CAMBRIDGESHIRE
& PETERBOROUGH
COMBINED AUTHORITY

Draft Methodology for Prioritising Bus Routes for Investment – Route Design Standard

September 2022

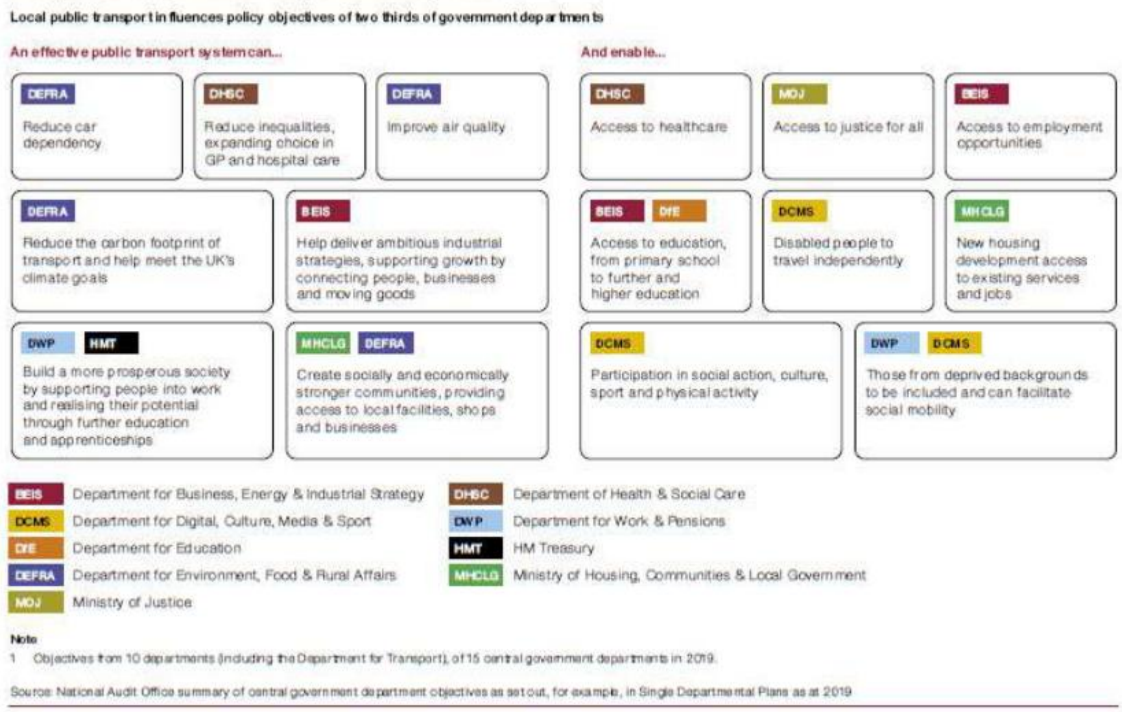
Introduction

The Cambridgeshire and Peterborough Combined Authority requires a method of assessing which bus routes to subsidise that considers a wide range of parameters including value for money, support for rural areas, support for deprived areas and contribution to sustainable travel to work, education, and access to vital services such as hospital and GPs.

A brief review of literature shows that there is no standard method for doing this. In the US[1] there was a study that split route assessment between a group of separate standards. Of relevance to this methodology was the proposed 'Route Design Standard' considering population density, employment density and equity of access alongside an 'Economic Design Standard' that considered aspects such as costs per passenger mile, subsidy per passenger per mile etc. There would then be a subsequent set of standards applied to the design of the timetable.

Other research focuses on evaluating the effectiveness of existing services against a specific policy objective. For example, the assessment of the 'Better Bus Area' initiative focused on the data needed to assess effectiveness against the stated aim of *"support the local economy and facilitate economic development through the improvement of local bus infrastructure, patronage numbers and services".* Other work by the National Audit Office (NAO) starts from a similar place, posing the question *'Why buses are important, and to whom?'*[2]. The question is then answered in relation to supporting the achievement of government policy.

