|  |  |
| --- | --- |
| ***Transport for Greater Manchester***  **Review of Look-up Tables from Centro Smartcard Data**  *17th February 2014* |  |

**EXECUTIVE SUMMARY**

Current DfT Guidance on concessionary travel reimbursement recommends that the average fare forgone is calculated using a look-up table built from smartcard data collected by four Lancashire districts in 2009. The Guidance states that a Travel Concession Authority may replace the NoWcard look-up table with one built from its own data, or taken from a similar area.

Centro has implemented smartcard recording of concessionary travel usage and has been able to build look-up tables based on its own area. Transport for Greater Manchester is currently unable to do the same because some of its operators have yet to implement smartcard systems sufficiently. However, it seems highly likely that it is more appropriate for TfGM to use Centro’s look-up table for calculating reimbursement rather than the NoWcard look-up table.

We have examined the raw data used by Centro and the method it has devised to construct its look-up table, and also used the raw data to construct our own equivalent to the Centro look-up table. The Centro data represents nearly 95% of all concessionary journeys in the period covered, which at 26 weeks is substantially greater (and hence more robust) than the 5 weeks of data from which the NoWcard table was built.

The process used by Centro to build its look-up table is consistent with the key principles on which such tables should be built, and the overall characteristics of the data set and resulting look-up table are as expected. However, there are differences in the detail of Centro’s process compared to that used in building the original NoWcard look-up table, which we have used in building our own version of a look-up table based on identical West Midlands data.

We have tested the alternative look-up tables by using them to calculate discount factors for some illustrative ticket prices supplied to us by TfGM. The tests demonstrate that using either of the West Midlands-based look-up tables (that built by Centro, and that built by ourselves) results in significantly higher discount factors than would be found from the NoWcard look-up table. The two Centro-based look-up tables provide similar results, with the table built by Centro giving slightly smaller discount factors (and consequently is more favourable to operators).

2011 Census data demonstrates that there are significant differences in the socio-economic and geographic characteristics of the 4 NoWcard districts and the PTEs as a whole, and in particular Greater Manchester and the West Midlands. These almost certainly mean that discount factors calculated from West Midlands data are much more appropriate for use in Greater Manchester than the NoWcard look-up table. In contrast to the NoWcard districts, Greater Manchester and the West Midlands have:

* much higher population densities;
* a very much smaller proportion of the older population living in rural (i.e. non-urban) areas;
* lower car ownership amongst the older population
* much greater use of the travel concession per head of elderly population

On most measures, the West Midlands and Greater Manchester are closely grouped, with differing levels of car availability appearing to account for much of the spread in relative concession usage both between the PTEs and also the NoWcard districts.

Ideally, TfGM would calculate reimbursement using look-up tables constructed from its own smartcard data, but this is not yet possible. It is highly likely that were TfGM to use look-up tables derived from West Midlands data rather than the NoWcard data, the results would be much closer to those which would be obtained if it had its own local data. Either of the West Midlands based look-up tables derived by Centro and ourselves could be used by TfGM with confidence, but since the former gives outcomes that are more favourable to operators, its selection would be least vulnerable to criticism of bias by the bus operators.

Our recommendation is therefore that TfGM make use of the Centro-derived, West Midlands-based look-up tables in calculating reimbursement for bus operators in Greater Manchester, rather than the look-up table based on the 4 NoWcard districts incorporated in the DfT Calculator.

**INTRODUCTION**

Current DfT Guidance on reimbursement of concessionary travel for older and disabled people[[1]](#footnote-1) recommends that Travel Concession Authorities use what it calls the “Discount Factor Method” for deriving an estimate of the Average Fare Forgone – the fare that would have been paid by concessionary passholders if the concession did not exist. The Discount Factor Method relies upon a look-up table that summarises the frequency of journey-making by concessionary pass holders, which is used to infer how much passholders would use cash fares, or daily or weekly discount tickets, in the absence of the concession. The look-up table must be built from smartcard data, which provides the only source of information on trip making that has the potential to be sufficiently complete.

