike * $distance_{sq}$) + (0.1812 * ebike * gradient).

where 'pcycle' is the proportion of cyclists expected; 'distance' is the fastest-route distance in km, '$distance_{sqrt}$' and '$distance_{sq}$' are, respectively the square-root and square of distance; 'gradient' is the fastest-route gradient (centred on 0.78%); 'Dutch' is a binary variable that takes the value '0' for the Government Target (Equality) scenario and '1' for the Go Dutch or the Ebikes scenario; and 'ebike' is a binary variable that takes the value '0' for the Government Target (Equality) and Go Dutch scenario and '1' for the Ebikes scenario.

Figure 4 shows the distribution of cycling propensity generated by Equation 1B, according to distance and hilliness.

Figure 4: Prevalence of cycling to work at baseline among 18,882,504 English and Welsh commuters travelling <30km to work, and modelled prevalence of cycling to work in Go Dutch and Ebikes scenarios, according to a) route distance and b) route gradient



For commuters with no fixed workplace, we similarly started with Equation 2A, and extended this as follows.

Equation 2B: logit(pcycle) = Equation 2A + mean Dutch parameter + mean Ebikes parameter
= --6.530 + (132.2 * $\text{meanpropensity}_{sq}$) + (11.47 * $\text{meanpropensity}_{sqrt}$) + (dutch * meandutch) + (ebike * meanebike)

where '$\text{meanpropensity}_{sq}$' is the square of the mean propensity to cycle among type 1 and type 2 OD pairs in the home LSOA in question, and '$\text{meanpropensity}_{sqrt}$' is the square root term; 'meandutch' is the average value of the Equation 1B Dutch parameters for commuters living in the same home LSOA; and 'meanebike' is the average value of the Equation 1B Ebikes parameters for commuters living in the same home LSOA.

iii. Gender Equality

Plain language overview

In the 2011 Census, women accounted for 48% of all English and Welsh commuters but only 27% of all cycle commuters. This gender disparity is seen across the country, with no local authority having a proportion of female cyclists greater than 50%. However, in places such as the Netherlands where cycling accounts for a high proportion of trips, women cycle at least as much as men [5, 6]. Places in England and Wales with higher overall levels of commuter cycling also tend to have smaller gender inequalities in commuter cycling [5, 6].

The 'Gender Equality' scenario seeks to capture a situation in which these gender disparities are eliminated. In this respect, it differs somewhat from the preceding four scenarios, as it does not use distance and hilliness data to model propensity to cycle. Instead it assumes that male propensity to cycle remains unchanged – i.e. there is no change in the number of male cycle commuters – and that female propensity to cycle rises to match male propensity. This scenario has the greatest relative impact in areas where the rate of cycling is highly gender-unequal.

Technical details

The Gender Equality scenario assumes that male propensity to cycle remains unchanged – i.e. there is no change in the number of male cycle commuters – and that female propensity to cycle rises to match male propensity in each LSOA-level OD pair. We estimated this number of cyclists in the OD pair in the scenario using the following equation:

<u>Equation 3:</u> $$SNcyclists = BNcyclists_m * (1 + (BNcommuters_f / BNcommuters_m))$$

Where 'SNcyclists' is number of cycle commuters in the Gender Equality scenario, '$BNcyclists_m$' is the recorded number of male cycle commuters at baseline, and '$BNcommuters_f$' and '$BNcommuters_m$' are the total numbers of females and males in the OD pair respectively.

To illustrate how this method works in practice, imagine an OD pair in which 50 from a total of 500 people commute by cycle, 35 males ($BNcyclists_m = 35$) and 15 females ($BNcyclists_f = 15$). 300 of the total trips in the OD pair are made by males ($BNcommuters_m = 300$) and 200 by females ($BNcommuters_f = 200$). Applying Equation 3:

$$\begin{aligned} SNcyclists &= BNcyclists_m * (1 + (BNcommuters_f / BNcommuters_m)) \\ SNcyclists &= 35 * (1 + (200 / 300)) \\ &= 58.3 \end{aligned}$$

All these extra 8.3 cyclists are female, giving a new total of 15 + 8.3 = 23.3 female cyclists (and still 35 male cyclists). Gender Equality in cycling has been reached, such that 11.7% of commute trips are made by cycling among both men (35/300) and women (23.3/200). These additional 8.3 cyclists expected at the OD pair are distributed deterministically across all females who are non-cyclists at baseline. In this worked example, the number of females who were non-cyclists at baseline is 200-15=185, meaning each is given a scenario increase in cycling of 8.3/185=0.045 (with the scenario increase in cycling among mailers all females who were already cycling at baseline being 0).

Equation 3 was applied to commuters with 'no fixed workplace' in the same way. As in other scenarios we assumed no change among commuters travelling >30km or outside England and Wales.

4. Estimating mode shift, health impacts and reductions in carbon emissions

i. Modelling scenario mode shift in walking, car driving and public transport

To estimate the health impacts of our scenarios, we needed to estimate the number of new cyclists who had previously commuted on foot. Similarly, to estimate the carbon impacts of our scenarios, we needed to estimate the number of new cyclists who had previously commuted as car drivers. We also estimated changes in numbers of car passengers, motorcyclists and public transport users, as we believed these would be of interest to some users. We assumed that within any given OD pair commuters were equally likely to shift to cycling from any baseline mode, and therefore the mode shift was proportional to mode share at baseline.

For example, take an OD pair containing 220 commuters at baseline, of whom 20 cycle, 80 walk, 50 are car drivers and 70 use other modes of transport. If the 'Government Target (Equality)' scenario number of cyclists rose to 50 in this OD pair, this would mean that the number of non-cyclists decreased to 170, giving a ratio of change among non-cyclists of 170 / 200 = 0.85. We assumed this 0.85 scenario relative decrease applies to all modes, and (as when calculating the scenario increase in cycling) we applied the scenario levels of walking and driving deterministically at the level of the individual. Thus, for example, each of the 80 individuals who walked to work at baseline have a scenario level of walking value of 0.85, giving an aggregate scenario level of walking across the OD pair of 0.85*80 = 68 walking commuters.

For the purposes of estimating health and carbon impacts of the current level of cycling relative to a 'no cycling' counterfactual, we made the same assumption. For example, again take the OD pair containing 220 commuters at baseline, of whom 20 cycle, 80 walk, 50 are car drivers and 70 use other modes of transport. In a 'no cyclists' counterfactual, the number of non-cyclists would increase to 220, giving a ratio of change among non-cyclists of 220 / 200 = 1.1. Thus in the 'no cyclists' counterfactual, the scenario level of walking among former pedestrians would also be 1.1, giving an aggregate scenario number of walkers of 80 * 1.1 = 88, and so on. When estimating mode split in the 'no cyclists' counterfactual in the small number of OD pairs that at baseline consisted entirely of cyclists, we assumed a mode split of 31% walking, 35% car drivers, 4% car passengers, 2% motorbike, and 28% other modes. These percentages correspond to the observed mode split among the 974 MSOA OD pairs in which 70-99% of individuals cycled in the 2011 Census.

ii. Estimating the physical activity health benefits

An approach based on the DfT TAG was used to estimate the number of a) premature deaths and b) Years of Life Lost (YLLs) that were avoided due to increased physical activity[7]. In this Manual we provide an overview of our methods, with a table of our input parameters in Appendix 5. Further details can be found in [1] and [8].

Trip duration was estimated as a function of the 'fastest' route distance and average cycling speed, with the latter being calculated as a function of gradient (see Appendix 4). In additional, the marginal METs per hour of cycling were also estimated to vary as a function of gradient (see Appendix 4). Trip duration and mMET/hour were combined to estimate the total mMETs per week for each new cyclist (further details in [1]). These mMET/week with then compared to reference TAG values (Appendix 5), and used to estimate relative reductions in background mortality. Specifically, we took the reference TAG mortality relative reduction value of 0.9 and scaled it by the power of the observed weekly mMETs versus the reference weekly mMET value of 8.75, i.e. relative reduction in mortality = 0.9 ^(observed weekly mMETs / 8.75).

The risk of death varies by sex and region, and increases rapidly with age. This was accounted for using age and sex-specific background mortality rates for each local authority in England and Wales. Note that as our update of the PCT with an individual-level synthetic population means we can now assign a mortality rate to each individual based on their own age and sex, rather than relying on the average age and sex distribution of commuter cyclists in their local authority.

To allow for the fact that cycling would in some cases replace walking trips, TAG estimates of the increase in premature deaths due to the reduction in walking were also calculated. The net change in the number of deaths avoided for each OD pair was estimated as the number of deaths avoided due to cycle commuting minus the number of additional deaths due to reduced walking. Note that for a trip of a given distance, walking involves more physical activity than cycling. This means that the observed health benefits can be negative in some areas or on some routes if a high proportion of new cyclists previously walked. This is particularly common in very short trips, and in these cases health (dis)benefits are presented in red.

We then converted our estimate of the net number of deaths avoided into an estimate of the net number of YLLs avoided. We did this by using Global Burden of Disease data from 2017 in England and Wales to estimate the average YLL loss per death, as previously described in [8]. This was done separately by age group, sex and region. As recommended in UK appraisal methods future benefits were discounted by 1.5% per year. This means discounting both premature deaths avoided in the future, and the on-going stream of benefits (YLLs saved) from each death avoided. We provide the YLLs per death already discounted.

