APPENDICES

APPENDIX 1: QUESTIONNAIRE FACSIMILE

The questionnaire form used in the survey is shown on the following pages. In practice, it was printed on a double A4 sheet, and presented to the respondents as a single folded piece of paper.

APPENDIX 2: INVESTIGATION OF DATA ACCURACY

Many studies based on data collected from travellers' own estimates have found cause for concern in the accuracy of the estimates. Particularly when travellers are asked to make estimates about journeys that they have not made, there must always be some suspicion about the processes of perception and reporting. There has been speculation about the effects of these processes* but little hard evidence to clarify their functioning.

Accordingly, we decided in this study to investigate the accuracy of the responses. Respondents had been asked for their home addresses, so it was possible in many cases to deduce exactly what journey the individual had made and what his alternative was. To reduce the problems of getting information on the public transport and highway systems, we took a sample entirely from the data collected in Sheffield. For these people, we traced out the journeys we thought they had made and the alternatives, and compared the answers they had given on the questionnaires with our estimates of the times and costs of the journeys.

Three parameters could be compared directly: the travelling time, the walking time and the cost of the public transport journey. The distributions of ratios shown in Table 2.1 show no significant differences between the public transport employees and the control group, except perhaps in the walking time, which we might have consistently underestimated for one of the public transport sites.

Comparing the actual journeys with the alternatives in this table, we see that actual journeys are overestimated slightly, alternative journeys a little more, but the differences are generally less than 10 per cent in each case.

---

* See, for example, Report of a Workshop on the Value of Travel Time, DOE Time Research Note 1975.

Table 2.1: Distributions of Ratios of Respondents' Estimates to Independent Estimates

	Public Transport Employees			Control Group		
	Mean	(Error)	SD	Mean	(Error)	SD
Actual journeys						
Travel time	1.06	(.03)	.22	1.08	(.04)	.26
Walk time	1.22	(.06)	.48	1.00	(.07)	.47
Cost	–	–	–	1.00	(.01)	.06
Alternative journeys						
Travel time	1.17	(.09)	.38	1.14	(.05)	.24
Walk time	1.22	(.12)	.52	1.11	(.09)	.47
Cost	–	–	–	1.04	(.02)	.11

Another parameter investigated that appeared to give similar results was the driving distance. Because we could not determine the route that drivers would take, the comparison was made between the stated driving distance and the crow-fly distance between home and work place. The introduction of a second parameter, since driving distance cannot be represented as proportional to crow-fly distance, but requires an individual constant to be added, adds another degree of uncertainty which removes any possibility of deriving significant results. Such indications as there are show results consistent with those for the public transport variables discussed previously, that is that actual journeys are overestimated slightly and alternative journeys slightly more.

The final parameter investigated in this way was the waiting time. Although we compared the waiting time with the headway, we could find no positive correlation; indeed both actual and alternative journeys show slight negative correlations between headway and waiting time for both public transport employees and the control group employees. A quite important difference was found between actual and alternative journey waiting times, the mean of the former being about five minutes (one-way), whereas the mean of the latter is about eight minutes.

This finding on waiting time is the only finding of the accuracy investigation that gives rise to any concern. For the other variables, the biases introduced are small and fairly uniform. The problems introduced by the waiting time bias, however, are severe, and reduce drastically the value of the investigations of that variable. These problems are discussed further in Appendix 7.

APPENDIX 3: ANALYSIS OF INCOMPLETE DATA

One of the main problems in analysing data collected from respondents who complete questionnaire forms voluntarily is that some of them will fail to answer some of the questions. In most previous studies the data that was received from such people was omitted. In this study, however, to omit all the incomplete records would have been too serious a loss of data. We therefore devised a method to enable these records to be included.

The method is applicable to maximum likelihood analysis of logit models of the form

$$Pr(y=1) = \phi(a_1x_1 + \dots + a_rx_r)$$

where y is a binary random variable taking the values 0 and 1;

ϕ is the logit function defined by

$$\phi(z) = 1 / (1 + \exp(-z));$$

$x_1 \dots x_r$ are known;

and $a_1 \dots a_r$ are the parameters to be calibrated.