In 2009 when the DfT methodology was developed, the only available source of data was from smartcards issued by the NoWcard consortium of local authorities in Lancashire and Cumbria. The current version of the DfT Calculator, which remains the recommended default mechanism for calculating reimbursement, still uses the look-up tables derived from the early 2009 NoWcard data. This was based on five weeks of data from four Lancashire Districts.

The DfT recognises that NoWcard data may not be fully representative of all areas in England, and that as smartcard implementation becomes more widespread, the feasibility of building alternative look-up tables has increased. Its Guidance states that ”Where such data is available TCAs may wish to replace the NoWcard data in the model with their smartcard data or smartcard data from another area which they can demonstrate to be representative of their own area.”

The robustness of the estimates of discount factors that emerge from the Discount Factor method is entirely dependent upon the extent to which the smartcard data that can be captured represents the vast majority of the concessionary journeys made by the passholders living in a given area. If a significant proportion of concessionary journeys are not recorded as smartcard transactions, then look-up tables built from the smartcard data alone could significantly under-represent journey frequencies, and hence underestimate discount factors and over-estimate reimbursement for revenue forgone. Some smartcard under-recording is inevitable because of transaction failures or other reasons, and if this is modest, it can be corrected by factoring up the smartcard data. But because of the very “lumpy” nature of the journey frequency distributions and how these are referenced in the DfT methodology, reliance on factors that are too large creates the danger of unreliable and inconsistent results, and should be avoided.

In the West Midlands, Centro has had access to substantial quantities of smartcard data for some time, and initially built a look-up table from its own data in 2011. An updated West Midlands-based look-up table has now been built using data from the first six months of the 2013-14 financial year. In this period, smartcard data represented nearly 95% of Centro’s best estimate of the total number of concessionary journeys made in that period.

In Greater Manchester, bus operators are still some way off from recording the vast majority of concessionary journeys via smartcards, and so it is not currently possible to build a robust look-up table from Greater Manchester data. However, there are a number of reasons why it is likely that a look-up table based on Centro data is more appropriate for use in Greater Manchester than one based on NoWcard data.

The Centro look-up table was produced by Centro using internal resources, and without a view to it being used more widely by other TCAs outside the West Midlands. This note is the output from a commission by TfGM to put Centro’s look-up tables under independent scrutiny, to review the data and process used to derive them, and to examine the case for TfGM using tables derived from concessionary journeys made in the West Midlands for reimbursement of bus operators in Greater Manchester.

The author is extremely grateful for the assistance given by Centro, through the provision of data and discussion of technical issues, and to TfGM for guidance and feedback. However, the current note is the sole responsibility of the author.

**DATA AND CONSTRUCTION PROCESS**

Centro extracted information about the older and disabled concessionary journeys starting in the West Midlands, and made using ENCTS passes issued to West Midlands residents, for the period from 1st April to 30th September 2013 – the “Reference Period”. The data extract was intended to include all relevant smartcard-recorded concessionary journeys. It is worth noting that the length of Reference Period is substantially greater than was the case for the NoWcard look-up table (26 weeks as opposed to 5 weeks) and consequently the Centro data represents a significantly more robust basis for a look-up table than the original NoWcard data.

All major and most minor bus operators in the West Midlands are smart-equipped and record and report concessionary journeys through the back-office operated on behalf of Centro. For the 6 months used for the data extract:

* 33.9 million concessionary journeys were reported as being carried by bus operators in the West Midlands
* Of these, 32.0 million concessionary journeys were recorded using smartcards
* The bus operators reported that an additional 1.9 million concessionary journeys were manually recorded (i.e. not recorded through the smartcard system) representing 5.6% of all concessionary journeys.
* About 80% of the manually recorded journeys arose from machine failures and other problems with smartcard transactions, and the balance were reported by operators who were not smart equipped.
* On the basis of the smartcard data, these journeys were made using 325,000 distinct passes (each of which was used at least once during the Reference Period). This represents about two-thirds of the total number of passes in circulation during this time (488,800) – implying that a third of passholders made no use of the concession during the six month period.