The monetary value of the mortality impact was calculated by multiplying the number of YLLs avoided by £57,965, which is the value of a statistical life year used by DfT, converted into 2010 prices. Note that this is the present, single-year value (although as noted above the YLLs saved from a death in one year occur over many years in the future). Users might want to implement a discounting method to sum the value across multiple years into the future. See https://cdn.rawgit.com/npct/pct-shiny/master/regions_www/www/static/03a_manual/pct-bike-eng-user-manual-c1-yll-discounting.xlsx for a spreadsheet template that allows users to do this while simultaneously specifying how many years it will take to achieve the scenario and how long the effects are expected to last for.

iii. Estimating the economic value of reduced sickness absence

We estimated the economic value of reduced sickness absence using an approach similar to that used to estimate the reduction in mortality. We used an identical approach to calculate the increase in weekly cycling mMETs and the decrease in weekly walking mMETs. These mMET/week were then compared to reference TAG values of a 0.25 relative reduction in short-term sickness absence associated with 8.75 mMETs/week (see Appendix 5). We scaled these values by the observed weekly mMETs as follows: relative reduction in sickness absence = 0.75 ^(observed weekly mMETs / 8.75).

Average hours of sickness absence are a function of sickness absence rate (the proportion of hours taken off sick) and total working hours. Both of these factors vary by sex, age and region.[8] We therefore calculated age and sex-specific average annual hours of sickness absence for regions in England and for Wales. The range was from 8.2 hours/year for men aged 16-24 in the East Midlands to 69.9 hours/year for men aged 50-64 in Wales (see Appendix 5 for details).

These average sickness hours were multiplied by the relative reduction in sickness absence. As when calculating the mortality impact, we calculated the reduction in sickness absence due to increased cycling and then subtracted the increase in sickness absence due to decreased walking. This generated the net change in annual sickness hours, which we converted to sickness days by dividing by 7.5.

Finally we multiplied the net change in annual sickness hours by mean hourly salary costs provided by TAG [[9], tab A1.3.1]. We used the market price 'average of all working persons', which was £19.27 using 2010 values. We scaled this figure to vary by region, defining scaling ratios using 2018 median salaries. This resulted in average salaries ranging from £17.16 in the North East to £24.15 in London.

iv. Estimating reductions in transport carbon dioxide emissions from car driving

When comparing each scenario to baseline, we estimated the reduction in transport carbon dioxide (CO_2) emissions as follows:

Change in CO_2-equivalent emissions (in kg) per year
= Change in no. car drivers * former distance travelled by former car drivers * mean cycle commute trips per cyclist per week * 52.2 * CO_2-equivalent emissions (in kg) per kilometre

The change in the number of car drivers was estimated using the mode shift calculations described in Section 4i. Note that we specifically focus on car drivers, not car passengers, as the standard practice in estimating transport CO_2 emissions is to attribute all emissions to the car driver, to avoid double-counting. Their average former distance was assumed to be equal to the new 'fastest-route' distance travelled by the cycle commuters. The mean cycle commute trips per cyclist per week was estimated, stratified by age and sex, from the National Travel Survey. The average CO_2-equivalent emission per kilometre car driving was taken as 0.182kg, which is the 2017 value for an 'average' car of 'unknown' size in the UK government's carbon conversion factors [10].

[8] Note that these two factors sometimes offset each other. For example, in the Annual Population Survey 2016-2018 women had a 50% higher sickness absence rate than men but also worked 25% fewer hours, such that their total annual hours of sickness absence was only 12% higher

5. Aggregate estimates to provide zone-level estimates and to form the Route Network

v. Aggregating OD pairs to give zone-level results, and to give bidirectional lines

Our synthetic population contains directional OD data, i.e. distinguishing between travel from origin A to destination B, and another for travel from origin B to destination A. After performing the modelling stages described above, we aggregated the values for individuals to the zone level by summing our outcome variables across all OD pairs with the same home LSOA. This gave us LSOA-level estimates of the total number of cycle, foot and car commuters living in each LSOA in each scenario, plus the total change in mortality and in CO_2 emissions resulting from behaviour change among residents of that LSOA. Equivalent aggregations were done for MSOA zones.

In addition, we aggregated individuals to generate bidirectional OD pairs at the a) LSOA and b) MSOA level by adding up the values in both directions between a given pair of locations (e.g. adding individuals making the A-to-B commute with individuals making the B-to-A commute). These bidirectional totals are what we present in our visualisation tool.

Note that the MSOA-level lines are therefore generated from the LSOA-level synthetic population. However, the distance and hilliness values assigned to each MSOA straight-line, fast route or quiet route are calculated directly at the MSOA level from CycleStreets, based on the population-weighted centroids of the MSOAs in question.

vi. The Route Network layer

Information about the *aggregate cycling potential* on the road network is shown in the Route Network (LSOA) layer. This layer was generated by aggregating overlapping LSOA-level 'fast' routes, and summing the level of cycling for each scenario using a new function, overline2, which has been published in the R package stplanr [11]. This layer therefore relates to the *capacity* that infrastructure may need to handle.

This layer is available in three complementary formats:

1. Online 'clickable' route network: Available in the Map tab. Users can click each line and see the estimated number of cyclists. This clickable version can be slow to load in large regions. In the largest regions, we only present segments of the route network above a certain minimum number of cyclists - see Region stats tab.
2. Online 'image' route network: Available in the Map tab. The route segments are colour-coded to show the banded number of cyclists (e.g. 10-49). This is much faster to load than the 'clickable' route network, and includes all segments with no minimum number of cyclists.
3. Downloadable route network files: Available from the Region data and National data tabs. This is the equivalent of the online 'clickable' route network, but with no minimum number of cyclists. Users can download this for their own analyses.

Note that more confidence can be placed in the relative rather than the absolute size of the numbers presented for the Route Network: i.e. one can say with more confidence that “the number of commuters increases approximately 5-fold under this scenario" than that "there are 1200 cycle commuters using this route under this scenario". The absolute numbers need to be treated with some caution because they are underestimates as the Route Network layer excludes within-zone commuters, commuters travelling over 30km and commuters with no fixed workplace. Of course, in reality the total number of cyclists would also include people travelling for non-commuting purposes.

5. References

1. Lovelace, R., et al., *The Propensity to Cycle Tool: An open source online system for sustainable transport planning.* Jounral of Transport and Land Use, 2017. **10**(1): p. 505–528.
2. Bundesamt für Statistik, B.f.R., *Mobilität in der Schweiz: Ergebnisse des Mikrozensus Mobilität und Verkehr 2010.* 2012, Neuchâtel: BfS, ARE.
3. Aldred, R., et al., *Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age.* Transport Reviews, 2016. **37**(1): p. 29-55.
4. Department for Transport, *Cycling Delivery Plan*. 2014, [Accessed 15/02/2016 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/364791/141015_Cycling_Delivery_Plan.pdf].
5. Aldred, R., J. Woodcock, and A. Goodman, *Does more cycling mean more diversity in cycling? An analysis of English and Welsh Census data on cycle commuting, exploring shifts in gender and age composition.* Transport Reviews, 2016. **36**(1): p. 28-44.
6. Pucher, J., J. Dill, and S. Handy, *Infrastructure, programs, and policies to increase bicycling: an international review.* Prev Med, 2010. **50 Suppl 1**: p. S106-25.
7. Kahlmeier, S., et al., *Health economic assessment tools (HEAT) for walking and for cycling: economic assessment of transport infrastructure and policies. Methods and user guide, 2014 update*. 2014, Copenhagen: World Health Organisation.
8. Tainio, M. and J. Woodcock, *SO17859 Research into valuing health impacts in Transport Appraisal*. 2016: Department for Transport: accessed 29/08/2019 from https://www.gov.uk/government/publications/transport-appraisal-valuing-health-impacts.
9. Transport, D.f., *TAG UNIT A4.1 Social Impact Appraisal*. 2019, London: Department for Transport; accessed 29/08/2019 from https://www.gov.uk/guidance/transport-analysis-guidance-webtag.
10. DEFRA, *DEFRA Carbon Factors: UK Government conversion factors for Company Reporting, 2015, V2.0*. 2015: Department for the Environment, Food and Rural Affairs [Accessed 18/2/2016 from http://www.ukconversionfactorscarbonsmart.co.uk/].
11. Lovelace, R. and R. Ellison, *stplanr: A Package for Transport Planning.* The R Journal, 2018. **10**(2): p. 7-23.
12. Prampero, P.E.D., et al., *Equation of Motion of a Cyclist.* Journal of Applied Physiology, 1979. **47**(1): p. 201-206.
13. Costa, S., et al., *Quantifying the physical activity energy expenditure of commuters using a combination of global positioning system and combined heart rate and movement sensors.* Prev Med, 2015. **81**: p. 339-44.

