## C. PCT methodology: schools layer

1. PCT input datasets ........ 2
   i. Core input dataset: The National School Census 2011 ........ 2
2. Modelling baseline propensity to cycle ........ 3
   ii. Plain language overview ........ 3
   iii. Technical details ........ 3
3. Modelling cycling across scenarios ........ 5
   i. Government Target scenario ........ 5
   ii. Go Dutch scenario ........ 5
   iii. Go Cambridge scenario ........ 7
4. Estimating mode shift, health impacts and reductions in carbon emissions ........ 9
   i. Modelling scenario mode shift in walking and car driving ........ 9
   ii. Estimating the physical activity impacts ........ 9
   iii. Estimating reductions in transport carbon dioxide emissions from car driving ........ 10
5. Aggregate to provide zone-level and school-level estimates, and to form the Route Network.... 10
6. References ........ 10

In this Section of the Manual we summarise the method used to create the PCT schools layer. Full details can be found in the technical appendix of our publication Goodman et al. [1]

## 1. PCT input datasets

### i. Core input dataset: The National School Census 2011

Our core input dataset was 2011 National School Census (NSC) in England. Each January, all state-funded schools in England have a statutory requirement to submit information on a range of pupil characteristics. For the PCT, we used data from the January 2011 NSC because this was the most recent year in which information was collected on pupils' usual, main mode of travel to school. 'Usual' mode of travel was defined as that used most frequently by the pupil throughout the year, and 'main' mode defined as that used for the longest distance – i.e. equivalent to the question from the UK household Census that was used to measure commuting to work in the PCT-commuting layer [2].

The information submitted by schools was used by the Department for Education to generate origin-destination (OD) pairs that linked each pupil's Lower Super Output Area (LSOA) of residence to their school, and disaggregated these OD pairs by travel mode. LSOAs are administrative regions designed to contain a population of around 1560 individuals (average 360 children age ≤18). This data was provided to us in an anonymized form by the Department for Education, with mode defined in four categories: foot, bicycle, car/van, and bus/train/other.

Schooling in England is divided into 'early years' (up to age 4), 'primary' (age 5 to 11) and 'secondary' (age 11 to 18). We created modelled propensity to cycle separately for early years/primary (henceforth 'primary') schools and secondary schools.

Independent (private) schools made up 9.9% of all schools in England in 2011, and contained 7.1% of all school children. These schools were not required to provide data on usual mode of travel to school in the NSC, and so were not present in the OD dataset. Of the 21,649 state-funded schools in England in 2011, we excluded 62 boarding schools plus 144 schools where home address or mode of travel to school was unknown for more than 25% of pupils. The PCT schools layer was created from the remaining 17,515 primary schools (4,188,769 pupils) and 3928 secondary schools (3,253,763 pupils). This corresponds to 98.8% of all pupils in state-funded schools (99.7% for primary and 97.7% for secondary); and 91.8% of all pupils in England including independent schools.

We enhanced these OD datasets by merging in route characteristic data on the distance and gradient of the 'fastest' routes. This was done using a routing algorithm 'developed for cyclists by cyclists' by the not-for-profit organisation CycleStreets (www.CycleStreets.net). Gradient was measured as the average slope experienced along the course of the route as a percentage. For example, a gradient of 2% indicates that for every 100m travelled horizontally the route involves a total change in vertical distance of 2m. This change of 2m could potentially reflect a rise of 2m or a fall of 2m or, for example, a rise of 1m followed by a fall of 1m.

*Complementary analyses of national travel surveys, to parameterise scenarios*

In addition to the NSC dataset, some of our analysis decisions and model parameterisation drew on analyses of the National Travel Surveys (NTS) in England (2010-2016, accessed from [http://discover.ukdataservice.ac.uk/](http://discover.ukdataservice.ac.uk/)) and the Netherlands (2010-2016, accessed from https://easy.dans.knaw.nl/ui/home). Both are nationally-representative surveys that include a travel diary, of duration 1 week in England and 1 day in the Netherlands.