Individual journey transactions were summarised into the total number of concessionary journeys made by each passholder on each day in the Reference Period. Passes are uniquely identified by a pass number, which is read and recorded by the on-bus smartcard equipment. It is likely that over the course of the Reference Period, more than one pass (and hence pass number) might be issued to some individual passholders, for example as a result of replaced passes. Although potentially leading to some ambiguity in the interpretation of passholder-related statistics, this is not likely to significantly affect the robustness of the look-up tables derived from the data, which are constructed by looking at the journeys made using distinctly numbered passes in individual calendar weeks, in isolation. Note that whereas the full data extract covered the six calendar months which represent 26 weeks and one day, the need to analyse the data in whole-week blocks led to look-up tables being built from data for the period 1st April – 29th September 2013.

It should be emphasised that the volume of data that is potentially available for analysis is extremely large, and this gives rise to a number of practical problems arising from computing limitations. These required Centro to produce an interim file from the raw data, which is used to summarise journey allocations in blocks of whole weeks (e.g. five weeks). The Centro process has been examined and in principle correctly assigns journeys, and counts imputed ticket sales, to the different ticket types for each realisation of day and weekly price ratios. However, as would be expected, there are detailed differences between the methodology used by Centro, and that used to derive the original NoWcard look-up table (to which Centro did not have access at the time).

These differences concern the treatment of fractional parts of journeys, which arise from the application of an expansion factor to reflect the small proportion of concessionary journeys not reported through the smartcard back-office systems. Not allowing for unreported journeys is likely to bias resulting discount factors, but the existence of non-integer journeys numbers creates some intuitive difficulties when building look-up tables. The Centro approach is to introduce some rounding in the process, whereas the NoWcard method does not. Both approaches represent a compromise in reconciling the need to accommodate non-integer numbers with the fundamental philosophy behind the method, but the consequence is that the resulting look-up tables gives rise to slightly different discount factors. Those derived from the Centro method are likely to be lower (i.e. more favourable to operators) than those derived using the NoWcard method, if applied to identical “raw” data.

To understand the scale of this sensitivity, and to confirm in an overall sense the correctness of the Centro methodology, we have built our own version of a look-up table from Centro’s data, with a methodology based on that used to construct the NoWcard look-up table. The input data for both was identical, which was provided by Centro to Minnerva in the Excel file named “201310 Jny Concentration v2 – Anon.xlsx”.

The Minnerva analysis exclusively used Centro’s data on journey numbers by individual passholders for each day. 324,850 individual pass numbers[[2]](#footnote-2) were identified, for each of which at least one journey was made on one of the 182 days in the data set. After expansion to allow for non-smartcard recorded journeys, this data set represents 32,620,484 concessionary journeys made in the 26 complete weeks ending 29th September 2013.

Appendices 2 and 3 show fragments of the look-up tables derived by Centro and Minnerva from the Centro data. The full look-up tables extend to a day ticket price ratio of 10 and a weekly ticket price ratio of 40, but the fragments illustrated are for price ratios of up to 3 and up to 20 for day and week tickets respectively. To provide interim logic checks and facilitate processing, the Minnerva analysis divided the data into two 13 week periods for which separate tables were built. The fragment of table shown in Appendix 3 is the aggregated 26 week version of the table derived by adding together the two 13 week tables.