14. Sperlich, B., et al., *Biomechanical, cardiorespiratory, metabolic and perceived responses to electrically assisted cycling.* European Journal of Applied Physiology, 2012. **112**(12): p. 4015-4025.
15. Hawley, J.A. and T.D. Noakes, *Peak Power Output Predicts Maximal Oxygen-Uptake and Performance Time in Trained Cyclists.* European Journal of Applied Physiology, 1992. **65**(1): p. 79-83.
16. Sperlich, B., et al., *Biomechanical, cardiorespiratory, metabolic and perceived responses to electrically assisted cycling.* Eur J Appl Physiol, 2012. **112**(12): p. 4015-25.
17. Kelly, P., et al., *Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship.* Int J Behav Nutr Phys Act, 2014. **11**: p. 132.
18. DEFRA, *Greenhouse gas reporting: conversion factors 2017* 2017: Department for the Environment, Food and Rural Affairs [Accessed 18/2/2018 from https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2017].

Appendix 1: Creating a synthetic population

A1.1 Probabilistically assigning information on car ownership

Our initial dataset of age * sex * travel to work load is available at the LSOA layer. Based on the home and work LSOAs, we assigned home and work MSOAs.

For each MSOA-level OD pair, some data is available on car ownership:

1. Selected MSOA OD pairs: number of car owners by travel mode. Only available for OD pairs containing 10+ commuters. (Dataset 'CT0599' available as a safeguarded dataset from https://wicid.ukdataservice.ac.uk/).
2. All MSOA OD pairs: number of car owners in total. (Dataset 'wu09buk_msoa_v1' available as a safeguarded dataset from https://wicid.ukdataservice.ac.uk/).

We used these data to probabilistically assign car ownership to individual commuters, such that the total number of people owning no car in each MSOA OD pair * mode combination was correct. For OD pairs where the number of car owners by mode was not available, we probabilistically assigned car ownership such that the total number of people owning no car in each OD pair was correct. Around 1% of commuters did not live in private households so were not eligible to be asked this question – here and below, they were treated as having no car.

The probabilistic assignment was done as a function of home region, age, sex, and mode. The probabilities used were calculated by combining the two 5% individual-level samples from the Census 2011 (available on the UK Data Archive, dataset IDs 7605 and 7682). We pooled these two datasets together to increase power, although note that because the samples are overlapping they will double count 1 in 40 of the individuals included. Among 2,675,558 commuters in these datasets, we estimated the probability of having no car in the household as a function of sex, age (categories: 16-24; 25-34; 35-49; 50-64; 65+) and mode of travel to work (categories: bicycle; walking; car driver; car passenger or motorcycle; train or underground; bus; taxi or other). We did this by fitting logistic regression models with "no car" as the outcome and with sex, age, and mode as the predictor variables. We ran these regression models stratified by 11 regions (10 standard regions plus London split into Inner and Outer London) to allow for geographical variation in the relationship between car ownership, sex, age, and mode of travel to work.

We then probabilistically assigned car ownership, with the probability of any individual being assigned the status of "no car in household" being proportional to the estimated probability of not owning a car for their age-sex-mode combination. For example, consider an OD pair in North-West England containing two commuters, of whom one is known to have no car in their household. Of the two commuters, one is a male cyclist age 50-64 (modelled probability of not owning a car 23.7% in regression analyses), and the other a female bus commuter age 25-34 (modelled probably of not owning a car 47.8%). One of these two individuals would probabilistically be assigned the status of not owning a car, with the female bus commuter being approximate twice as likely to get this as the male cyclist.

A1.2 Probabilistically assigning information on ethnicity

We categorised ethnicity as a binary variable: "White" (White British, White Irish and Other White) and "non-White" (including Asian, Black, Mixed ethnicity and Other ethnic groups). We chose this categorisation because all the non-White ethnic groups had a considerably lower odds of cycling to work than White ethnic groups in adjusted analyses (Table 3).

Table 3: Odds ratios for cycling to work among commuters in Census 2011 (N=2,078,441 individuals)

Ethnic group	N	Adjusted odds ratio (95%CI)
White British	1,692,751	1
White: Irish	20,103	1.02 (0.95, 1.09)
White: Other White	116,665	0.93 (0.90, 0.96)
Mixed: White + Black Carib./ African	14,127	0.69 (0.63, 0.76)
Mixed: White + Asian/Other mixed	16,254	0.82 (0.76, 0.89)
Asian/Asian British: Indian	59,603	0.25 (0.23, 0.27)
Asian/Asian British: Pakistani	26,775	0.15 (0.13, 0.17)
Asian/Asian British: Bangladeshi	10,798	0.13 (0.11, 0.16)
Asian/Asian British: Chinese	13,058	0.60 (0.54, 0.66)
Asian/Asian British: Other Asian	30,963	0.43 (0.40, 0.47)
Black/Black British: African	30,122	0.28 (0.26, 0.31)
Black/Black British: Carib./Other Black	30,390	0.43 (0.39, 0.46)
Other ethnic group: Any other ethnic group	16,832	0.44 (0.40, 0.48)

Analyses adjust for distance to work, sex, age, household car ownership, and region of England and Wales. The analyses include all individuals in the two anonymised 5% datasets who have no missing data for these variables and who travelled <40km to work.

We assigned ethnicity to individuals in MSOA OD pairs using a very similar procedure to that used for car ownership. Again, for each MSOA-level OD pair, some data was available on ethnicity:

1. Selected MSOA OD pairs: number of non-white individuals by mode. Only for OD pairs containing 5+ white commuters and 5+ non-white commuters. (Dataset 'CT600' available as a safeguarded dataset from https://wicid.ukdataservice.ac.uk/).
2. All MSOA OD pairs: number of non-white individuals in total. (Dataset 'wu08cew_msoa_v1' available as a safeguarded dataset from https://wicid.ukdataservice.ac.uk/).

Again, we used this data to probabilistically assign ethnicity to individual commuters, such that the total number of non-white individuals in each MSOA OD pair * mode combination is correct. For OD pairs where ethnicity by mode is not available, we probabilistically assigned ethnicity such that the total number of non-white individuals in each OD pair was correct. The probabilistic assignment was done as a function of home region, age, sex, mode, and car ownership. The probabilities used were calculated by combining the two 5% individual-level samples that have been made available from the Census.

A1.3 Comparison of synthetic population with true Census data

We conducted tests in Greater Manchester comparing the cycling, walking, and driving mode share of our simulated population with the true Census data. True Census data on total mode according to car ownership and ethnicity was extracted from cross tabs available at the level of local the local authority (datasets DC7201EWla and DC7401EWla available from https://www.nomisweb.co.uk/census/2011). As shown in Table 4, the mode share in our simulated population generally showed a close match to the true Census data for both car ownership and ethnicity. As illustrated in in relation to ethnicity, there was also a good match for the patterning by age and sex, as judged against the 5% anonymized sample available for Greater Manchester.

Table 4: Commute mode share for cycling, walking, and driving among commuters in Greater Manchester: comparison of the true Census data to our simulated population

Group	Mode share	True Census data	Our simulated population
N		1,124,157 / 1,119,467†	1,124,157
No household car	Bicycle Foot Car driver	4.81% 27.2% 16.1%	4.67% 27.5% 15.6%
One or more household cars	Bicycle Foot Car driver	1.81% 7.96% 71.4%	1.82% 7.93% 71.4%
White	Bicycle Foot Car driver	2.35% 10.8% 64.3%	2.35% 10.7% 64.6%
Non-white	Bicycle Foot Car driver	1.47% 11.4% 54.7%	1.50% 12.0% 52.5%

† Car ownership data missing for 0.4% (4960/1124157) of commuters in Greater Manchester

Figure 5: Prevalence of cycling to work by age, sex, and ethnicity: comparison of our simulated population and the true Census data (N= 58,972 commuters from the anonymized 5% local authority sample)

Appendix 2. Modelling baseline propensity to cycle as a function of individual, area, and trip characteristics, as an input for the Government Target (Near Market) scenario

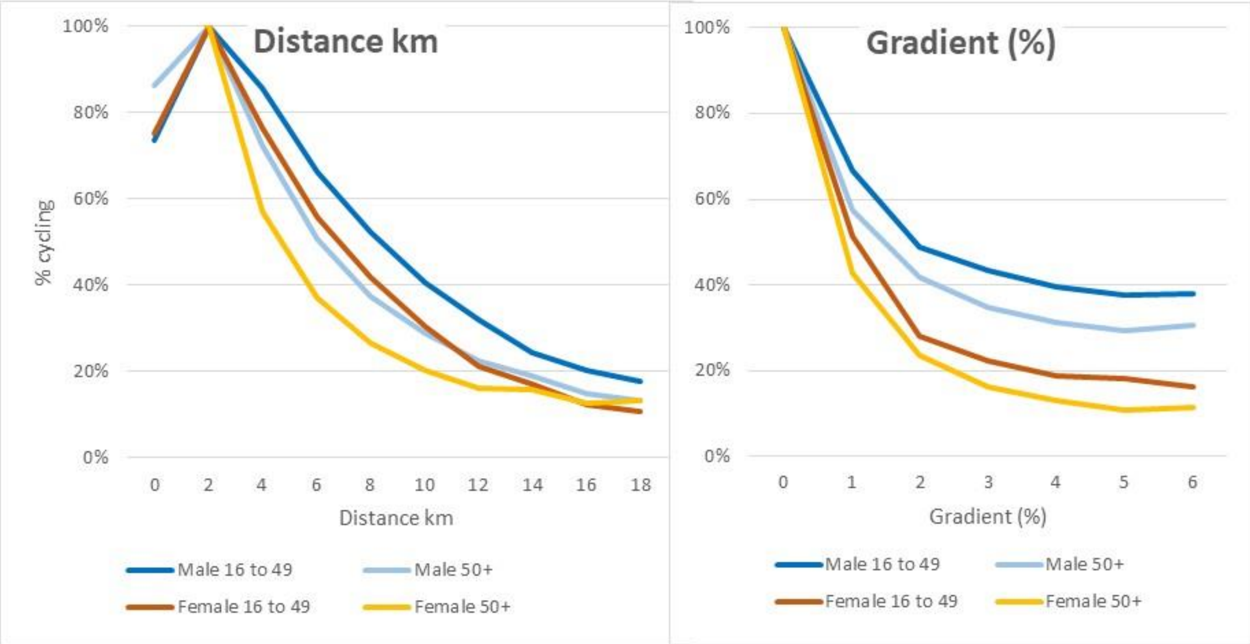
A2.1 Modelling baseline propensity to cycle for within-LSOA flows or between-LSOA flows <30km