Figure 1
Local public transport and wider government objectives



[1] BUS ROUTE EVALUATION STANDARDS (trb.org)
[2] Improving local bus services in England outside London (nao.org.uk)

Based on this simple read through the proposed methodology is based on the following:

1. Having an approach that acknowledges each route needs to meet an 'Economic Design' standard and a 'Route Design Standard'.

    a. The Economic Design Standard (addressed elsewhere) – but to include population data.

    b. The Route Design Standard being based on how well subsidising the route helps to meet the policy objectives of the combined authority as defined by the Sustainable Growth Ambition Statement.

Route Design Standard – Methodology

The Sustainable Growth Ambition statement (SGAS)[3] was reviewed alongside the proposed monitoring paper for the SGAS[4] and the draft Local Transport and Connectivity Plan (LTCP)[5]. The main points of each were then used to draw up a series of headings under which data could be identified for the assessment of the value of each route toward supporting the aims of the Combined Authority. This was done noting that not every aim within the SGAS is immediately relevant to the direct provision of bus services e.g., building a knowledge-based economy.

The following list was chosen:

1. *Reducing inequalities*: Supporting disadvantaged communities, with access to education, employment, and work.

2. *Productivity:* Supporting people to connect with larger labour markets / better employment.

3. *Connectivity:* Supporting links between places and access to services (e.g., Hospital).

4. *Climate:* Potential future value of the route for supporting mode shift from car to bus.

Against these headings a data list was drawn up. There were restrictions on the data that could be

- It needed to be available at a small area level to enable a fine grain analysis of each route.
- The data needed to be recognised as authoritative.
- The data needed to be relatively contemporary (noting the time delay in releasing national statistics).
- Data needed to available within the CPCA Corporate GIS system for analysis *(see appendix one and two of this report for a detailed account of the analytical process).*

In addition, it was thought that some data could be used outside of a ranking calculation to provide useful context. For example, to understand the rural / town / urban coverage of the routes, or to understand the relative size of the labour markets (not available at small area level) that the route was linking people to.

[3] Document.ashx (cmis.uk.com)
[4] Agenda Item No (cmis.uk.com)
[5] Draft-LTCP.pdf (yourltcp.co.uk)

Theme	Reducing Inequalities			Productivity		Connectivity			Climate
Data Set	Index of Multiple Deprivation 2019 – IMD Score	Index of Multiple Deprivation 2019 - Ed–cation and Skills Domain	*Context only* Rural / Urban Classification 2016[6]	Income Estimates of small areas 2018[7]	*Context only* Relative labour market size[8]	Journey time statistics of England and Wales 2019[9]	Index of Multiple Deprivation 2019 – Access to services Score[10]	Significant service points on route (based on Ordnance Survey data) - Hospital - Higher Education (inc sixth forms) - *Any others?*	Relative traffic volumes on route – based on Basemap 2019 (TRACC data)
Route x						*A choice of one of the above – to be tested*			

[6] Rural Urban Classification - GOV.UK (www.gov.uk) Awaiting an update based on the 2021 census (current classification is 2011 based)
[7] Income estimates for small areas, England and Wales - Office for National Statistics (ons.gov.uk) noting that 2018 is the latest release (made in 2020). A calculation will be considered to understand the ratio of earnings between different areas.
[8] Dft Journey time statistics include modelled employment centre data – ratio of connection could be used e.g. a route connecting a small labour market area to a much bigger one.
[9] Journey time statistics, England: 2019 - GOV.UK (www.gov.uk) Published in 2021
[10] This is to be tested for overlap with the DfT Access to Service model used to develop the Journey Time Statistics. There may be so much overlap that one indicator is sufficient