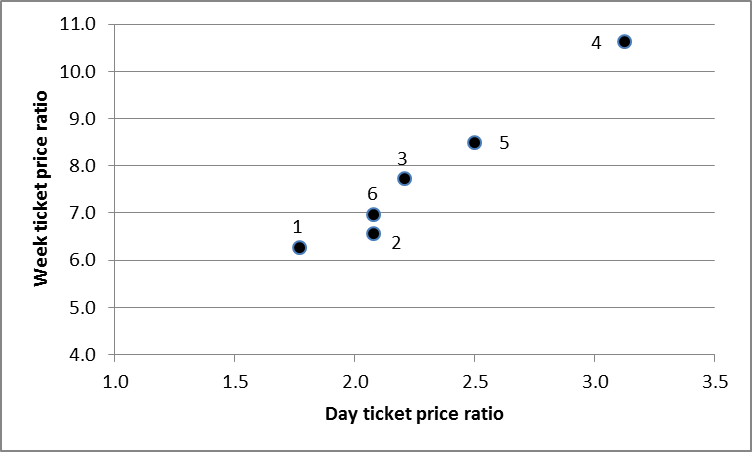
**DISCOUNT FACTOR IMPLICATIONS OF ALTERNATIVE LOOK-UP TABLES**

TfGM have estimated the average cash fare per journey, and the average price of day tickets and weekly tickets for six different bus operators serving Greater Manchester. These estimates are used in this note to illustrate the average fare implications of alternative look-up tables, but otherwise are not intended to have any other significance. They are summarised in Table 1 below. Individual operators have been identified using an arbitrary number in order to preserve anonymity.



**Table 1 Illustrative ticket prices for different operators in Greater Manchester, 2013-14 prices**

Cash fares vary by about +/- 15% between operators, whereas the prices of day tickets and weekly tickets vary slightly less. It is interesting to note that when measured as price ratios (the number of journeys that need to be made on a day or weekly ticket to become equivalent to the cash fare), day and weekly ticket prices are strongly correlated across all operators i.e. relatively cheap day tickets are associated with relatively cheap week tickets, and likewise for more expensive day and weekly tickets. This is shown in the scatter diagram in Figure 1, where the numbers show the operator number.



**Figure 1 Relationship between day and ticket price ratios**

So compared to the price of a cash fare, Operator 1 has relatively inexpensive day and weekly tickets, whereas Operator 4 has equally expensive day and weekly tickets. The expectation is therefore that the discount factor will be larger (i.e. a greater proportion of passengers would purchase discount tickets in the counter-factual) in the case of Operator 1 than Operator 4.

Three alternative look-up tables have been used to calculate average fare discount factors from these illustrative ticket prices. The look-up tables are as follows:

* NoWcard look-up table as incorporated in the current version of the DfT Calculator. This is based on five weeks of data collected in four Lancashire districts in early 2009. It represents the default look-up table that must be used to calculate a discount factor, if a Travel Concession Authority does not have access to an alternative, and wishes to use the “Discount Factor” method.
* The Centro 2013 look-up table built by Centro from smartcard data for the West Midlands covering a six month period from April 2013.
* The alternative look up table built by Minnerva from the same six months data for the West Midlands.

There are a number of ways of summarising the output from the application of the Discount Factor method. The value that is used to calculate the best estimate of the average fare forgone, and hence reimbursement for revenue forgone is the “final” discount factor. Within the DfT Calculator, this is applied to the average cash fare per journey that was input into the discount factor calculation. The “headline” impact of the alternative look-up tables, using the DfT calculation has been illustrated in the following table.



**Table 2 Final discount factors (DfT calculation) for different operators using alternative look up tables**

Looking at the results for individual operators, the NoWcard look-up table leads to discount factors that are typically half of those estimated from either of the look-up tables based on West Midlands data. The look-up table built by Centro gives discount factors that are up to 1 percentage point different to that estimated by the Minnerva-derived table. This pattern is consistent across all the operators shown, though it is likely to be less so were day and weekly ticket price ratios less strongly correlated.

It is clear from the above that look-up tables sourced from different areas will lead to significantly different estimates of discount factors for a given set of input fare and ticket prices. This is to be expected, because the propensity to use public transport will vary from area to area for a combination of reasons, and this will inevitably be reflected in the frequency with which individual passengers (both concessionary and non-concessionary) make journeys. That in turn will influence the distribution of passengers between ticket types that is implicit in the look-up table.

It should also be noted that we have some concerns that the detailed arithmetic built into the DfT Calculator spreadsheet may incorrectly compute the final discount factor quoted in Table 2. Although what appear to be errors go some way to cancel each other out, in our opinion the net effect is that it is probable that the calculated final discount factor underestimates the correct value. If this is the case, then the differences between the use of NoWcard and West Midlands-based look-up tables would be even greater than that shown.