To generate the Government Target (Near Market) scenario, we again first sought to model current (baseline) propensity to cycle. As in the previous section, we estimated propensity to cycle among these 19 million commuters by fitting logit regression models with cycling as the outcome. We again included the same predictor variables to capture the effect of distance (linear, square-root and square terms), gradient (centred on the LSOA? mean of 0.78%), and the interaction between distance and gradient.

The differences were that:

1. We additionally included the following predictors: age category (16 to 24; 25 to 34; 35 to 49; 50 to 64; 65 to 74; 75+); non-White ethnicity (binary); having a household car (binary); fifth of income deprivation; urban-rural status (Urban major conurbation; Urban minor conurbation; Urban city and town; Rural town and fringe; Rural village and dispersed); and a sparsity index, identifying the sparsest 5% of areas in terms of population (binary).
2. We stratified the regression models by sex and into two broad age categories (16 to 49, and 50+). We did these because age and sex show interactions with several of the other predictive models. For example, as illustrated in Figure 6, the deterrent effect of distance and of hilliness is stronger in women than in men, and in older people than in younger people. We also stratified the regression model by region (the 10 standard regions of England and Wales, subdividing London into Inner and Outer London). We did this because there exists regional variation with respect to how strongly our predictor variables are associated with cycle commuting. For example, car ownership is less strongly associated with cycling in London than in other regions of England and Wales. Specifically, in Inner London non-car owners are 1.1 times more likely to cycle than car owners (8.0% versus 7.0% mode share) and in outer London non-car owners are 1.5 times more likely to cycle; whereas in all other regions of England and Wales non-car owners are 2.2-3.3 times more likely to cycle. In total, therefore, we parameterised the Government Target (Near Market) scenario by running 44 regression models (male/female * 2 age categories * 11 regions). The sample size across these analyses ranged from 91,475 to 1,056,721 commuters. The coefficients for all the regression equations in all the 44 strata are shown in the Appendix in Table 5 - Table 8.

Figure 6: Effect of distance and hilliness on relative probability of commuter cycling, stratified by age and sex (N=18,882,504 English and Welsh commuters travelling <30km to work)



A2.2 Modelling baseline propensity to cycle for other types of commuters

For commuters with no fixed workplace, we modelled propensity to cycle as a function of the average propensity to cycle among commuters living in the same LSOA and commuting <30km. Specifically, we modelled it as a function of a) the square of the mean propensity to cycle among type 1 and type 2 OD pairs in the home LSOA in question, and b) the square root term of that propensity. This is equivalent to what we did for route-based propensity to cycle. We stratified by region in running these models: regression coefficients can be found in the Appendix in Table 9.

Finally, as in the Government Target (Equality) scenario, we did not model baseline propensity to cycle among individuals living more than 30km from their place of work or commuting outside England or Wales. Instead, given the considerable uncertainties about where the cycling reported by these individuals was taking place, we assumed no increase in cycling levels among these commuters in our scenarios.

A2.3 Applying scaling factors to facilitate comparisons with the Government Target (Equality) scenario

Like Government Target (Equality) scenario, the Government Target (Near Market) scenario models and approximate doubling of cycling nationally, corresponding to the proposed target in the UK government's draft Cycling Delivery Plan to double cycling between 2013 to 2025 [4]. The Government Target (Equality) scenario models a doubling of cycling across England and Wales as a whole by adding “observed cycling” to “expected cycling”. In any given region, however, cycling may more than double or less than double. For example, cycling more than doubled in regions like the West Midlands which had observed cycling levels that were below what were expected. By contrast, because the initial Government

Target (Near Market) baseline propensities were generated in models stratified by region, adding “observed cycling” to “expected cycling” would double cycling *within* each region (as well as doubling cycling nationally). This would complicate comparisons between the two scenarios. For example, if scenario levels of cycling in Solihull were lower in the Government Target (Near Market) than in the Government Target (Equality) scenario, this might be because Solihull had comparatively few "near market" individuals, but it would also partly reflect the fact that the overall level of cycling in the West Midlands increases less in the Government Target (Near Market) scenario (increasing 2-fold) than the Government Target (Equality) scenario (increasing 2.3-fold).

To correct this, we applied regional scaling factors to the propensities generated from the 44 logistic regression models such that the overall increase in cycling in each region was the same in the Government Target (Near Market) scenario as in the Government Target (Equality) scenario. For example, the scenario increase in cycling in the West Midlands was 2.86% in the Government Target (Equality) scenario but initially only 2.07% in the Government Target (Near Market) scenario. The scaling factor for the West Midlands was 2.86/2.07=1.38: a full list of scaling factors is given in the Appendix in Table 10.

The scenario captured by the Government Target (Near Market) scenario can therefore be described as one in which:

- Cycling doubles overall nationally.
- The cycling increase in each region is a function of that region’s distance and hilliness (i.e. the regional increase is the same as in the Government Target (equality) scenario, because of the application of scaling factors).
- Within regions, the cycling increase in each area and in each flow is a function of the age, sex, ethnicity, and car ownership of the constituent commuters; the income deprivation, urban-rural status, and population sparsity of their home LSOA; and the distance and hilliness of their commute trip.

Table 5: Regression coefficients of the Government Target (Near Market) propensity models for male commuters age 16 to 49, stratified by region (analysis restricted to commuters travelling within LSOA or <30 km)

		North East	North West	York-shire & Humber	East Mid-lands	West Mid-lands	East of England	Inner London	Outer London	South East	South West	Wales
Age	16 to 24	0	0	0	0	0	0	0	0	0	0	0
	25 to 34	0.292	0.264	0.249	0.161	0.245	0.282	0.468	0.481	0.245	0.325	0.400
	35 to 49	0.524	0.464	0.450	0.347	0.433	0.387	0.517	0.676	0.392	0.482	0.626
Ethnicity	White	0	0	0	0	0	0	0	0	0	0	0
	Non-White	-0.931	-0.918	-0.984	-0.927	-1.174	-0.567	-0.931	-1.192	-0.644	-0.435	-0.365
Any car in household	Yes	0	0	0	0	0	0	0	0	0	0	0
	No	0.758	0.788	0.801	0.880	0.884	0.934	0.014	0.391	0.753	0.564	0.781
Income deprivation	Fifth 1 (poorest)	0	0	0	0	0	0	0	0	0	0	0
	Fifth 2	-0.050	0.073	0.077	-0.045	0.031	0.082	0.167	0.154	0.007	0.112	-0.026
	Fifth 3	0.035	0.048	0.125	0.050	0.068	0.255	0.121	0.205	0.107	0.101	0.027
	Fifth 4	0.031	0.103	0.227	0.097	0.016	0.247	-0.107	0.302	0.145	0.236	0.178
	Fifth 5 (richest)	0.094	0.038	0.325	0.156	0.017	0.500	-0.242	0.468	0.170	0.305	0.164
Urban-rural status	Urban major conurbation	0	0	0	0	0	0	0	0	0	-	-
	Urban minor conurbation	-	-	0.104	-0.121	-	-	-	-	-	-	-
	Urban city and town	-0.095	0.281	0.568	-0.246	0.378	0.665	-	-0.577	0.283	0	0
	Rural town and fringe	-0.226	0.384	0.187	-0.381	0.319	0.400	-	-0.198	0.146	-0.276	-0.021
	Rural village and dispersed	-0.458	0.225	0.212	-0.494	0.304	0.296	-	-0.458	0.196	-0.401	-0.261
Sparse population	No	0	0	0	0	0	0	0	0	0	0	0
	Yes	0.386	0.140	0.183	-0.122	-0.171	0.181	-	-	-	-0.042	0.386
Fast-route distance	Linear term	-0.652	-0.644	-0.751	-0.708	-0.766	-0.832	-0.612	-0.376	-0.708	-0.767	-0.705
	Square root term	2.083	2.089	2.327	2.192	2.367	2.564	2.510	1.430	2.202	2.400	2.385
	Squared term	0.009	0.008	0.011	0.010	0.011	0.012	0.005	0.003	0.010	0.011	0.009
Gradient	Linear term	-0.224	-0.197	-0.397	-0.143	-0.101	-0.322	0.210	-0.286	-0.266	-0.256	-0.002
Distance* gradient interactions	Distance* gradient	0.000	0.030	-0.009	0.049	0.039	0.017	0.081	0.006	-0.006	-0.003	0.044
	Square root distance* gradient	-0.015	-0.141	0.078	-0.231	-0.189	-0.122	-0.473	-0.010	-0.019	0.053	-0.267
Constant		-4.517	-4.553	-4.705	-3.666	-4.555	-4.772	-4.580	-4.317	-4.171	-4.064	-4.945

Gradient entered as a term centred on 0.78. Cells marked '-' are empty, for example there are no 'urban minor conurbations' in the North East.