Appendix 1

Capturing Geographical Information

Appendix 1: Capturing Geographical Information

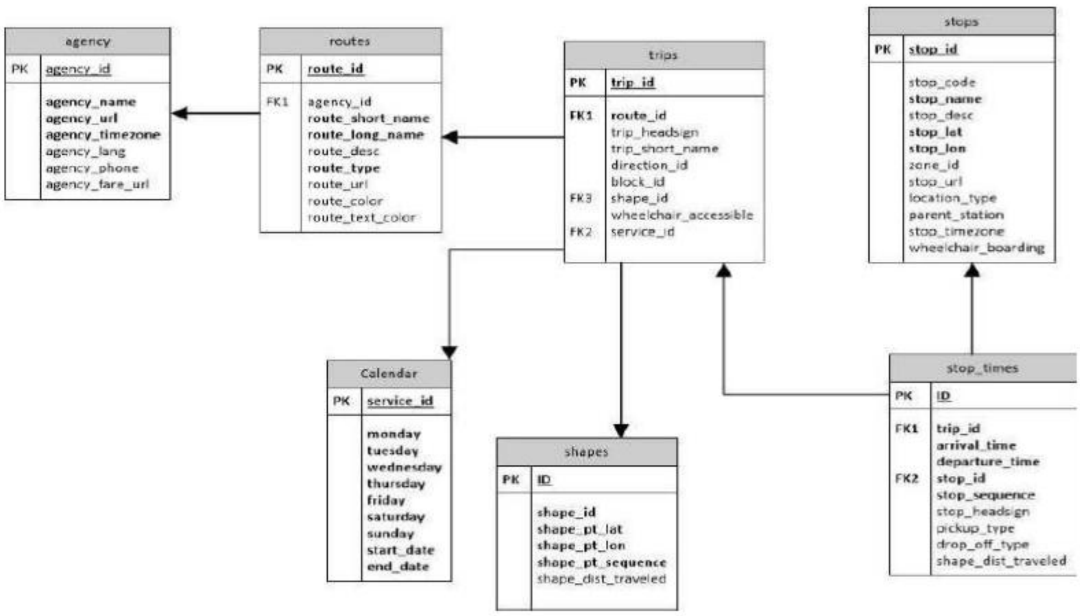
Initially accessing the information required for the assessment was straight forward, however the processing and management of the data requires several steps each of which require separate quality assurance to ensure that no errors are introduced.

1. The most recent bus service data is downloaded in GTFS[11] format data from the DfT [All timetables data (dft.gov.uk)](https://www.dft.gov.uk).

   Most of the Cambridgeshire routes are part of the East Anglian Dataset whereas the routes starting in Peterborough centre on the East Midlands Dataset.

2. The data is not comprehensive. It is only as good as the information that is submitted by the bus companies and commissioning authorities. Past reviews have shown that routes are sometimes missed or incomplete. This is so, for those delivered by smaller operators.

3. GTFS data has a very specific format and relationship between the tables of information.



One very specific challenge in this case is to extract the right 'Shapes' based on the 'Routes' that are being analysed. This is because there isn't a direct relationship between the two within the GTFS format. Rather the relationship is made via the 'Trips' table. When running the join care needs to be taken to extract the desired shape as there may be multiple shapes for each route / trip based on real world differences in the timetable e.g., some bus trips having extensions at peak times.

[11] General Transit Feed Specification

4. The data was processed using a GTFS loader12 in QGIS (Desktop GIS) to dissemble the GTFS file. Into its various tables of routes / trips / times / shapes.

5. The selection of Cambridgeshire specific routes is then made (the process is doubled up as Peterborough routes are on the East Midlands dataset). Identification needs to be done with care as there are multiple routes with the same shortcode e.g., there are eight route 5s in the database. The field 'tripheadsign' can be used to separate these out e.g., selecting the route 5 with the Cambridge Headsign rather than Ipswich.

6. As per section 3 above, the query relationship between the tables is then carried out within QGIS. The linking code lines are randomly generated unique long integers, and the route numbers, operator codes etc are also integers / short text. Therefore, additional information is introduced at this stage to enable the data to be understandable, for example a common-sense descriptor (created by the public transport team) to go alongside the route number.

7. The process was then checked for quality assurance purposes. As per expected the query structure returned ten correct routes and eleven with errors. The errors included incomplete routes or there being multiple (possibly historic versions of the same route on the database) and routes with alternative spurs depending on the time of day or alternative routes depending on 'there or back'. Corrections were then made to the queries.

8. Where there were partial closures to routes of, they were missing entirely from the GTFS download (e.g., V5) then these were captured by hand drawn lines (with reference to the timetable / stop information).