From the perspective of Transport for Greater Manchester, look-up tables derived from concessionary journeys in its own area would give most assurance that it’s reimbursement calculations delivered “no better off/no worse off” payments. However, it is currently not possible to build reliable estimates of look-up tables from Greater Manchester smartcard data because bus operator implementation of smartcard recording of concessionary journeys is significantly incomplete. In the absence of look-up tables from its own area, the next-best option for TfGM is to use a look-up table from another area.

The inclusion of the Nowcard-based look-up table in the DfT Calculator reflects the fact that the vast majority of Travel Concession Authorities will not be in a position to build their own look-up tables from smartcard data generated within their own areas. However, the availability of look-up tables from the West Midlands, and potentially from other PTEs, gives TfGM a choice. The likelihood is that in general, large urban areas such as those for which the PTEs are responsible will have more in common with each other than with the Lancashire districts from which the NoWcard data was obtained. A number of key indicators confirm that this is the case, as is now discussed.

**COMPARISONS OF KEY CHARACTERISTICS OF THE DIFFERENT AREAS**

The main difference between the large conurbations such as the West Midlands and Greater Manchester, and the NoWcard districts, is greater population density, spread more consistently over a large area. This creates the conditions under which public transport services can be provided economically, and which will be used by a greater proportion of people, more frequently. Greater frequency of use will in turn provide more scope for passengers to make use of discount tickets, and will encourage operators to make such tickets available – a tendency that will be enhanced by the greater level of inter-operator competition that is likely to be more sustainable with a larger overall market for public transport use.

Some of the differences in the socio-economic characteristics of the PTE areas and the NoWcard districts[[3]](#footnote-3) can be identified from 2011 Census data. Selected key indicators are summarised in Table 3 below.

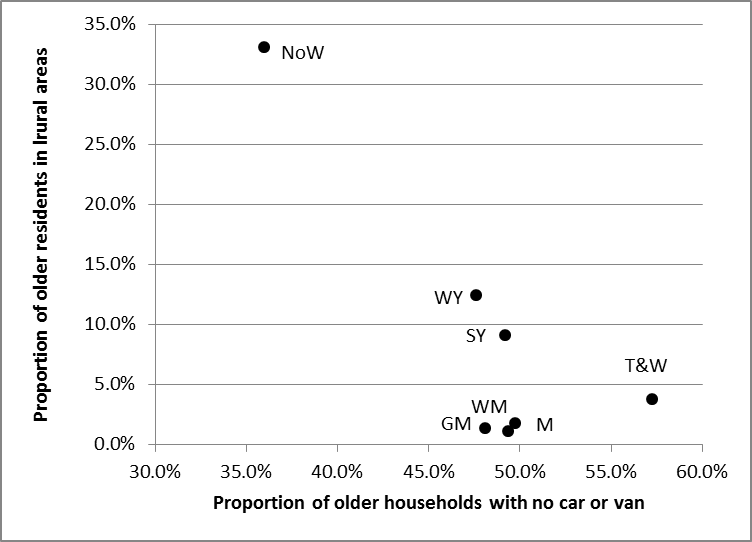
**Table 3 Key indicators from 2011 Census for the NoWcard districts and PTE areas**

Relative to all the PTEs, the 4 Lancashire Districts have:

* a somewhat larger proportion of the population aged 60 or over;
* very substantially lower population densities;
* a very much larger proportion of the older population living in rural (i.e. non-urban) areas;
* higher car ownership amongst the older population.

These indicators have drawn upon the most readily available summaries of the 2011 Census data and do not take account of the precise eligibility criteria for older people, nor the number of disabled persons who potentially qualify for the concession. But more precise data is unlikely to significantly affect these overall conclusions.

Two indicators that particularly demonstrate the differences between the NoWcard districts and the PTEs are the proportion of older residents in rural areas (and hence unlikely to have access to higher frequency bus services), and the proportion of older households without access to a car or van (and hence dependent upon public transport). These are plotted in Figure 2, which is a scatter diagram of these indicators, which show the similarity of the PTE areas (and in particular between the West Midlands and Greater Manchester), and the difference between all of the PTEs and the NoWcard districts.