Table 6: Regression coefficients of the Government Target (Near Market) propensity models for female commuters age 16 to 49, stratified by region (analysis restricted to commuters travelling within LSOA or <30 km)

		North East	North West	York-shire & Humber	East Mid-lands	West Mid-lands	East of England	Inner London	Outer London	South East	South West	Wales
Age	16 to 24	0	0	0	0	0	0	0	0	0	0	0
	25 to 34	0.522	0.439	0.487	0.379	0.466	0.462	0.632	0.803	0.466	0.545	0.554
	35 to 49	0.557	0.452	0.667	0.533	0.496	0.428	0.519	0.769	0.444	0.490	0.365
Ethnicity	White	0	0	0	0	0	0	0	0	0	0	0
	Non-White	-0.414	-0.584	-0.756	-0.972	-0.970	-0.501	-1.036	-1.291	-0.611	-0.338	-0.176
Any car in household	Yes	0	0	0	0	0	0	0	0	0	0	0
	No	0.635	0.803	0.725	0.725	0.775	0.927	0.011	0.552	0.815	0.555	0.732
Income deprivation	Fifth 1 (poorest)	0	0	0	0	0	0	0	0	0	0	0
	Fifth 2	0.054	0.144	0.090	0.056	0.196	0.303	0.146	0.257	0.159	0.141	0.041
	Fifth 3	0.234	0.074	0.164	0.151	0.333	0.551	0.085	0.328	0.38	-0.012	0.054
	Fifth 4	0.283	0.191	0.368	0.199	0.249	0.528	-0.070	0.392	0.417	0.211	0.730
	Fifth 5 (richest)	0.476	0.142	0.537	0.124	0.209	0.881	-0.220	0.689	0.433	0.271	0.499
Urban-rural status	Urban major conurbation	0	0	0	0	0	0	0	0	0	-	-
	Urban minor conurbation	-	-	0.312	0.040	-	-	-	-	-	-	-
	Urban city and town	-0.213	0.270	1.093	-0.090	0.604	1.369	-	-0.724	0.557	0	0
	Rural town and fringe	-0.280	0.251	0.538	-0.266	0.374	0.964	-	-0.567	0.123	-0.504	-0.266
	Rural village and dispersed	-0.489	0.305	0.439	-0.243	0.706	0.938	-	-0.306	0.245	-0.390	-0.324
Sparse population	No	0	0	0	0	0	0	0	0	0	0	0
	Yes	0.371	0.828	0.183	-0.199	0.215	0.356	-	-	-	0.207	0.606
Fast-route distance	Linear term	-0.924	-0.777	-0.975	-0.842	-0.86	-1.016	-0.891	-0.486	-0.961	-1.044	-1.151
	Square root term	2.633	2.409	2.668	2.268	2.386	2.950	3.395	1.857	2.766	2.975	3.694
	Squared term	0.016	0.011	0.018	0.014	0.014	0.016	0.010	0.003	0.016	0.018	0.016
Gradient	Linear term	-0.911	-0.494	-1.048	-0.590	-0.418	-0.422	0.148	-0.321	-0.333	-0.439	-0.236
Distance* gradient interactions	Distance* gradient	-0.062	0.045	-0.097	0.033	0.031	0.053	0.086	0.013	0.005	-0.023	0.088
	Square root distance* gradient	0.401	-0.190	0.505	-0.142	-0.106	-0.360	-0.503	-0.068	-0.100	0.164	-0.378
Constant		-6.601	-6.411	-6.414	-4.988	-6.336	-6.755	-5.958	-6.149	-6.076	-5.516	-7.285

Gradient entered as a term centred on 0.78. Cells marked ‘-‘ are empty, for example there are no ‘urban minor conurbations’ in the North East.

Table 7: Regression coefficients of the Government Target (Near Market) propensity models for male commuters age 50+, stratified by region (analysis restricted to commuters travelling within LSOA or <30 km)

		North East	North West	York-shire & Humber	East Mid-lands	West Mid-lands	East of England	Inner London	Outer London	South East	South West	Wales
Age	50 to 64	0	0	0	0	0	0	0	0	0	0	0
	65 to 74	-0.625	-0.609	-0.600	-0.592	-0.627	-0.487	-0.751	-0.776	-0.549	-0.708	-0.612
	75+	-0.336	-0.376	-0.445	-0.176	-0.374	-0.482	-0.457	-0.393	-0.402	-0.481	-0.249
Ethnicity	White	0	0	0	0	0	0	0	0	0	0	0
	Non-White	-1.234	-1.071	-0.871	-1.063	-1.212	-0.612	-1.089	-1.312	-0.626	-0.456	-0.360
Any car in household	Yes	0	0	0	0	0	0	0	0	0	0	0
	No	1.084	1.085	1.099	1.185	1.210	1.291	0.126	0.682	1.083	0.865	1.059
Income deprivation	Fifth 1 (poorest)	0	0	0	0	0	0	0	0	0	0	0
	Fifth 2	0.027	0.122	0.031	0.110	0.062	0.147	0.256	0.071	0.042	0.083	-0.105
	Fifth 3	0.074	0.113	0.127	0.127	0.144	0.224	0.324	0.247	0.157	0.168	0.013
	Fifth 4	0.084	0.165	0.187	0.239	0.131	0.226	0.242	0.329	0.234	0.223	0.189
	Fifth 5 (richest)	0.215	0.178	0.407	0.277	0.205	0.467	0.053	0.488	0.213	0.293	0.258
Urban-rural status	Urban major conurbation	0	0	0	0	0	0	0	0	0	-	-
	Urban minor conurbation	-	-	0.234	0.215	-	-	-	-	-	-	-
	Urban city and town	-0.059	0.397	0.846	0.133	0.486	0.684	-	-0.456	0.308	0	0
	Rural town and fringe	-0.131	0.557	0.49	-0.108	0.453	0.44	-	-0.265	0.062	-0.199	-0.003
	Rural village and dispersed	-0.271	0.373	0.262	-0.314	0.257	0.315	-	0.031	-0.069	-0.386	-0.296
Sparse population	No	0	0	0	0	0	0	0	0	0	0	0
	Yes	0.406	0.257	0.129	0.196	0.156	0.488	-	-	-	-0.215	0.351
Fast-route distance	Linear term	-0.595	-0.593	-0.752	-0.628	-0.763	-0.766	-0.561	-0.449	-0.643	-0.713	-0.537
	Square root term	1.851	1.739	2.052	1.645	2.101	2.115	2.228	1.421	1.792	1.950	1.520
	Squared term	0.008	0.008	0.013	0.009	0.011	0.012	0.005	0.005	0.010	0.011	0.008
Gradient	Linear term	-0.310	-0.301	-0.617	-0.355	-0.254	-0.365	0.133	-0.402	-0.327	-0.273	-0.278
Distance* gradient interactions	Distance* gradient	0.024	0.022	-0.034	0.030	0.038	-0.009	0.023	0.000	0.009	0.008	0.014
	Square root distance* gradient	-0.079	-0.080	0.204	-0.106	-0.132	-0.04	-0.296	0.014	-0.054	0.004	-0.064
Constant		-4.273	-4.164	-4.321	-3.380	-4.181	-4.151	-4.342	-3.852	-3.661	-3.344	-3.878

Gradient entered as a term centred on 0.78. Cells marked '-' are empty, for example there are no 'urban minor conurbations' in the North East.