## 2. Modelling baseline propensity to cycle

### ii. 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 children who cycle to school. We did this using OD data from the 2011 NSC, and modelling cycling to school as a function of route distance and route hilliness. We modelled cycling at baseline so using logistic regression applied at the individual level, modelling the relationship between the proportion of children cycling (the dependent variable) and the fastest route distance and route gradient (the two explanatory variables). We fit these equations separately for primary and secondary school children.

This model of baseline propensity to cycle formed the basis of both our scenarios (Government Target and Go Dutch), as described in more detail in the next section.

### iii. Technical details

To capture the highly non-linear relationship between distance and propensity to cycle to school, distance was modelled using linear, square-root and square terms (Equations 1.1 and 2.1). The 'gradient' variable was entered as the original gradient derived from CycleStreet.net minus 0.63%, which is the estimated average route gradient of children travelling to school 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. We also fitted an interaction term between distance (as a linear term) and the centred measure of gradient. The resulting equations were:

Equation 1.1 (primary school children):

$$\text{logit (pcycle)} = -4.813 + (0.9743 * \text{distance}) + (-0.2401 * \text{distance}_{sq}) + (-0.4245 * \text{centred\_gradient})$$

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

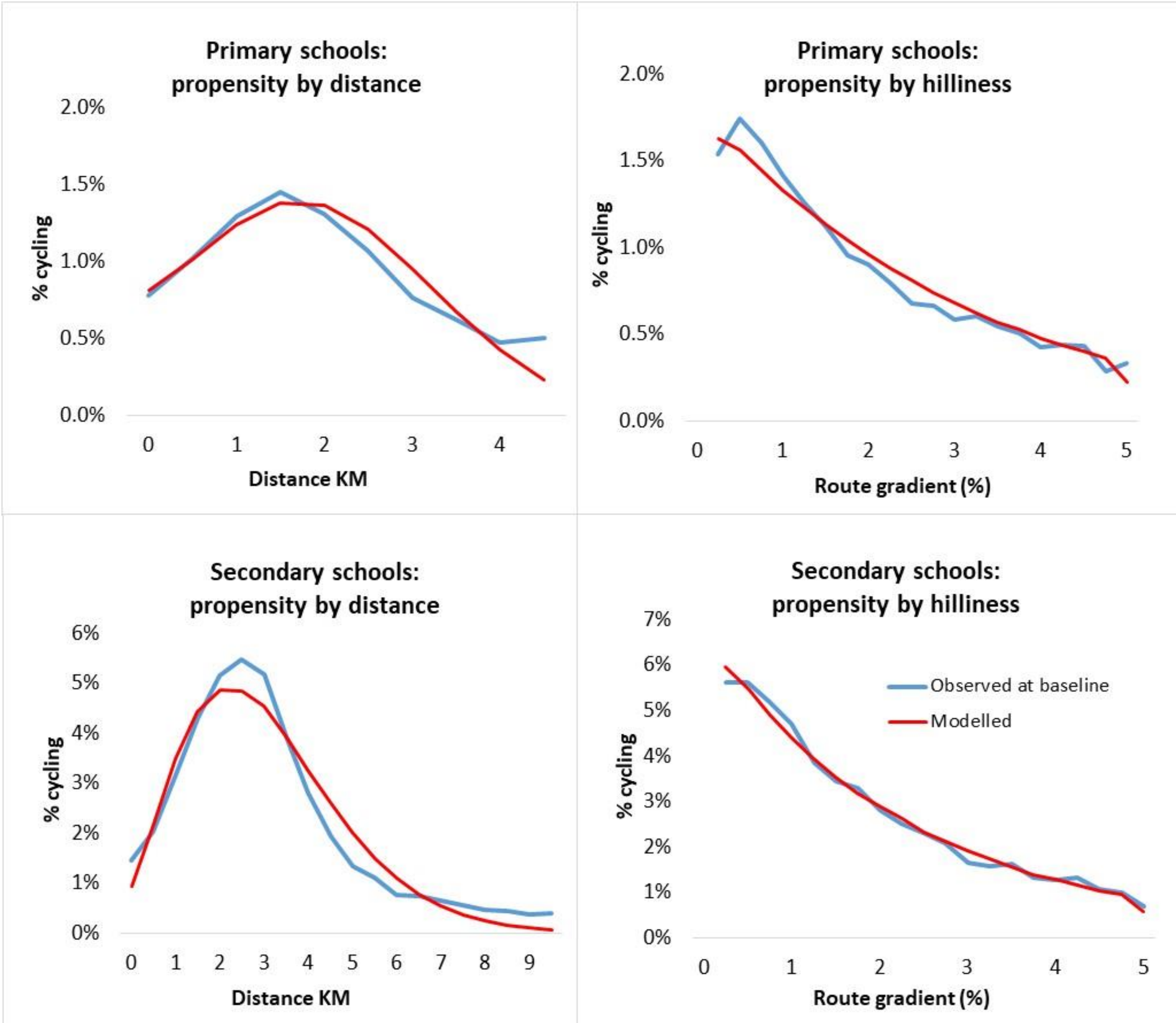
Equation 2.1 (secondary school children):