**Figure 2 Indicators of area types from 2011 Census**

These indicators reflect the socio-economic and geographic characteristics of the different areas but do not directly reflect public transport usage. Unfortunately, good data on, for example, public transport journeys is not readily available for the NoWcard districts, but it is possible to compare the scale of concessionary journey making. Some assumptions are required to achieve like-for-like comparisons, because of the significant concessionary use of rail and tram in PTE areas.

In some PTEs, continuous monitoring surveys provide well-established means of estimating concessionary journey numbers on the different public transport modes. In the West Midlands there is more limited survey data because of the widespread availability of smartcards for monitoring concessionary bus usage. Data on concessionary tram and rail use is poorer, and it has been assumed that the concessionary use of each public transport mode is similar to that for all public transport passengers.

For the four Lancashire Districts, Metro is not an option and no data is available on concessionary rail use, which is assumed to be small. Estimates of older and disabled concessionary bus journeys have been based on the five weeks of smartcard data collected in February-March 2009, which we have expanded to a full year by multiplying by 52/5.

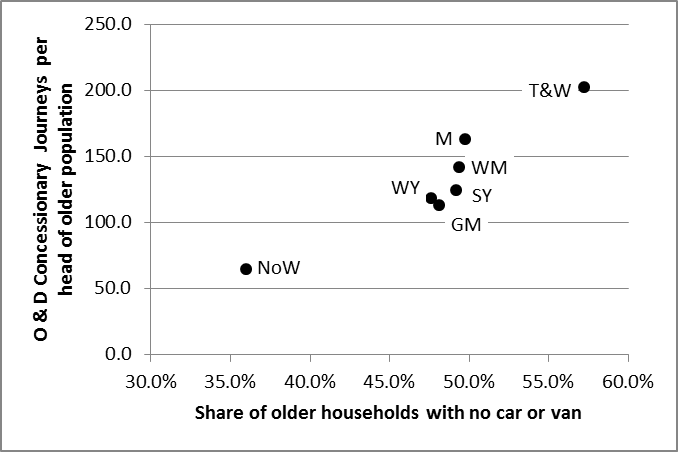
The data is summarised in Table 4 below, which also includes values for the other PTEs (2012-13 data).



**Table 4 Concessionary journey volumes**

When considering which of the look-up tables is most likely to be appropriate to Greater Manchester, the most significant issue is probably the average number of concessionary journeys per head. This is most likely to be well correlated with the frequency of daily and weekly trip making and hence will have the greatest influence on the discount factors estimated for a given set of ticket prices. The measure shown in the table is the average number of older and disabled concessionary journeys per head of population, defined (in this instance) as usual residents aged 60 or over.

The table clearly shows a wide gap between the 4 Lancashire districts and the PTEs, although there is quite wide variation between the PTEs. It is not obvious why the concessionary trip rate in Greater Manchester should be the lowest of the PTEs, although it is much closer to most PTEs than it is to the NoWcard districts. Journey rates are strongly correlated with the level of car ownership, as is illustrated in Figure 3.



**Figure 3 Correlation between journey rate and car availability**

On these measures, five of the six PTE areas, including both Greater Manchester and the West Midlands, clearly form a group of similar characteristics. Tyne and Wear represents the extreme, with particularly high usage of the concession, but accompanied by especially low levels of car ownership amongst the older population. In contrast, the NoWcard districts have both relatively low usage of the concession, which is highly correlated with relatively fewer older households not having access to a car or van.

**CONCLUSIONS**

Overall, our conclusion is that there is much more in common between the PTEs as a whole than there is between the four NoWcard districts. This is supported by both 2011 Census data, and indicators reflecting the relative volume of concessionary journeys made per head of population. Given the availability of robust look-up tables derived from West Midlands data, it is highly likely that were TfGM to use look-up tables derived from West Midlands data, the results would be much closer to those that would be obtained if it had its own local data, than if it used the NoWcard look up tables.