The problem arises when some of the x variables are not given in the data.

The procedure involves four steps.

1. The means $\bar{x}_1 \dots \bar{x}_r$ and covariance matrix Σ of the x variables are calculated.

2. For each individual with incomplete x data estimates are made of the distribution of the missing x values conditional on the known x values.

3. For each individual a 'weighting factor' λ is calculated, taking values between 0 and 1, that reflects the completeness of the data.

4. The estimated data and weighting factors are then used in the logit analysis to estimate the parameters a.

It turns out that the weighting factors λ also depend on the parameters a, so that steps 3 and 4 have to be performed iteratively to achieve a stable solution.

In this Appendix we do not give proofs of the results, but merely outline the methods that are used at each stage in the process and the most critical assumptions that are required.

In the first stage of the process, the means and covariances of the x variables are calculated using the relevant complete data. That is the mean of x_i is calculated only from the records for which x_i is given; the covariance of x_i and x_j is calculated only from the records for which x_i and x_j are both given. This procedure clearly requires the assumption that the presence or otherwise of x_i values is independent of the value itself. We can, however, correct for dependence of the presence of x_i values on y by carrying out separate calculations for records with y=0 and those with y=1, then combining these calculations in the overall ratio of y=0 and y=1 before proceeding to the next stage.

For the second stage of the process, estimating the distribution of the unknown x values conditional on known x values, we require the assumption that the overall distribution of the x variables is multivariate normal. Given this assumption, it is fairly straightforward to calculate the conditional mean $\bar{x}^*$ and conditional covariance matrix Σ^* of the unknown values, but it must be noted that both of these conditional estimates are in general different from the unconditional estimates.

The values of λ calculated in the third stage of the process, are given by

$$1/\lambda^2 = 1 + 3 \text{ var } (a_1x_1 + \dots + a_rx_r) / \pi^2$$
$$= 1 + 3 \underline{a}'\Sigma^*\underline{a} / \pi^2$$

using vector notation, where Σ^* is the conditional covariance matrix of the unknown x values for this individual. It can be seen from this equation that if all the x values are known, $\Sigma^*=0$ and $\lambda=1$. On the other hand if all the values are unknown, Σ^* becomes infinite and $\lambda=0$.

In the iterative process, where we solve for λ and $\underline{a}$, we use at each stage the best available estimates of $\underline{a}$ to estimate λ. For the first iteration, if no previous estimates of $\underline{a}$ are available, we set λ to a constant value for all incomplete records to obtain a first estimate of $\underline{a}$.

The final step of the procedure uses a conventional logit analysis. The weighting by λ is applied to the x variables before the logit analysis, so that the model calibrated is

$$Pr(y=1) = \emptyset(a_1\lambda x_1 + \ldots + a_r\lambda x_r)$$

This weighting process is somewhat unusual, but it can be proved to be very close to the correct result for $\underline{a}$ under fairly general assumptions.

The weakest assumption is this procedure is that the x variables are multivariate normally distributed. Although this assumption clearly cannot be exactly true for the type of data collected from transportation questionnaires, we felt that the value of the incomplete data rehabilitated by this process outweighed the small errors that might be introduced.

APPENDIX 4: SPECIFICATION OF POPULATIONS AND VARIABLES FOR ANALYSIS

A total of 3365 questionnaires were returned from the survey, but, for a number of reasons, not all of them could be used in the analyses. In this Appendix we detail the reasons why records were deleted from the data, and give the numbers excluded for each reason. We then go on to define precisely the variables used in the analyses.

Data Excluded

A maximum number of 2361 individuals were considered for inclusion in the analyses. Other records were excluded for five reasons.

A total of 66 individuals had either not come to work on the day surveyed, or had not (one case only) stated the mode used.