$$\text{logit (pcycle)} = -7.178 + (-1.870 * \text{distance}) + (5.961 * \text{distance}_{sqrt}) + (-0.5290 * \text{centred\_gradient})$$

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

where 'pcycle' is the proportion of cyclists expected in an OD pair (or, equivalently, the probability of cycling for an individual); 'distance' is the fastest route distance in km, '$\text{distance}_{sqrt}$' and '$\text{distance}_{sq}$' are, respectively the square-root and square of distance; and 'gradient' is the fastest-route gradient (centred on 0.63%). Equation 1.1 showed a very good fit to the observed data with respect to distance and hilliness (**Error! Reference source not found.**). Equation 2.1 showed relatively good fit to the observed with respect to distance and very good fit with respect to hilliness (**Error! Reference source not found.**).

**Figure 1: Observed versus predicted prevalence of cycling to school among 7,442,532 English school children, according to a) route distance and b) route gradient**




Sample restricted to primary school children travelling <5km to school and secondary school children travelling <10km

## 3. Modelling cycling across scenarios

We created 'Government Target' and 'Go Dutch' scenarios using an equivalent approach to what is done in the PCT-commuting layer. We subsequently added a further scenario, 'Go Cambridge'.

### i. Government Target scenario

The 'Government Target' scenario models a 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 . To model the total number of cyclists in the Government Target scenario propensity to cycle ('pcycle') in each OD pair was estimated using the equations set out in Section A2.1. This value was multiplied by the total number of children in the OD pair, this added to the recorded number of cyclists in the 2011 Census (see Table 1).

**Table 1: Summary of scenario generation rules, as applied to OD pairs <5km (primary)/<10km (secondary)**

| 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 | Recorded no. in Census 2011. | (no. children* Baseline propensity to cycle) + Column A | • Cap Column B1 at 100%. | Column B2 minus Column A |
| Go Dutch | Recorded no. in Census 2011 | no. children * 'Go Dutch' propensity to cycle‡ | • Set Column B1 as equal to Column A if B1 is less than A. | Column B2 minus Column A |
| Go Cambridge | Recorded no. in Census 2011 | no. children * 'Go Cambridge' propensity to cycle¥ | • Set Column B1 as equal to Column A if B1 is less than A. | Column B2 minus Column A |

† Using equations 1.1 + 2.1 or, equivalently, using equations 1.2 + 2.2 with 'dutch'=0