Our recommendation is therefore that TfGM make use of the West Midlands-based look-up tables in calculating reimbursement for bus operators in Greater Manchester. TfGM could use either the table built by Centro, or the alternative that we have constructed. The latter is more consistent with the default NoWcard table currently included in the DfT Calculator. However, because the Centro-derived table leads to somewhat lower discount factors, its selection would recognise that residual uncertainties associated with average fare calculations are inevitable, and gives the benefit of these doubts to the bus operators.

The residual uncertainties include:

* fundamental questions about the extent to which the discount factor method adequately represents the counter-factual;
* the best way of dealing with partial journeys when building look-up tables from expanded data; in our view the Centro look-up table may slightly under-estimate discount factors;
* concerns that the arithmetic within the DfT Calculator itself may tend to under-estimate the “final” discount factor.

There remains a degree of uncertainty about how close a look-up table built from Centro data will fully reflect the equivalent that would be built from data on concessionary journeys in Greater Manchester were it to be available. However, in our view it is far more likely that the look-up table built from West Midlands data will be more representative of Greater Manchester than the look-up table based on NoWcard data. Consequently, we recommend that TfGM use the Centro-derived, West Midlands based look-up table for its reimbursement calculations.

Andrew Last, Minnerva Ltd

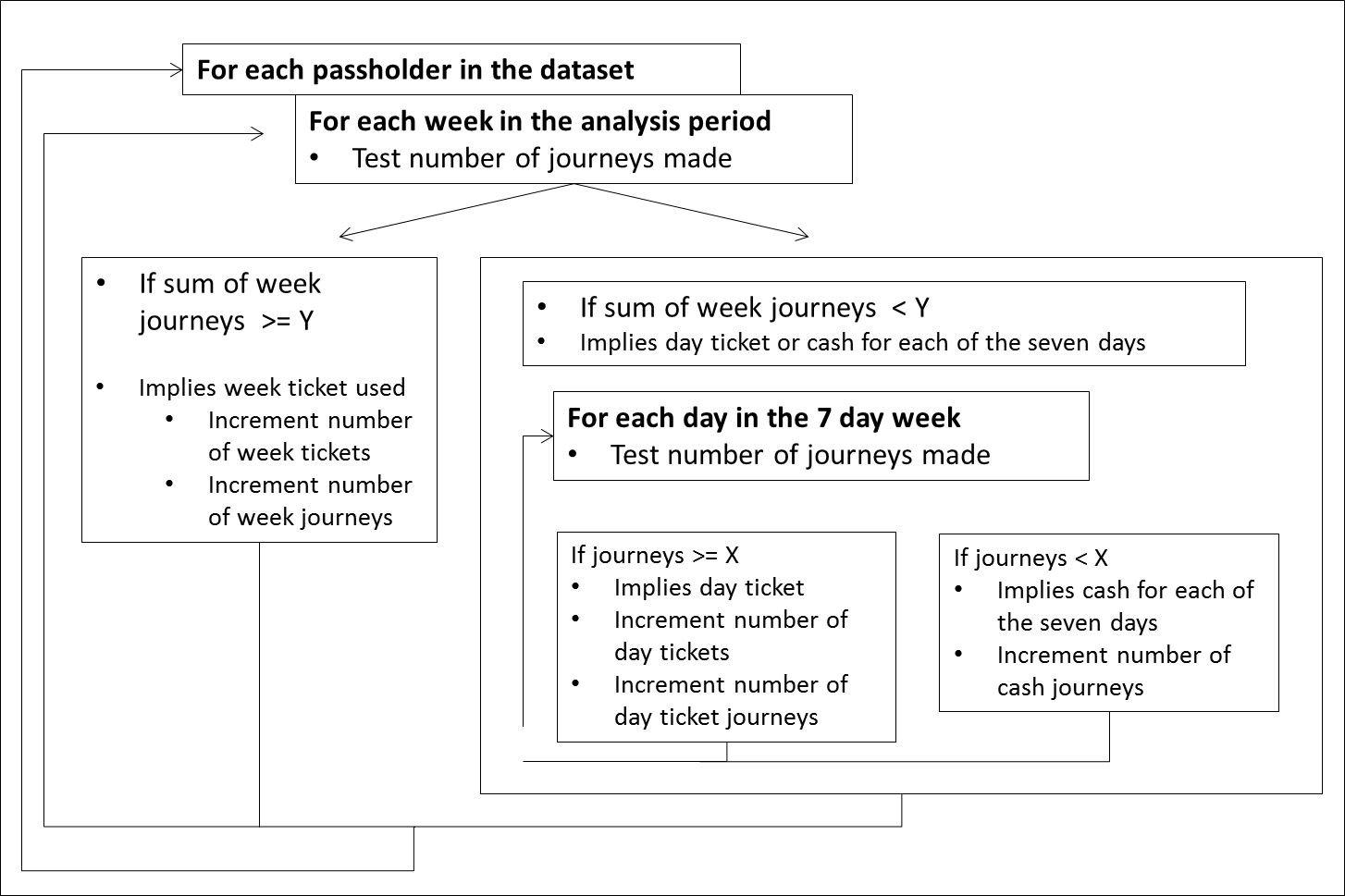
17th February 2014

**APPENDIX 1 LOOK-UP TABLE CONSTRUCTION: MINNERVA/NoWCard process**

In concept, the construction of the look-up table itself from the journeys-by-passholder-by-day dataset involves an iterative process that requires the following steps:

* For each passholder
* For each week ticket price ratio Y (i.e. 1, 2, 3 etc)
* Then for each day ticket price ratio X (i.e. 1, 2, 3 etc)
* For each week in the analysis period
  + If the sum of journeys in the week is equal to or exceeds Y, assume a weekly ticket would have been sold, and record the weekly ticket sale and journeys
  + Otherwise for each day in the particular week