A total of 617 individuals made their journeys to work by modes other than car driving or public transport (bus, train and ferry). The main modes for these people are walking (complete journey), car passenger, cycle and motor cycle. Those that walked constitute 158 records, mainly shorter journeys by non-car owners. Journey data was not coded in detail for these trips.

A total of 93 individuals made trips at times outside normal commuting hours, taken to be arrival at work 0600-0959 and departure from work 1500-1959. Individuals not stating times were left in the data. These people mostly travelled by car or had no car available.

Only 9 individuals were removed because their public transport journeys (or public transport alternatives) contained unusual values of certain components. These were total walking or waiting times in excess of 60 minutes, or a total fare in excess of £5. These people all went by car.

A total of 219 car drivers were excluded from all the analysis because they stated that they needed their car for work. The questions of availability and mode choice would, we thought, be overruled by the requirements of the employer.

Further exclusions were made from specific analyses for specific reasons that made the records unusable for that particular analysis.

The further systematic exclusions from the car ownership analyses were of those individuals who had failed to answer the car ownership question. A total of 126 had failed to answer, but in 49 cases we were able to deduce that the household either did or did not (one case) own a car, so that only 77 records had to be excluded. A further 4 records were excluded as outliers,* leaving 2280 records for this analysis.

For the multi-car ownership, which was conditional on car ownership, we obviously excluded the 648 non-car owners. We also excluded those 48 records where we had been able to deduce positive car ownership but not the exact number of cars. This left 1584 records.

For both car availability and mode choice we included the records where the car ownership question had not been answered. Three groups of individuals were, however, excluded from both analyses.

A total of 207 public transport users (to work) were excluded because their journeys were 'impure'. Usually this meant that the journey <u>from</u> work was not by public transport (e.g. car passenger), although those who took public transport to work and walked home were included in all analyses. Other trips of this type would include mixed journeys - e.g. driving or being driven to public transport stations. Most of these people did not have a car available.

The final systematic exclusion from both car availability and mode choice analyses was of the 129 public transport users who failed to answer the question "could you have driven?". For these people the

---

* Outliers are records whose value of the dependent variable is extremely unlikely. They are excluded since they might otherwise distort the model.

question of car availability was, we felt, not resolved. For car availability, therefore, we could not determine their value of the dependent variable, and for mode choice non-car availables were excluded anyway.

A further 9 individuals were excluded from both analyses as outliers. That is, their car availability or mode choice was highly unusual and the reasons for this situation as revealed by the questionnaire were not those we were intending to model - e.g. their car was being repaired. This left 2016 records for the car availability analysis.

From the mode choice only we also excluded 37 people whose car journeys or car alternatives contained unusual components. These were additional walking of more than 40 minutes in total, a distance driven (one way) of more than 20 miles or parking and toll charges more than £2. We also excluded the records of 812 individuals who did not have a car available, since the mode choice was conditional on car availability. This left 1167 records for mode choice analysis.

These exclusions are summarized in Table 4.1.

Table 4.1: Records Excluded from the Analysis

Total records			3365
No journey made or no mode stated	66		
Mode used other than drive or pt	617		
Journey outside normal times	93		
Unusual pt journey	9		
Car needed for work	219		
Total Considered for analyses			2361
Car Ownership			
Ownership not known or deduced	77		
Outliers	4		
Included for car ownership		2280	
Not car owners	648		
Ownership not known but positive	48		
Included for multi-car ownership		1584	
Multi-car owners	304		
Car Availability and Mode Choice			
Impure pt journey	207		
Car availability not known	129		
Outliers	9		
Included for car availability		2016	
Car not available	812		
Unusual car journey	37		
Included for mode choice		1167	
Pt users	231		

Specification of Variables

Dependent Variables

- Mode choice was taken primarily from the individual's completion of the questionnaire. This defines the modes used as "drove a car or van" or "travelled by public transport" for the journey "all or most of the way to work". Inconsistencies are thus possible if the journey home was radically different from the journey to work.
- Car availability: an individual was taken to have a car available unless: either the number of cars or vans in regular use by the household is zero; or he went to work by public transport and stated either that he could not have driven to work or that he did not drive because someone else required the car. This definition avoids any dependence on a decision by the respondent whether he 'could have driven' in spite of the requirements of another household member.
- Car ownership was taken primarily from that question on the questionnaire, which includes having regular use with car ownership. In some cases where the question was not answered it was possible to deduce that the household owned or did not own at least one car.
- Multiple car ownership was taken from the car ownership question.