9. The resulting GIS dataset is then uploaded into ArcGIS online for display, sharing and analysis.

[12] GTFS Loader — QGIS Python Plugins Repository This plugin allows to load a GTFS ZIP file, that will be extracted into the GeoPackage with individual tables. For stops and shapes it creates vector layers.

Appendix 2:
Query Building

Appendix 2: Query Building

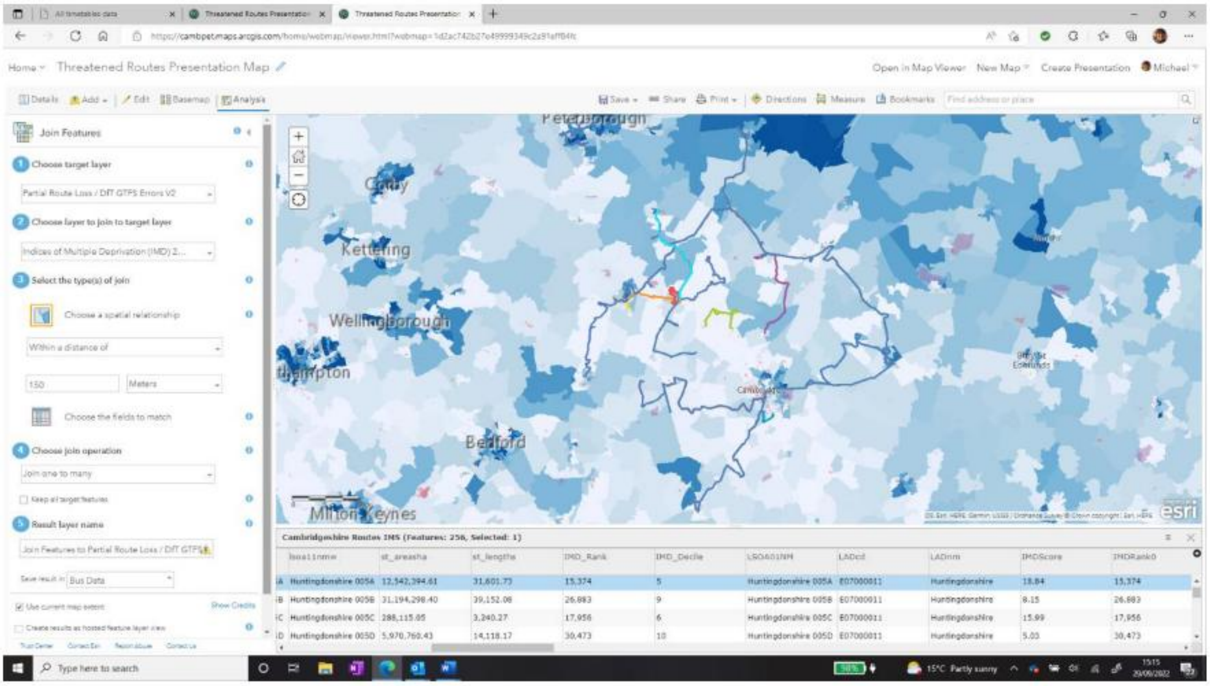


Figure 1 - Query building to identify IMD data for each route

As per the screen shot above a query was built within ArcGIS Online (AGOL) to identify where each bus route passes within 150m of a Lower Super Output Area (LSOA) as the crow flies, and to gather up the Index of Multiple Deprivation (IMD 2019) data for that LSOA.

The distance of 150 metres was selected subjectively, but with reference to the limited study literature available on the distance people are willing to walk to a bus-stop[13]. Keeping in mind the route only need be with 150ms of an LSOA boundary, with the LSOA extending an even great distance. 150m appeared to be a compromise between selecting too narrow an area or too much (noting studies identify a range of walk to stop of 100 to 700 metres, depending on area geography).

It should be noted that the route was used rather than individual bus stops so as to identify fully the potential reach of the service given future options may consider adding stops to routes to gain additional value.

An illustration is provided overleaf (see figure 2) for the results of the selection process in Peterborough. For each of the three routes:

- Route 29 was within 150 metres of the boundary of 23 LSOAs
- Route 23 was within 150 metres of the boundary of 12 LSOAs
- Route 24 was within 150 metres of the boundary of 15 LSOAs

[13] (PDF) Acceptable walking distance accessible to the nearest bus stop considering the service coverage (researchgate.net)

As the routes were relatively close some of the LSOA selected were the same. This is not seen as a problem, but rather a reflection of the ground conditions.