‡ Using equations 1.2 + 2.2 with 'dutch'=1

¥ Using equations 1.3 + 2.3 with 'cambridge'=1

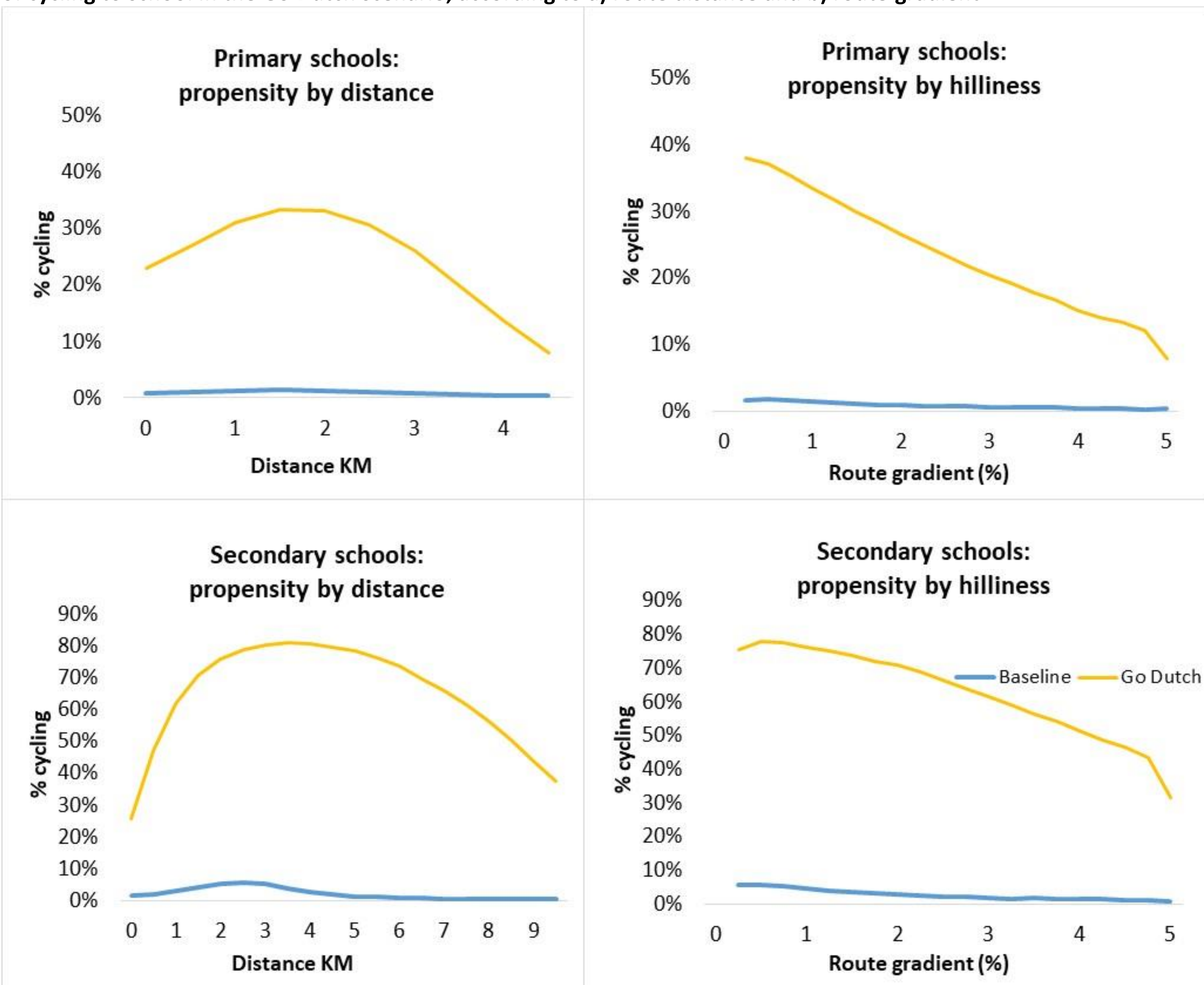
### ii. Go Dutch scenario

*Plain language overview*

While the Government Target scenario models relatively modest increases in cycling to school, the Go Dutch scenario is an ambitious vision for what cycling in England could look like. The 'Go Dutch' scenario models the level of cycling expected if English school children cycled to school as much as children in Netherlands, taking into account differences in the distribution of hilliness and trip distances. For this scenario, our approach was to start from the Equations estimating baseline propensity to cycle (Section A2.1) and add additional parameters corresponding to the increased in propensity to cycle among English versus Dutch children.