Table 8: Regression coefficients of the Government Target (Near Market) propensity models for female commuters age 50+, stratified by region (analysis restricted to commuters travelling within LSOA or <30 km)

		North East	North West	York-shire & Humber	East Mid-lands	West Mid-lands	East of England	Inner London	Outer London	South East	South West	Wales
Age	50 to 64	0	0	0	0	0	0	0	0	0	0	0
	65 to 74	-0.279	0.014	-0.228	-0.149	-0.036	-0.136	-0.629	-0.638	-0.13	-0.253	-0.478
	75+	0.510	0.496	0.379	0.421	0.768	0.291	-0.296	-0.045	0.518	0.525	0.705
Ethnicity	White	0	0	0	0	0	0	0	0	0	0	0
	Non-White	-1.170	-0.556	-0.914	-1.332	-0.963	-0.704	-1.178	-1.433	-0.659	-0.490	-0.364
Any car in household	Yes	0	0	0	0	0	0	0	0	0	0	0
	No	0.496	0.717	0.550	0.668	0.576	0.736	0.080	0.642	0.770	0.431	0.711
Income deprivation	Fifth 1 (poorest)	0	0	0	0	0	0	0	0	0	0	0
	Fifth 2	0.071	0.098	0.052	0.098	0.092	0.144	0.444	0.381	0.125	0.167	-0.063
	Fifth 3	0.004	0.077	0.000	0.148	0.266	0.316	0.571	0.439	0.254	0.050	-0.119
	Fifth 4	0.044	0.192	0.115	0.242	0.232	0.211	0.534	0.613	0.330	0.248	0.318
	Fifth 5 (richest)	0.220	0.214	0.215	0.181	0.282	0.465	0.379	1.062	0.352	0.342	0.323
Urban-rural status	Urban major conurbation	0	0	0	0	0	0	0	0	0	-	-
	Urban minor conurbation	-	-	0.614	1.082	-	-	-	-	-	-	-
	Urban city and town	0.008	0.755	1.550	1.249	0.932	1.355	-	-0.293	0.500	0	0
	Rural town and fringe	-0.008	0.695	1.221	1.271	1.074	1.356	-	-0.028	0.352	-0.116	0.261
	Rural village and dispersed	-0.023	1.119	0.955	1.109	1.075	1.228	-	-1.017	0.421	-0.070	-0.270
Sparse population	No	0	0	0	0	0	0	0	0	0	0	0
	Yes	0.615	0.511	0.194	0.146	0.198	0.698	-	-	-	-0.159	0.736
Fast-route distance	Linear term	-0.882	-0.602	-0.857	-0.778	-0.765	-0.809	-0.746	-0.626	-0.759	-0.815	-0.78
	Square root term	2.497	1.442	1.969	1.723	1.666	1.852	2.764	1.778	1.733	1.746	2.043
	Squared term	0.015	0.011	0.018	0.015	0.014	0.016	0.009	0.009	0.014	0.016	0.012
Gradient	Linear term	-0.781	-0.792	-1.291	-0.910	-0.680	-0.579	-0.487	-0.615	-0.610	-0.768	-0.717
Distance* gradient interactions	Distance* gradient	0.007	-0.006	-0.096	0.011	0.032	0.021	-0.150	0.024	0.013	-0.034	0.03
	Square root distance* gradient	0.023	0.056	0.543	0.000	-0.051	-0.185	0.377	-0.071	-0.048	0.271	0.002
Constant		-6.004	-5.288	-5.191	-5.054	-5.112	-4.928	-5.662	-5.300	-4.441	-3.705	-5.426

Gradient entered as a term centred on 0.78 Cells marked '-' are empty, for example there are no 'urban minor conurbations' in the North East.

Table 9: Regression coefficients of the Government Target (Near Market) propensity models for commuters with no fixed workplace

		North East	North West	York-shire & Humber	East Mid-lands	West Mid-lands	East of England	Inner London	Outer London	South East	South West	Wales
Mean propensity in the LSOA†	Squared term	351.2	-509.1	7.3	-178.9	182.5	66.5	-178.4	83.5	118.2	-93.0	114.6
	Square root term	18.33	26.06	10.74	19.14	8.37	5.79	28.68	19.06	7.65	15.50	14.82
Constant		-7.350	-8.131	-6.321	-7.612	-5.982	-5.708	-9.639	-7.453	-5.982	-7.140	-6.878

†'Mean propensity in the LSOA' is the average modelled propensity to cycle among within-LSOA commuters or commuters travelling less than 30 km

Table 10: Regional scaling factors applied to Government Target (Near Market) propensities, to generate the same scenario increase in cycling at the regional level as in the Government Target (equality) scenario

	Scenario increase in cycling (%), Government Target (equality) scenario (A)	Scenario increase in cycling (%), Government Target (Near Market) scenario, before scaling (B)	Scaling factor applied to Government Target (Near Market) propensities (A/B)
North East	3.12%	1.80%	1.731
North West	3.43%	2.21%	1.550
Yorkshire and Humber	2.72%	2.64%	1.032
East Midlands	2.99%	2.87%	1.040
West Midlands	2.86%	2.07%	1.382
East of England	3.03%	3.68%	0.824
Inner London	4.30%	7.27%	0.592
Outer London	3.29%	2.34%	1.406
South East	2.86%	3.15%	0.906
South West	2.48%	3.73%	0.665
Wales	2.08%	1.47%	1.422

Appendix 3: Modelling mode shift: a consideration of two possible approaches

We considered two choices in how to model an increase in cycling:

1. **Switch a fraction of every non-cycling commuter to cycling in a deterministic manner**. This gives the average expected impact of each scenario. For example, a certain flow might have a modelled increase of 0.3 cyclists, of which 0.06 cyclists were young white women, 0.01 were young non-white women etc. This is comparable to what we have done previously in the PCT (and was the only approach feasible in previous versions, which were based on aggregate data in which OD pairs with the units of analysis).
2. **Switch some whole individuals from not cycling to cycling in a probabilistic manner**. This takes the average expected impact of each scenario and probabilistically applies it to individuals. For example, a certain flow with an expected increase of 0.3 cyclists, would be probabilistically given an actual increase of 0 or 1 cyclists (or possibly more). Any new cyclists would have their own individual age, sex, ethnicity, and car ownership characteristics. This is similar to the approach used in the Impacts of Cycling Tool

One advantage of the first approach is that it is comparable to what we have done previously in the PCT, and so provide continuity over time. It also may lend itself better to flow-level and small-area-level comparisons, as at these small scales the random influence of probabilistic assignment might sometimes be large. On the other hand, the second approach may be more intuitive to some users, since it deals with the switching of whole individuals. The second approach might also facilitate the implementation of more sophisticated health calculations in the future (health and carbon calculations we are currently implementing in the PCT are compatible with both approaches).

On balance, we considered it best to adopt the first approach to enhance continuity over time and facilitate local analyses. However, we suggest that the second option might become more valuable if the PCT is ever integrated with the impacts of Cycling Tool and/or a more sophisticated approach to health calculations is implemented.

Appendix 4. Modelling speed and energy expenditure as a function of trip hilliness

We are very grateful to Dr Tessa Strain from the MRC Epidemiology Unit, University of Cambridge, for her help in doing this hilliness work.

We sought to estimate the marginal METs (mMETs), and the associated average speed, involved in cycling on routes of different average gradient. In doing this we were attempting to generate a plausible distribution of mMET values by hilliness while not making large changes to the population average values for mMETs and speeds that are recommended by HEAT and TAG, and that PCT had so far been using. In other words, we were seeking to refine the PCT approach to be more sensitive to differential effects across areas according to their hilliness, while retaining broadly similar overall estimates of health impact. It is for this reason that some of our assumptions were made with a view to back-fitting the output values to ultimately be consistent with the TAG and HEAT assumptions previously used in PCT.

Below we outline our methods for doing this, and the associated assumptions. These are also captured in the spreadsheet posted on GitHub at https://github.com/npct/pct-inputs/blob/master/02_intermediate/03_hilliness_calculations/EngWales_mmetspeed_hilliness.xlsx

Target range of hilliness values

In the 2011 Census, 99.9% of all commute routes had an average gradient of ≤7%. We therefore focused on estimating mMETs in this range, applying the 7% incline values to the very small proportion of commuters travelling on steeper slopes.

Power required in cycling

We used the equation from di Prampero et al. [12] that calculates the power (in Watts) required by a cyclist to move. The equation can be broken up into three parts that we have termed road resistance, wind resistance, and gravity. These are then summed.

- Road resistance = CoefficientofRollingResistance * Weight * GroundSpeed
- Wind resistance = CoefficientofAirResistance * BodySurfaceArea * (BarometricPressure/AirTemp) * AirVelocity^2 * GroundSpeed
- Gravity = Gravity *Weight * SineofAngleofIncline * GroundSpeed
- Power (watts) = Road resistance + Wind Resistance + Gravity

Speed assumptions

Uphill moving speed by gradient

We used the data points in Table 11 to develop a decay function for uphill moving speed with incline. We fit this decay function using linear regression, with speed as the outcome and the square root of incline as the predictor.

Table 11: Input or assumed speeds for uphill movement, used to develop decay function

Incline (%)	Speed (km/hr) based on data	Speed (km/hr) based on our decay function	Comments on data source
0	20	20	In 2015, the average moving speed of rides designated as commutes on Strava was 23.7 km/hr[9] but these are likely to be those going faster, with better bikes, over longer distances. Other Strava data from other cities outside the UK also gave average speeds of 20-25 km/hr[10] – the same caveats apply. We took 20km/hour to be conservative, and this made it easier for us to match the observed NTS data.
0.75	16	16.3	0.75% is the average 2-way gradient for Cambridge. Average total journey speed when cycling for transport in Cambridge has been reported to be 16.1km/hour [13]. 16km/hour is approximately the **uphill moving** speed one needs to assume to get this overall journey average if a) downhill speed is 20km/hour and b) 15% of the journey spent stationary (slightly lower than the assumed national average of 20%).
2.8	12.6	12.9	A study of 8 sedentary women averaged a speed of 12.6 km/hr on a 3% short gradient [14]. As these were sedentary women, we expect this slightly to underestimate the average commuter on this gradient.
5.0	9.9	10.4	A study of 8 sedentary women averaged a speed of 9.9 km/hr on a 5% short gradient [14]. As these were sedentary women, we expect this slightly to underestimate the average commuter on this gradient.
7.0	8	8.6	The lowest possible speed for a bike is in the range of 7.2 km/hr[11] but given many cycle up slopes of 10-15%, we expect the speed at 7% gradient to be higher than this minimum

From these data points we used the following formula for our assumption of uphill moving speed as average gradient (as a percentage) increased:

Speed = 20 – 4.3 * (gradient ^ 0.5)

As shown in the third column of Table 11, this provided a relatively good fit to our input data

Stationary time in each journey

We further assumed there was a proportion of each journey spent stationary (e.g. waiting for traffic lights and stuck in traffic) and so speed when moving would be different from total journey speed.