  + If the journeys made on the day exceed X, assume a day ticket would have been sold, so record the day ticket sale and associated journeys
  + Otherwise record the journeys made as cash ticket sales.

The diagram below illustrates this procedure.



There are various ways in which this process can be implemented. Some sort of automated process to control some of iterative loops is likely to be required, as might be constructed using Microsoft Visual Basic, either on a standalone basis or incorporated into an Excel spreadsheet. This is the approach adopted for re-analysis of the Centro data, in which an Excel Visual Basic macro controls successive calculation of each row of the look-up table, iterating over the range of weekly price ratios from 1 to 40.

The macro calculates the number of weekly tickets and journeys, daily tickets and journeys, and cash tickets, for each passholder for a given weekly price ratio. The summation of these, over all passholders, is then written to a summary worksheet as one row of the look up table. The macro controls this process over the range of weekly price ratios. Some subsequent manipulation is then required to put the resulting output data into the format that exactly matches the row and column structure of the look-up table incorporated in the DfT Calculator.

**APPENDIX 2 CENTRO-DERIVED WEST MIDLANDS 2013 LOOK-UP TABLE (FRAGMENT)**

Data: concessionary journeys made on 1st April – 29th September 2013

Built by: Centro

Unique card numbers with at least one journey recorded during period: 324,850



**APPENDIX 3 MINNERVA-DERIVED WEST MIDLANDS 2013 LOOK-UP TABLE (FRAGMENT)**

Data: concessionary journeys made on 1st April – 29th September 2013, separately analysed in two distinct 13 week analyses

Built by: Minnerva

Unique card numbers with at least one journey recorded during period (average of two 13 week periods): 290,393



1. Most recently published in October 2013 for schemes commencing in April 2014 [↑](#footnote-ref-1)
2. This figure is likely to overstate the number of individual passholders making these journeys, since there will be an element of duplication where cards have been reissued to individual passholders for various reasons. [↑](#footnote-ref-2)
3. The four Districts for which passholder and journey data were extracted in the 2009 ITS study were Chorley, Lancaster, South Ribble and Wyre [↑](#footnote-ref-3)