**Error! Reference source not found.** shows the distribution of 'Go Dutch' cycling propensity generated according to distance and hilliness. As it shows, cycling to school is transformed from being a very small minority of children to (at least at short distances) a large majority. These Go Dutch propensities are somewhat lower than the cycle mode shares observed in the Netherlands because they take into account the fact that distance is England is a hillier country. For example, the cycle mode share for primary school trips of distance 2.0-2.9km is 52% in the Dutch NTS data versus 32% in the Go Dutch scenario. For secondary school trips of distance 2.0-2.9km, the cycle mode share is 91% in the Dutch NTS data versus 77% in the Go Dutch scenario.

**Figure 2: Prevalence of cycling to school at baseline among 7,442,532 English school children, and modelled prevalence of cycling to school in the Go Dutch scenario, according to a) route distance and b) route gradient**




Sample restricted to primary school children travelling <5km to school and secondary school children travelling <10km

### *Technical details*

We estimated these additional parameters using trip-level analysis of the English and Dutch National Travel Surveys. We estimated the parameters for primary school children by restricting the analysis to trips for the purpose of education of <5km by children aged 3-10 (N=57,362 trips among 6,357 children in the 2010-2016 English data; N=39,835 trips among 16,528 children in the 2010-2016 Dutch data). We estimated the parameters for secondary school children by restricting the analysis to trips for the purpose of education of <10km by children aged 11-18 (N=54,246 trips among 6,411 children in the 2010-2016 English data; N=54,246 trips among 14,110 children in the 2010-2016 Dutch data).

In estimating the increased propensity to cycle among Dutch secondary school children, we included a main effect term and an interaction term with distance (as a linear term). In the primary school children we only included a main effect term, as the interaction was not significant. As hilliness data was not available in the Dutch survey, we weighted the data so that the English sample of school children lived in areas with the same hilliness profile as the Dutch children. This was done to allow comparisons that were not affected by differences in average hilliness between England versus the Netherlands.

In the primary school logit model, based on 97,197 trips among 22,885 English and Dutch children, the coefficient for a main effect with Dutch (versus English) status was 3.642. In the secondary school logit model, based on 83,481 trips among 20,521 English and Dutch children, the coefficient for a main effect with Dutch (versus English) status was 3.574, while the interaction term between Dutch status and distances was 0.3438. Adding these 'Go Dutch' parameters gave rise to the following propensity to cycle equations:

Equation 1.2 (primary school children):

logit(pcycle) = Equation 1.1 + Dutch parameters

logit (pcycle) = -4.813 + (0.9743 * distance) + (-0.2401 * $distance_{sq}$) + (-0.4245 * centred_gradient) + (3.642 * dutch)

pcycle = exp ([logit (pcycle)]) / (1 + (exp([logit(pcycle)])

Equation 2.2 (secondary school children):

logit(pcycle) = Equation 1.2 + Dutch parameters

logit (pcycle) = -7.178 + (-1.870 * distance) + (5.961 * $distance_{sqrt}$) + (-0.5290 * centred_gradient) + (3.574 * dutch) + (0.3438 * dutch * distance)

pcycle = exp ([logit (pcycle)]) / (1 + (exp([logit(pcycle)])

where 'pcycle' is the proportion of cyclists expected in an OD pair (or, equivalently, the probability of cycling for an individual); '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 0.63%); and 'dutch' is a binary variable that takes the value '1' for the Go Dutch scenario and '0' for all other scenarios.

### iii. Go Cambridge scenario

*Plain language overview*

In response to user feedback, we subsequently developed a further scenario 'Go Cambridge'. This scenario models the level of cycling expected if English school children cycled to school as much as children in the local authority of Cambridge, taking into account differences in the distribution of hilliness and trip distances.