Journey Variables

Journey variables were taken directly from the questionnaire. Except for the driving distance, which was specifically one-way (to work), they referred to round journey times, though in some cases we assumed a return journey with the same times and costs as the outward journey. For the public transport journey (made or given as an alternative), five variables were used:

walking time (minutes)
waiting time (minutes)
in-vehicle time (minutes) in bus, train or ferry
fares (pence) or a daily average if a season ticket was used
number of buses and trains (not ferries).

For the car journey in mode choice analyses, three further variables were used:

driving time (minutes)

driving distance (one way miles)

parking cost (pence) plus tolls for the Mersey tunnels where appropriate.

Any walking time in this journey was subtracted from that for the public transport journey.

Location Variables

The sites surveyed were divided into ten groups, and nine dummy variables were used to represent the differences between Liverpool Edge Lane sites and the nine other groups. The groups were:

	(Liverpool Edge Lane)	6.	Stoke
2.	Liverpool City Centre	7.	Sheffield City Centre
3.	Norwich Eastern Counties	8.	Sheffield Queens Road
4.	Norwich Colman Foods	9.	Bristol
5.	Bolton	10.	Manchester

The numbers given for each group of sites correspond with the variable numbers used in the analyses, reported in Appendix 6.

Other Dummy Variables

Dummy variables were also used for five other factors in the analysis.

- Job type, taking the value 1 for manual workers, 0 for clerical and administrative workers.
- Employment, taking the value 1 for 'control group' employees, 0 for public transport employees.
- Sex, taking the value 1 for women, 0 for men.
- Age was available in four ranges: under 21, 21-30, 31-50 and over 50. The age group 31-50 was arbitrarily selected as the standard, and up to three variables were introduced to represent differences between the other age groups and this group.

- Income was available in six ranges, five covering £1000 per annum ranges up to £5000, and the last covering annual incomes over £5000. We found no significant differences between the two lowest income groups (probably because of a lack of data), and merged these for all the analyses. Up to four dummy variables were then used to represent the differences between this group and the other income groups.

APPENDIX 5: JOURNEY LENGTHS OF PUBLIC TRANSPORT EMPLOYEES AND CONTROL GROUP

In Chapter 2 we indicated that the public transport employees and the control group were closely comparable in terms of a number of descriptive parameters. A further parameter of interest is the length of the journey to work. Partly this is of interest to determine whether the journey lengths might have different effects on the response to fare changes by the two groups, partly the investigation aims to detect any effect that the fare concession might have on the journey length itself.

The investigation is summarized in Tables 5.1 and 5.2, which show the comparison of journey lengths defined respectively by driving distance and by public transport travel time. Only male respondents are included. These tables show the numbers of respondents for whom data was available in each of the two groups of employees and the t statistic for the difference between the groups.

The results in these tables show no clear effects. At some sites, particularly in Liverpool, there are significant differences in the journey lengths between the groups, but no trend that covers all the sites.

Table 5.1: Comparison of Driving Distances

	Car Drivers			Public Transport Users		
Site and Job Type	PT	Control	t	PT	Control	t
Liverpool Edge Lane						
Manual	22	18	-3.8	4	2	-0.5
Non-Manual	34	91	-3.0	5	6	-2.6
Liverpool Centre						
Non-Manual	19	22	-0.2	24	20	-0.2
Norwich						
Non-Manual	14	40	-1.8	11	3	-0.7
Manual	29	56	2.1	3	0	-
Bolton						
Non-Manual	9	19	-1.3	7	2	0.6
Stoke						
Non-Manual	17	53	-2.2	5	0	-
Manual	33	46	1.5	6	1	-
Sheffield Centre						
Non-Manual	25	84	0.3	17	32	-1.6
Sheffield Queens Road						
Non-Manual	6	16	-0.2	3	1	-
Manual	47	23	0.7	8	1	-
Bristol						
Non-Manual	17	48	-0.2	5	0	-
Manual	40	36	-1.0	6	0	-
Manchester						
Non-Manual	74	53	-0.6	33	5	-2.9