Total journey speed = speed when moving * proportion of time spent moving

In practice this is a simplification of the reality in which a cyclist spends some time stationary, some time travelling slowly e.g. because of traffic, and some time travelling at a steady-state speed.

[9] https://bikmo.com/magazine/results-are-in-strava-reveals-average-british-cycle-commute-length/
[10] https://www.vox.com/2015/10/8/9480951/bike-commute-data-strava
[11] https://www.cyclist.co.uk/in-depth/682/how-steep-is-too-steep-when-cycling-uphill

We assumed the MET value of stationary time was equal to the MET value for steady-speed cycling. We did this to balance out the low energy requirements of waiting stationary at a traffic light (plausibly around 2-3 MET) versus the higher energy requirements involved in the start-stop nature of cycling when interacting with other road users.

Accelerating requires more energy than holding a constant speed. However, we do not have the data to model in detail acceleration and waiting. Thus we assumed that on average the lost time was at the average MMET rate as for the whole journey, rather than at a resting rate. This also provided a much better fit with observed objective data e.g. Costa than assuming the time was spent resting.

We assumed 20% of total journey time could be spent stationary, based on numbers discussed in various London cycling blogs of 10-30%.[12] We selected 20% within this range as a value that gave a fairly close match between average speed in this new method and the average previously used of 14km/hour.

It is likely that better cycling infrastructure e.g. under a Go Dutch scenario would reduce this time but we lack data to include this in the quantitative model.

Other assumptions

Ground Resistance Coefficient

We assumed this to be 0.007 because 0.005 is typical for standard road surface with clincher tyres[13]; we assumed a worse road surface and poorly maintained tyres on commuter cyclists

Weight of rider

We assumed this to be 76.9 kg as an average between the English average male (83.6kg) and female (70.2kg).[14]

Weight of bike and bags

We assumed this to be 16 kg as a good commuter bike can weigh ~11-12kg[15] and we added on 3-4kg for a bag and other bike accessories.

Wind Resistance Coefficient

We assumed this to be 0.5, which indicates no head or tailwind.[16]

[12] https://www.londoncyclist.co.uk/how-much-time-do-you-waste-waiting-at-a-traffic-light/ and http://www.croydoncyclist.co.uk/time-spent-at-traffic-lights/

[13] http://theclimbingcyclist.com/gradients-and-cycling-how-much-harder-are-steeper-climbs/

[14] https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/theaveragebriton

[15] https://inews.co.uk/ibuys/sports-and-fitness/best-bikes-commuting-london-electric-road-hybrid-folding-under-1000/

[16] http://theclimbingcyclist.com/gradients-and-cycling-how-much-harder-are-steeper-climbs/

Frontal Area

We assumed this to be 0.8 m^2 as 0.63 is typical for "tops" position[17]; we estimate a bit higher for upright commuters with non-aerodynamic bags and clothing. This is fractionally higher than the value given in theclimbingcyclist blog[18] (0.6 m^2) but commuter cyclists are more likely to sit very upright even when on "tops" and so this rounding up is probably warranted.

Air density and gravity

These were set at 1.225 kg/m^2 (roughly sea level and 15 degrees temperature) and 9.8m/sec^2. Air density is pressure/air temperature.

Efficiency

Not all power generated will be transferred to the bike. A well-maintained bike is thought to be about 95%.[19] The lower end of the range (where we expect commuters to be) is around 93%;[20] this was the value we used.

Conversion from Watts to mMETs

We used the equation from the Hawley and Noakes (1992) paper showing a very high correlation between Max power output (Wmax) and VO_2 max to convert Watts to L/min of O_2 [15].

$$VO_2 \text{ max} = 0.01141 \times \text{Wmax} + 0.435$$

We then converted from L/min to kcal/min by multiplying by 5.[21] This was then divided by bodyweight and multiplied the time spent moving in hours.

Marginal METs were calculated by subtracting 1 MET.

Calculating average METs and speeds for two-way trips

The PCT is based on average gradients, i.e. a gradient of 1.5% means an average uphill gradient of 1.5% in one direction, and average downhill gradient of 1.5% in the other direction. We assumed that energy expenditure and speed when going downhill was equal to energy expenditure and speed when travelling on the flat.

Estimating energy expenditure for walking and ebiking

We assumed the relative effort of walking on a hill was directly proportional to the relative effort of cycling on a hill. We therefore multiplied all our cycling mMET values by 0.663, the ratio that gave an overall average walking mMET value of 3.6. This is the value that has been reported in the literature, and that we have been using so far in PCT [13].

We assumed that having an ebike halved the additional effort required when going uphill, which is in line with our previous observation that the deterrent effect of hills for ebike-owners was around half the size as non-ebike owners [1]. We further assumed that cycling

[17] https://www.cyclingpowerlab.com/CyclingAerodynamics.aspx
[18] http://theclimbingcyclist.com/gradients-and-cycling-how-much-harder-are-steeper-climbs/
[19] http://theclimbingcyclist.com/gradients-and-cycling-how-much-harder-are-steeper-climbs
[20] https://www.cyclingpowerlab.com/DrivetrainEfficiency.aspx
[21] https://sites.google.com/site/compendiumofphysicalactivities/help/unit-conversions

on the flat was 1.8 mMET less effort on an ebike than on a bicycle. Together this approximately generated the average ebiking mMET of 3.5 that has been reported in the literature, and that we have been using so far in PCT [16].

Estimating speeds for walking and ebiking

We assumed the relative speed penalty of walking on a hill was directly proportional to the relative speed penalty of cycling on a hill. Thus far in PCT we have been assuming cycling speeds of 14 km/hour and walking speeds of 4.8 km/hour, based on HEAT guidance [7, page 16].[22] We multiplied all our newly-calculated cycling speeds values by 4.8/14 = 0.34 to give an updated estimate of walking speed by gradient.

For ebiking, we thought it plausible that the relative speed penalty of travelling on a hill would be smaller than for a traditional bike. We had previously been assuming that ebiking speed was 1.17 times faster than cycling speed. This was based on the Dutch NTS 2013-2016, in which mean cycling speed was 15.0km/hr for bicycle commute trips and 17.5 km/hr for ebike commute trips (17.5/15.0=1.17). We have also estimated that the average route gradient in the Netherlands is 0.78%. We therefore applied this ratio of 1.17 to routes with an average gradient of 0.75%. For lower and higher gradients, we scaled this such that the hilliness effect was half that observed for cycling.

[22] 4.8km/hour is also consistent with NTS 2010-2016, in which the mean speed was 4.6 km/hr for commute walk trips among those for whom walking is usual main mode and excluding trips with implausible speeds (defined as >10km/hr).

Results

Based on the calculations described above, Table 12 presents the average marginal METs and speeds assigned for different modes to routes of different gradients. Note that for the average gradient for commuter cyclists in 2011 was 1.5%, i.e. corresponding to:

- A 2-way speed of 13.9km/hr. This is similar to the HEAT value of 14 km used previously in PCT. It is also consistent with NTS 2010-2016, in which the mean speed was 13.6 km/hr for commute cycle trips among those for whom cycling is usual main mode and excluding trips with implausible speeds (defined as <2km/hr or >25km/hr).
- A mMET value of 5.39. This is similar to the HEAT value of 5.4 used previously in PCT.

Note also that higher gradients are associated with both higher mMET values and slower speeds. In other words, on a hilly route one is expending more energy and doing so for a longer time than on a flatter route of the same distance. These factors both contribute to a greater total energy expenditure for a more hilly commute of a given distance.