This represents an intermediate level of change between the Government Target and the Go Dutch scenarios, because the proportion of children cycling to school in Cambridge is high (30% in the National School Census, twice as high as the second highest local authority), but is not at the level observed in the Netherlands (58% in the Dutch National Travel Survey). In this respect, the level of ambition represented by the Go Cambridge scenario in the schools layer could be seen as analogous to the Go Dutch scenario in the commuting layer.

Our method of modelling the Go Cambridge scenario was very similar to that we used for the Go Dutch scenario. Once again, we started from the Equations estimating baseline propensity to cycle (Section A2.1) and add additional parameters corresponding to the increased in propensity to cycle among children living in Cambridge versus children living elsewhere in England. The main difference is that in the Go Dutch scenario we estimated the additional parameters using trip-level analysis of the English and Dutch National Travel Surveys. By contrast, we were able to model Go Cambridge directly from the National School Census, as described in the next section.

### *Technical details*

We estimated the additional Go Cambridge parameters using the same National School Census data that we used in estimating the baseline propensity to cycle equations. The only difference was that we introduced a binary variable indicating whether the child lived in the local authority of Cambridge or not.

In estimating the increased propensity to cycle among primary school children in Cambridge, we included a main effect term and an interaction term with distance (as a linear term). We did this as the interaction was highly significant ($p<0.001$), and inspection of the data indicated a clear pattern for a higher relative propensity to cycle in Cambridge at higher trip distances. For example, for trips <0.5km there was an 11-fold difference between Cambridge and the rest of England; for trips 2-2.5 km there was a 23-fold difference; and for trips over 4.5 km there was a 33-fold difference. In the secondary school children, we only included a main effect term, as the interaction was only weakly significant ($p=0.012$) and not convincing upon visual inspection.

In the primary school logit model, the coefficient for a main effect with Cambridge (versus other parts of England) status was 2.334, while the interaction term between Cambridge status and distances was 0.2789. In the secondary school logit model, the coefficient for a main effect with Cambridge status was 3.049. Adding these ‘Go Cambridge’ parameters gave rise to the following propensity to cycle equations:

<u>Equation 3.1 (primary school children)</u>:

logit(pcycle) = Equation 1.1 + Cambridge parameters

logit (pcycle) = -4.813 + (0.9743 * distance) + (-0.2401 * $\text{distance}_{sq}$) + (-0.4245 * centred_gradient) + (2.334 * Cambridge) + (0.2789 * Cambridge * distance)

pcycle = exp ([logit (pcycle)]) / (1 + (exp([logit(pcycle)])

<u>Equation 3.2 (secondary school children)</u>:

logit(pcycle) = Equation 1.2 + Cambridge parameters

logit (pcycle) = -7.178 + (-1.870 * distance) + (5.961 * $\text{distance}_{sqrt}$) + (-0.5290 * centred_gradient) + (3.049 * Cambridge)

pcycle = exp ([logit (pcycle)]) / (1 + (exp([logit(pcycle)])

where ‘pcycle’ is the proportion of cyclists expected in an OD pair (or, equivalently, the probability of cycling for an individual); ‘distance’ is the fastest route distance in km, ‘$\text{distance}_{sqrt}$’ and ‘$\text{distance}_{sq}$’ are, respectively the square-root and square of distance; ‘gradient’ is the fastest-route gradient (centred on 0.63%); and ‘Cambridge’ is a binary variable that takes the ‘1’ for the Go Cambridge scenario and ‘0’ for all other scenarios.

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

### i. Modelling scenario mode shift in walking and car driving

The next stage was to convert the cycling uptake estimates into estimates of mode shift, and from this estimate health and carbon impacts. Specifically, to estimate the health impacts of our scenarios, we needed to estimate the number of pupils who had previously travel to school on foot. Similarly, to estimate the carbon impacts of our scenarios, we needed to estimate the number of new cyclists who had previously travelled by car. As in the PCT-commuting layer, we assumed that within any given OD pair children were equally likely to shift to cycling from any baseline mode, and therefore the mode shift was proportional to mode share at baseline.