Table 5.2: Comparison of Public Transport Travel Times

Site and Job Type	Car Drivers PT	Car Drivers Control	Car Drivers t	Public Transport Users PT	Public Transport Users Control	Public Transport Users t
Liverpool Edge Lane						
Manual	21	21	-2.1	12	7	-0.7
Non-Manual	26	84	-2.8	21	55	-0.5
Liverpool Centre						
Non-Manual	22	20	0.9	50	42	1.5
Norwich						
Non-Manual	13	40	0.4	24	4	0.7
Manual	27	42	2.7	14	6	-0.6
Bolton						
Non-Manual	8	17	-1.6	13	11	1.5
Stoke						
Non-Manual	16	51	-1.1	16	9	1.1
Manual	30	38	1.1	23	26	0.7
Sheffield Centre						
Non-Manual	27	74	2.9	38	62	0.1
Sheffield Queens Road						
Non-Manual	7	14	-0.1	6	3	-1.7
Manual	46	21	0.6	26	11	0.8
Bristol						
Non-Manual	16	50	-1.1	25	6	0.3
Manual	28	27	-0.6	44	0	-
Manchester						
Non-Manual	74	52	0.9	62	17	1.4

APPENDIX 6: LOGIT RESULTS

In this appendix we give the full results of the main logit model calibrations carried out in this study. Six main calibrations were carried out:

Car ownership
Multiple car ownership
Car availability (including passenger takers)
Car availability (excluding passenger takers)
Mode choice (including passenger takers)
Mode choice (excluding passenger takers).

The results from these calibrations are given in detail in the following pages.

The populations used in each case have been described in detail in Appendix 4, except that 273 car drivers who took passengers were excluded for the second car availability and mode choice runs. The variables used were also given in detail in that Appendix. In these tables they are referred to by abbreviations given at the end of the tables.

The tables show first the best estimate of the weight attached to the variable and the standard deviation of that weight. Finally, notes are given on the correlations of the errors in these estimates.

Table 6.1: Car Ownership and Multiple Car Ownership: Logit Coefficients and Standard Errors (in brackets)

		Car Ownership		Multiple Car Ownership	
		probability no car owned		probability one car owned	
1.	CON	.62	(.20)	-.22	(.19)
2.	S2	.55	(.24)	.35	(.32)
3.	S3	-.29	(.30)	-.38	(.41)
4.	S4	-1.30	(.38)	.21	(.33)
5.	S5	-.34	(.31)	-.07	(.39)
6.	S6	-.27	(.22)	-.04	(.29)
7.	S7	-.13	(.24)	.44	(.33)
8.	S8	-.35	(.28)	.93	(.43)
9.	S9	.09	(.23)	-.11	(.29)
10.	S10	.37	(.21)	.40	(.29)
11.	AG1	.40	(.19)	-.88	(.24)
12.	AG2	.19	(.15)	-.03	(.17)
13.	AG4	.50	(.14)	.51	(.21)
14.	IG2	-1.16	(.23)	.06	(.55)
15.	IG3	-1.57	(.25)	-.29	(.54)
16.	IG4	-2.19	(.29)	-1.20	(.53)
17.	IG5	-2.60	(.30)	-2.06	(.53)
18.	WK	-.0064	(.0053)	.0004	(.0055)
19.	WT	-.0357	(.0059)	-.0164	(.0054)
20.	PT	-.0120	(.0026)	.0027	(.0027)
21.	COST	-.0242	(.0050)	-.0016	(.0026)
22.	VEH	.137	(.068)	.067	(.082)
23.	FEM	.27	(.14)	-.24	(.18)
24.	MAN	.23	(.15)	-.30	(.20)
25.	CONS	1.26	(.36)	2.54	(.64)