Table 12: Average speeds and marginal METs assigned to commute routes of different gradients

Average gradient (%)	Average 2-way marginal METs			Average 2-way speed (km/hr)		
	Cycling	Walking	Ebiking	Cycling	Walking	Ebiking
0.00	4.89	3.10	3.09	16.0	5.5	17.8
0.25	5.23	3.31	3.43	15.1	5.2	17.3
0.50	5.27	3.34	3.47	14.8	5.1	17.1
0.75	**5.37**	**3.40**	3.57	14.5	5.0	17.0
1.00	5.49	3.48	3.69	14.3	4.9	16.8
1.25	5.63	3.56	3.83	14.1	4.8	16.7
1.50	**5.76**	**3.65**	3.96	13.9	4.8	16.6
1.75	5.90	3.73	4.10	13.7	4.7	16.5
2.00	6.03	3.82	4.23	13.6	4.7	16.4
2.25	6.17	3.90	4.37	13.4	4.6	16.3
2.50	6.30	3.99	4.50	13.3	4.6	16.3
2.75	6.42	4.07	4.62	13.1	4.5	16.2
3.00	6.54	4.14	4.74	13.0	4.5	16.1
3.25	6.66	4.22	4.86	12.9	4.4	16.0
3.50	6.77	4.29	4.97	12.8	4.4	16.0
3.75	6.88	4.36	5.08	12.7	4.3	15.9
4.00	6.98	4.42	5.18	12.6	4.3	15.8
4.25	7.08	4.48	5.28	12.5	4.3	15.8
4.50	7.17	4.54	5.37	12.4	4.2	15.7
4.75	7.26	4.60	5.46	12.3	4.2	15.7
5.00	7.34	4.65	5.54	12.2	4.2	15.6
5.25	7.42	4.70	5.62	12.1	4.1	15.5
5.50	7.49	4.74	5.69	12.0	4.1	15.5
5.75	7.56	4.78	5.76	11.9	4.1	15.4
6.00	7.62	4.82	5.82	11.8	4.0	15.4
6.25	7.68	4.86	5.88	11.7	4.0	15.3
6.50	7.73	4.89	5.93	11.6	4.0	15.3
6.75	7.78	4.92	5.98	11.5	4.0	15.2
7.00 +	7.82	4.95	6.02	11.4	3.9	15.2

Appendix 5: Updated table of input parameters for health and carbon calculations

Table 13: Input parameters for estimation of health impacts using TAG, and for estimation of carbon impacts

Parameter description	Used for health or carbon or both?	Parameter value	Parameter source	Comment
Cycling commute distance	Both	Variable by OD pair	CycleStreets fastest route or route average - see final column of Table 1.	
Former walking commute distance	Health	Variable by OD pair	Assumed equal to cycling commute distance.	We assumed former pedestrians previously used the same route, rather than walking a shorter distance to reach the same destination.
Former driving commute distance	Carbon	Variable by OD pair	Assumed equal to cycling commute distance.	We assumed former car drivers previously used the same route, rather than driving a longer distance to reach the same destination.
Mean cycle commute trips per cyclist per week	Carbon	5.46 (men <50); 5.23 (men 50+); 4.13 (women <50); 4.88 (women 50+)	English and Welsh NTS, 2010-2016.	This is the average number of cycle commute trips reported per week among people who say cycling is their usual main mode. It includes respondents who said cycling was their usual main commute mode but reported no cycle commute trips in the past week. Used in calculating car trips/distances and carbon impacts, plus cycling duration impact.
Mean cycle commute trips per cyclist per week in a typical week	Health	7.24 (men <50); 7.32 (men 50+); 6.31 (women <50); 7.23 (women 50+)	English and Welsh NTS, 2010-2016.	This is the number of cycle commute trips reported per week among people who say cycling is their usual main mode*, and who reported at least one cycle commute trip in the past week.* The latter restriction is in place because the TAG input data on mortality risk reduction is largely based on studies asking about a 'typical week' – which we assume will include at least one cycle commute trip for those who say they use cycling as their usual main mode of travel to work.
Mean cycling speed	Health	Variable by trip hilliness, range 11-16 km/hr	See Appendix 4	
Mean walking speed	Health	Variable by trip hilliness, range 4-5.5 km/hr	See Appendix 4	
Mean ebike speed	Health	Variable by trip hilliness, range 13-19 km/hr	See Appendix 4	
Percent cycle trips made by ebikes in Go Dutch scenario	Health	Variable by OD pair, according to route distance	Dutch NTS, 2013-2016.	In the Go Dutch scenario, we assumed the percent of trips made by ebike corresponded to the recorded percentages among all cycle commute trips in the Dutch NTS 2013-2016. The values were 7% cycle trips by ebikes for trips <5km, 13% for trips 5-9.9km, 23% 10-19.9km, 23% for trips 20-30km.
Percent cycle trips made by ebikes in Ebikes scenario	Health	Variable by OD pair, according to route distance	Dutch NTS, 2013-2016.	In the Ebikes scenario, we assumed the percent of trips made by ebike corresponded to the recorded percentages among cycle commute trips made by ebike owners in the Dutch NTS 2013-2016. The values were 71% cycle trips by ebikes for trips <5km, 91% for trips 5-9.9km and 93% 10-19.9km. We assumed 100% for trips 20-30km.

Parameter description	Used for health or carbon or both?	Parameter value	Parameter source	Comment
Reference weekly cycling and walking mMETs	Health	8.75 min/week	Systematic review [17]	Note that 8.75 mMET per week approximately corresponds to achieving the World Health Organisation guidelines of 150 minutes of moderate-to-vigorous physical activity per week, or 11.25 METs.
Mortality reduction for reference levels of cycling and walking	Health	0.9	Systematic review [17]	Reduced relative risk = 1-0.9 = 0.1 or a 10% reduction for the reference weekly mMETs or walking or cycling. After scaling for the actual observed weekly mMETs, this reduced relative risk was capped at a 45% reduction for cycling, and 30% for walking.
Sickness absence reduction for reference levels of cycling and walking	Health	0.75	TAG guidance [9], paragraph 3.2.17	After scaling for the actual observed weekly mMETs, this reduced relative risk was capped at a 50% reduction.
Background annual mortality rate for commuters	Health	Variable by age category, sex, and local authority	Mortality rate for adults aged 16+ in England and Wales.	Calculated using data published by the Office for National Statistics on deaths and the mid-year population for each local authority in England in 2016 (downloaded from https://www.nomisweb.co.uk/). For each local authority, we took mortality rates for males and females in five-year age bands and weighted these by the age profile of commuters. In this way we calculated mortality rates for the age categories available in the Census 2011 data (16-24, 25-34, 35-49, 50-64, 65-74, 75+).
Discounted, average YLL loss per death	Health	Variable by age, sex and region	Global Burden of Disease Study, 2017	These are calculated from the Global Burden of Disease Study 2017, using results for England and Wales. They are then discounted over 100 years using a 1.5% discount rate For males the averages are: 16-24 = 43.0, 25-34 = 39.5, 35-49 = 33.3, 50-64 = 25.3, 65-74 = 18.2, 75+ = 9.4. The females the averages are: 16-24 = 43.1, 25-34 = 39.4, 35-49 = 33.0, 50-64 = 25.5, 65-74 = 18.2, 75+ = 8.5. The values vary only modestly across regions, with a range of up to 0.5 years.
Average sickness hours per year	Health	Variable by age, sex and region	Annual Population Survey, 2016-2018[23]	Hours of sickness absence in the past week was calculated as the difference between total usual working hours and actual working hours, for people who said that the difference was due to sickness or injury (method described by ONS here[24]). This was then multiplied by 52 to give the average number of hours in the past year. We restricted the analysis to commuters, i.e. people with a job who did not work from home. The range was from 8.2 hours/year for men aged 16-24 in the East Midlands to 69.9 hours/year for men aged 50-64 in Wales.

[23] Available from https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8489

[24] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/articles/sicknessabsenceinthelabourmarket/2014-02-25

Parameter description	Used for health or carbon or both?	Parameter value	Parameter source	Comment
Change in no. cycle commuters	Both	Variable by scenario	Equal to the 'scenario-increase in cycling', see Table 2	
Change in no. former pedestrians	Health	Variable by scenario	Mode shift estimation described in Section 4	
Change in no. former car drivers	Carbon	Variable by scenario	Mode shift estimation described in Section 4	Note that we specifically focus on car drivers, not car passengers, as the standard practice in estimating transport CO_2 emissions is to attribute all emissions to the car driver, to avoid double-counting
Value of a statistical life year	Health	£57,965	Provided by DfT	The DfT uses £60,000 as the cost of a statistical life year in 2012 prices. In line with TAG guidance, we used GDP deflator values to convert this to 2010 prices, by dividing by 1.0351
Economic cost per hour of sickness absence	Health	Variable by region	TAG ([9], tab A1.3.1) plus regional salary data	We took the TAG average value for all working persons in 2010 prices (choosing 2010 in line with TAG guidance), and scaled this up or down according to regional salary differences in 2018 derived by ONS from the Annual Survey of Hours and Earnings.[25]
CO_2-equivalent emissions, kg per km	Carbon	0.182	DEFRA 2017	This is the 2017 value for an 'average' car of 'unknown' size and fuel type in the UK government's carbon conversion factors [18].

CO_2= carbon dioxide; DEFRA=Department for the Environment, Food and Rural Affairs; DfT=Department for Transport; mMET=marginal Metabolic Equivalent Task; NTS=National Travel Survey; OD pair =origin-destination pair, YLL = Years of Life Lost, TAG

Note that we assumed a single constant value across all individuals for:

- Average emission factor of a car (DEFRA, 2017)
- Value of a statistical life year

Plausibly any of these values may vary by age, sex or region, but DEFRA and DfT do not provide values disaggregated by these characteristics. We likewise only included regional variation, rather than variation by age or sex, for salaries in the sickness absence calculations. Possibly we could make some improvements on this in future iterations.

[25]https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2018