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 – i.e. that any current cyclists would be equally distributed across the other modes currently observed. When estimating mode split in the 'no cyclists' counterfactual in the 2901 of OD pairs that at baseline consisted entirely of cyclists, we assumed a mode split of 49% walking, 31% car drivers and 20% other modes. These percentages correspond to the observed mode split among the 794 OD pairs in which 50-94% of individuals cycled in the 2011 Census.

### ii. Estimating the physical activity impacts

In this Manual we provide an overview of our methods: for full details see Goodman et al. [1].

We quantified physical activity impacts in terms of physical activity energy expenditure. Physical activity energy expenditure when walking and cycling was estimated using marginal METs. A MET or 'Metabolic Equivalent Task' is a measure of energy expenditure, with a value of 1 corresponding to resting. We retrieved MET values for children using the Youth Compendium of Physical Activities [3]. We converted these MET values into marginal METs (mMETs) by subtracting 1, to generate a measure of additional energy expenditure above resting.

Our first step in estimating the change in physical activity was to calculate the average annual physical activity energy expenditure a child who was cycling to school in each OD pair. We then estimated total additional physical activity per week in each OD pair that resulted from more children taking up cycling in a given scenario.

In calculating the physical activity impacts of a given scenario, one needs to offset the physical activity gained through increased cycling against the physical activity lost through any decrease in walking, since some new cyclists will formerly have walked to school. Among these former pedestrians, we calculated the weekly physical activity energy expenditure that was displaced by cycling. Note that because cycling is more energy-efficient than walking, it is possible for the total physical activity in a given OD pair to decrease in a scenario where cycling increases, if most of the children who start cycling previously walked.

We then calculated the net change in physical activity at the level of the OD pair. To put these findings in context, we also present this net change in physical activity energy expenditure relative to the estimated total baseline active travel energy expenditure in each OD pair.

Finally, note that among children in the 'primary' schools, 8% are aged 2 or 3. At this age, it is likely that many of the children reported to be cycling or walking to school will in fact be being cycled/carried/pushed in a pushchair for at least some of the distance. Among primary school children, our calculations may therefore overestimate to some extent both the physical activity increase due to increased cycling and the physical activity decrease due to decreased walking. The likely result is a modest overestimate of the total net change in physical activity. On the other hand, we are not seeking to model any of the potential health benefits to adults of switching to using a bicycle to escort younger children to school - so in that sense our model is conservative in terms of the scope of which physical activity impacts it estimates.

### iii. 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 users * former distance travelled by former car users * mean cycling to school trips per cyclist per week * 52.2 * mean car driver escort trips per child car trip to school * CO2-equivalent emissions (in kg) per kilometre

The change in the number of children travelling by car was estimated using the mode shift calculations described above. Their average former distance was assumed to be equal to the new 'fastest route' distance travelled by the children cycling to school in that OD pair. The mean number of cycling trips to school per cyclist per week was estimated to be 2.3 for primary school children, meaning that the mean number of cycling to school trips per year was 2.3 * 52.2 = 120.1. The corresponding value for secondary school children was 5.1 * 52.2 = 266.2. The average number of car driver escort trips for each of these child car trip to school was estimated as 1.2. Note that this value is greater than 1 because many parents drive their child to school and then make the return trip home, but is less than 2 because some parents drive two or more children to school in the same trip and/or drive to the school and then go on to make a trip for different purpose. 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 [4].

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

As in the PCT commuting layer, we aggregated the results from OD pairs to give zone-level results, school-level results, and to create a route network: for detail see User Manual C1

## 6. References


1. Goodman, A., et al., *Scenarios of cycling to school in England, and associated health and carbon impacts: Application of the 'Propensity to Cycle Tool'.* . Journal of Transport & Health, 2019. **12**: p. 263-278.
2. 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.

3. NCCOR. *Youth Compendium of Physical Activity*. 2017.
4. 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].