Correlations

In both calibrations all the site variables have moderate correlations with each other, nearly all falling in the range 0.3 to 0.5. None have any substantial correlations with any other váriable. The control variable is strongly negatively correlated with the cost in both experiments (-.77 and -.52). The only other noticeable correlations of journey variables are the public transport time with cost (-.18 and -.35) and with the vehicles (-.45 and -.47). The income variables are strongly correlated with each other (.51 to .68 and .81 to .90). Female and manual variables are moderately correlated with each other (.33 and .35).

Table 6.2: Car Availability: Logit Coefficients and Standard Errors (in brackets)

Car availability, probability car not available

		Excluding passenger takers		Including passenger takers	
1.	CON	.26	(.21)	.13	(.20)
2.	S2	.50	(.28)	.72	(.26)
3.	S3	-.57	(.35)	-.78	(.32)
4.	S4	-1.17	(.36)	-1.11	(.35)
5.	S5	-.54	(.34)	-.49	(.33)
6.	S6	-.43	(.26)	-.41	(.23)
7.	S7	-.17	(.27)	-.07	(.25)
8.	S8	-.78	(.31)	-.68	(.29)
9.	S9	.38	(.27)	.33	(.24)
10.	S10	.02	(.24)	.27	(.22)
11.	AG1	1.28	(.24)	1.26	(.21)
12.	AG4	.29	(.14)	.37	(.13)
13.	IG2/3	-.85	(.27)	-1.14	(.26)
14.	IG4/5	-1.50	(.30)	-1.76	(.28)
15.	WK	-.0075	(.0055)	-.0129	(.0053)
16.	WT	-.0380	(.0065)	-.0552	(.0063)
17.	PT	-.0111	(.0028)	-.0136	(.0026)
18.	COST	-.0163	(.0047)	-.0157	(.0045)
19.	VEH	.114	(.075)	.124	(.070)
20.	FEM	1.50	(.17)	1.50	(.16)
21.	MAN	.37	(.17)	.33	(.16)
22.	CONS	1.70	(.43)	2.03	(.40)

Correlations

In both calibrations, all the site variables have moderate correlations with each other, most falling in the range 0.3 to 0.5. None have any substantial correlation with any other variable. None of the age, income, sex and job type variables show any notable correlations, except that the female and manual variables are moderately correlated with each other (.33 and .34). The control variable is strongly correlated with the cost in both experiments (-.73 and -.74), but the only other substantial correlation is of public transport time with the number of vehicles (-.43 and -.43).

Table 6.3: Mode Choice: Logit Coefficients and Standard Errors (in brackets)

		Mode choice, probability car chosen			
		Excluding passenger takers		Including passenger takers	
1.	CON	.52	(.40)	.62	(.43)
2.	S2	-1.07	(.64)	-.90	(.58)
3.	S3	-1.31	(.69)	-.86	(.60)
4.	S4	1.17	(1.47)	1.96	(1.67)
5.	S5	-.95	(.62)	-1.01	(.59)
6.	S6	-.22	(.58)	-.08	(.55)
7.	S7	-1.32	(.57)	-1.17	(.52)
8.	S8	-.62	(.57)	-.72	(.56)
9.	S9	-2.27	(.66)	-1.54	(.59)
10.	S10	-1.39	(.47)	-1.56	(.46)
11.	MAN	.54	(.31)	.80	(.29)
12.	IG2/5	.89	(.71)	1.49	(.71)
13.	WK	.043	(.013)	.041	(.012)
14.	WT	.144	(.017)	.153	(.015)
15.	PT	.039	(.008)	.044	(.007)
16.	COST	.037	(.011)	.042	(.010)
17.	PARK	.061	(.016)	.063	(.014)
18.	VEH	-.42	(.15)	-.33	(.14)
19.	DR	.017	(.011)	.011	(.010)
20.	DIST	.024	(.005)	.027	(.005)
21.	CONS	-1.76	(.96)	-2.70	(.93)

Correlations

The correlations between the estimates in these calibrations show the same general features as those in the previous tables: the moderate correlations between the site variables, the high correlation between the control variable and the public transport cost, and the absence of other correlations among the dummy variables. Again the important correlations among the journey variables are between the public transport time and the number of vehicles (-.35 and -.36) and there is also a correlation between the public transport time and the driving distance (.43 and .45).