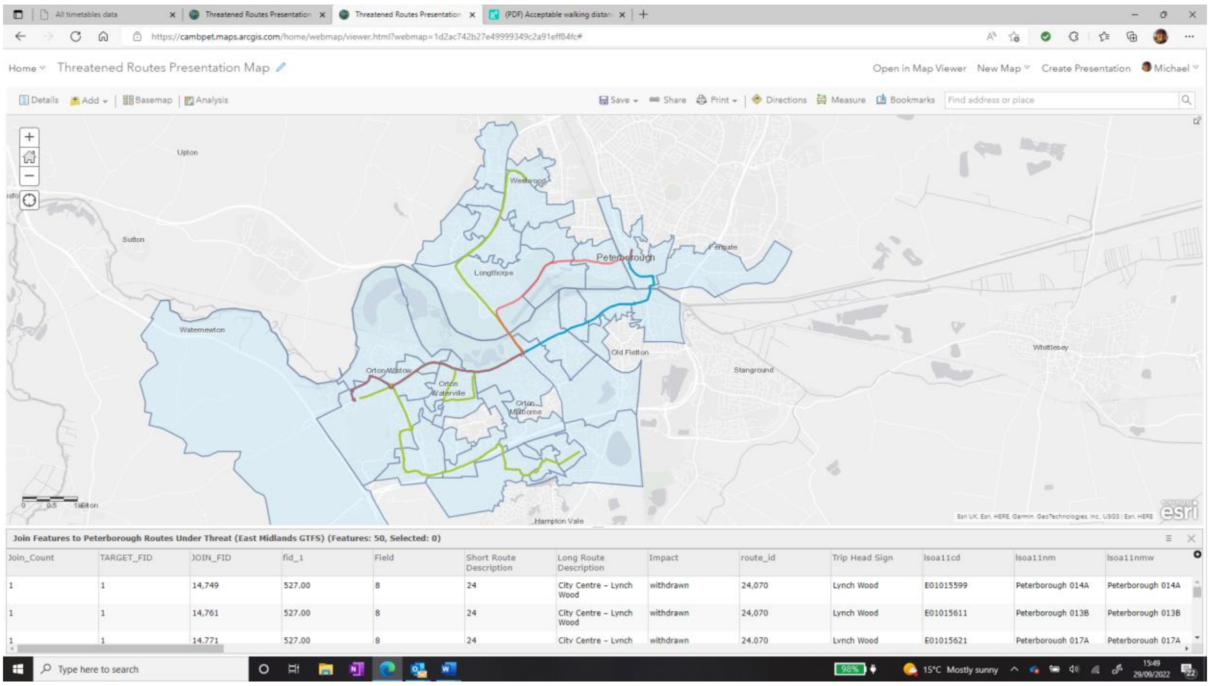


Figure 2 - Selection of LSOAs based on the three threatened Peterborough City Routes

The results of the example query were as per the table below with route 29 running closer to more deprived areas compared to routes 24 and 23. (Minimum and Maximum values were included to ensure no exceptionally deprived areas.

Bus Route	Average of IMD Score	Max of IMD Score	Min of IMD Score
23	15.31	38.37	8.15
24	19.90	41.70	6.00
29	24.48	57.29	8.03

Finally, a visual sense check was carried out on the selection process. This identified a problem in a very limited number of cases where a route ran a distance along dual carriageways / motorways where there wasn't the possibility of stopping, with the route not serving the communities it went through. In those cases, e.g., route 29 above in Peterborough a separate custom query was built and used to adjust the captured data.

It should also be noted that the analysis process could have been exercised using bus stop data. Which would help resolve the problem identified above. However, the analysis by stop, particularly where those stops are close together would yield a significantly larger dataset with a lot of overlap and duplication (on a many to many query). The subsequent cleaning of the data would greatly lengthen the analysis process, so given that speed of response was a priority a simpler route-based method was chosen.