Abbreviations

CON	–	Control group dummy variable	
Sn	–	Dummy variable for site group n	
IGn	–	Dummy variable for income group n:	IG2 – £2000–£3000 IG3 – £3000–£4000 IG4 – £4000–£5000 IG5 – £5000 +
AGn	–	Dummy variable for age group n:	AG1 – Under 21 AG2 – 21–30 AG4 – over 50
WK	–	Walk time	} mins
WT	–	Wait time	
PT	–	Time spent in public transport	
COST	–	Cost of public transport journey	p
VEH	–	Number of public transport vehicles used	
PARK	–	Park and toll cost of car journey	p
DR	–	Driving time	mins
DIST	–	One way driving distance	miles /10
FEM	–	Dummy variable for women	
MAN	–	Dummy variable for manual workers	
CONS	–	Constant	

APPENDIX 7: INTERACTION OF WAITING TIME AND INTERCHANGE VALUATIONS

The results of the logit calibrations given in Table 7 of Chapter 3 showed consistently a very high value apparently attached to waiting time and a negative value attached to interchanges. The latter result particularly requires explanation, since it is intuitively clear that an interchange imposes additional inconvenience on the traveller over and above the additional waiting time, walking and fare he may incur.

In this appendix we suggest that the negative value of interchanges is indeed spurious. We believe it to result from two effects, which together thoroughly confuse the valuation of waiting time and with it the valuation of interchange with which it is obviously closely linked. We conclude that the valuation of waiting time by analysis of transport choices is not likely to give useful results in most of the circumstances in which it has been applied.

The first difficulty is that car drivers state that they would experience very high waiting times if they were to travel by public transport. This was one of the main findings of the investigation of data accuracy reported in Appendix 5. We are not able to say whether these high stated waiting times are the result of failures in perception or simply that, not knowing the timetable, regular car users would in fact experience high waiting times if they were occasionally to use public transort.* Either way, the effect is that the waiting time acquires a weight as a discriminant additional to its weight as an actual disbenefit to the traveller.

The second difficulty is that although the frequency and reliability of public transport services are known to be important measures of their attractiveness,+ they are not measured in the data we collected, nor in most comparable data sets. These properties of the public transport system are however closely correlated with the waiting time, which will accordingly acquire additional weight to accommodate them.

Thus we see that the use of waiting time in the usual way in a choice model will lead to overvaluation of waiting time because of these effects. In this study, however, we are able to detect these effects because of the inclusion of a representation of interchanges. The overvaluation effects apply in full only to the first public transport vehicle used. Subsequently, regular travellers are not able to time their arrivals so accurately, and waiting times themselves reflect more fully the frequency and reliability of the services. But because most public transport journeys involve the use of only one vehicle, the value attached to waiting time will be determined primarily by such journeys i.e. will be too high. The more realistic valuation attached to subsequent waiting periods can then only be represented by a negative weight attached to interchanges.

It seems clear that in this study, as in most others, the value of waiting time has been overestimated. The representation used in this study (with waiting time and interchange) was maintained because it gives a more accurate representation of the total value of the effects discussed than the more usual representation using only the waiting time. For future studies, however, we must recommend that waiting time estimates are obtained only from regular public transport users, that frequencies or headways are incorporated explicitly in models, and that measures of reliability are developed.

* Regular travellers can achieve very low waiting times, e.g. see A J Daly and S Zachary Bus Passenger Waiting Times in Huddersfield LGORU Report T71.

\+ Literature reviewed by Daly and Gale in Reference 2 of the main part of this report.

APPENDIX 8: CALCULATION OF DIVERSIONS

In this Appendix we outline the methods used to predict the diversions likely to result from zero fares. We find that the calculation of the best estimate of diversion is comparatively straightforward, but that estimation of confidence limits is more difficult.

Diversions were predicted for three issues: mode choice, car availability and car ownership. In each case we used the calibrated binary logit model to make the prediction for non-public transport employees by setting the fare component to zero. Because of the proxy effects explained in Section 2.2 it was also necessary to adjust the employment variable (distinguishing public transport employees from other workers) to take the value appropriate to public transport employees.

The diversion was then estimated by

$$D = \sum_i \left(\phi(\underline{a}^{*\prime} \underline{x}_i) - \phi(\hat{\underline{a}}^{\prime} x_i) \right)$$

where the sum is taken over all the individuals i in the control group;

ϕ is, as usual, the logit function;

$\hat{\underline{a}}$ is the vector of maximum likelihood estimates of the parameters;

$\underline{a}^*$ is the maximum likelihood vector adjusted for zero fares as indicated above;

and $\underline{x}_i$ is the vector of independent variables for individual i.

It can be shown that, as the maximum likelihood vector $\hat{\underline{a}}$ varies around the true value $\underline{a}$, the variance of D is given approximately by

$$\text{var } D \doteq \underline{r}' \sum r$$

where $\sum$ is the covariance matrix of the maximum likelihood estimates;

and $\underline{r}$ is given by

$r_j = 0$ if j is not one of the deleted components;

$r_j = - \sum_i \phi' (\hat{\underline{a}}' \underline{x}_i) \; x_{ij}$ if j is one of the deleted components;

where ϕ' is the logit frequency function

$$\phi'(z) = \frac{d}{dz}\phi(z) = \phi(z)\ (1 - \phi(z))$$

These estimates were used for the diversions on each of the three issues separately. To predict the overall mode choice change, combining mode choice with car availability, we used the model

$$\Pr(y_1=1 \text{ and } y_2=1) = \phi(z_1)\ \phi(z_2)$$

giving the probability that a car was available and was used, where z_1 and z_2 are the logit arguments respectively for the car availability and mode choice models. Assuming that the choice probabilities are independent, we may estimate the overall diversion by

$$D_o = \sum_i (\phi(z_{i1}^*)\ \phi\ (z_{i2}^*) - \phi\ (\hat{z}_{i1})\ \phi\ (\hat{z}_{i2}))$$

where z_{ij}^* denotes the amended logit argument
and $\hat{z}_{ij}$ the logit argument with maximum likelihood a values for this individual.

This assumption of independence in the individual's choices, given the values of z_{i1} and z_{i2}, is quite plausible, and gives us little cause for concern.

In practice, however, we could not estimate D_o as above, because the mode choice argument z_2 was systematically unavailable for those travellers who did not have a car. Accordingly, we had to make the further approximation

$$D_o = (\sum \phi(z_{i1}^*) \sum \phi(z_{i2}^*) - \sum \phi(\hat{z}_{i1}) \sum \phi(\hat{z}_{i2}))\ /\ n_2$$

where n_2 is the number of travellers for whom the z_{i2} are defined, and the second sum in each product is taken only over such travellers. This approximation requires the further assumption that z_{i1} and z_{i2} are independent for each individual, which is untrue, but which seems the best that can be done in the circumstances.

This expression for D_o may be rewritten as

$$D_o = n_1 \; (p_1 * p_2 * - \hat{p}_1 \hat{p}_2)$$

where n_1 is the total number of travellers in the car availability model, and p_j is defined by

$$p_j = \sum \phi(z_{ij}) \; / \; n_j$$

The variance of D_o may be estimated by

$$\text{Var } D_o = n_1 (\hat{p}_1^{\;2} \text{ var}(p_2 * - \hat{p}_2) + \hat{p}_2^{\;2} \text{ var}(p_1 * - \hat{p}_1))$$

and $\text{var}(p_j * - \hat{p}_j)$ is calculated as above for the single-stage models.