

Appendix E – Technical note for Workstream C

Technical Note

Project Title:	Child Road Safety and Poverty Research
MVA Project Number:	C3A25400
Subject:	Workstream C Extension of the analysis to include child pedestrians and cyclists together, and child car occupants
Note Number:	Final Version
Author(s):	Henry Burroughs
Reviewer(s):	Stephen Wood
Last updated:	13 June 2011

1 Introduction

1.1 This note summarises the results and findings from Workstream C. Workstream C consists of four tasks:

- Task C1 - An analysis of whether the collision location can be used as a proxy for location of residence, using the last-3-year dataset.
- Task C2 - An analysis of whether or not a relationship also exists between deprivation and the casualty rates for pedestrian / cyclist and car occupant, using a 3 year or 11 year dataset as appropriate
- Task C3 - An analysis of the relevant variables involved in that relationship, using the statistical model.
- Task C4 - analysis of the poorest and best performing Super Output Areas (SOAs) using the model constructed with 11 year and 3 year datasets.

1.2 The note is structured around these tasks with a section dedicated to each.

2 The data

2.1 Analysis was carried using the following two main types of dataset:

- SOA level data
- Aggregated data – an aggregated version of the SOA data

2.2 The SOA level data was used as the input data for developing 'Level 2' models. The Aggregated data was used for casualty rate analysis and 'Level 1' models.

2.3 The SOA level data comprises of data from a number of sources including:

Technical Note Final Version

- Casualty data from the Police Service Northern Ireland (PSNI)
- Census data from the Northern Ireland Statistics Research Agency (NISRA)
- Road infrastructure data from the Department for Regional Development (DRD)
- Meteorological data from Meteorological Office

2.4 The same SOA data used in the previous study commissioned by the Department of Environment NI (DOENI), the Deprivation and Child Pedestrian Road Casualties study (January 2010) was used, but updated with more recent data where practical.

Casualty data

2.5 Two datasets were assembled from the latest casualty data provided by PSNI covering the period January 1999 and March 2010:

- 11 year (11Y) data – the full dataset, covering casualties between January 1999 and March 2010
- 3 year (3Y) data – data with details of SOA of home, covering casualties between April 2007 and March 2010

2.6 The datasets were used to determine the number of pedestrian+cyclist and car passenger casualties that occurred in each SOA by categories of age and gender for both 3-year and 11-year periods. This produced four sets of casualty numbers (by collision SOA, age and gender):

- 11Y pedestrian+cyclist casualties
- 11Y car passenger casualties
- 3Y pedestrian+cyclist casualties
- 3Y car passenger casualties

2.7 Two further sets casualty numbers were produced from the 3-year data where the number of casualties per SOA were determined by allocating casualties to SOA according to the location of the casualties home:

- 3Y pedestrian+cyclist casualties (by home SOA)
- 3Y car passenger casualties (by home SOA)

2.8 These sets of casualty data were then used to update the casualty numbers in the SOA level data used in the January 2010 DOENI study.

SOA level data

2.9 The following table details the contents of the SOA level data along with a short summary of the changes made since this data was used in previous study.

Table 2.1 Summary of the SOA level dataset

Variable name	Description	Change
cNNS	Number of casualties per SOA by age group (NN=0-4, 5-7, 9-11, 12-15, 0-15) sex (S=male, female, male+female) - 15 variables in total.	Taken from new 3Y and 11Y data for both pedestrian+cyclist and car passengers
pNNS	Population per SOA by age group (NN=0-4, 5-7, 9-11, 12-15, 0-15) sex (S=male, female, male+female) - 15 variables in total.	Updated with 2008 mid-year population (MYE) estimates for 3Y casualty numbers and 2004 MYE for 11Y
<u>Area</u>		
Lgd	Local Government District	No change
Area	Square area (ha)	No change
popden2	Population density per hectare	No change
Urbrur	Urban-Rural Definition Group (URDG) Classification	No change
Rain	Annual rainfall (mm)	No change
Sun	Annual sunshine hours	No change
<u>Demographic</u>		
nocar*	Percentage of households with no car	No change
tenren*	Housing tenure: percentage of households rented	No change
tenown*	Housing tenure: percentage of households owned by resident	No change
tenN*	Housing tenure (1=Owned, 2=mortgaged, 3=Shared own, 4=NIHE let,5=HA let, 6=Private let, 7=Other)	No change
Empwrk	Number of workplace population: the number of persons aged 16-74 in employment in the SOA	No change
Empres	Number of employed people resident in SOA	No change
Noqualspc	Percentage of population with no qualification	New - Census table KS13
scplNN	Number of school places (NN=5-7, 9-11, 12-15, 5-15)	No change
fsmPPP	Number of school places (P=Primary [5-7, 9-11], PP=[12-15], PPP= [5-15])	No change
<u>Deprivation</u>		
Mdm	Multiple Deprivation Measure (MDM) score	Updated with 2010 NIMDM Results
inc*	MDM Domain – Income	Updated with 2010 NIMDM Results
emp*	MDM Domain – Employment	Updated with 2010 NIMDM Results
health*	MDM Domain - Health deprivation and disability	Updated with 2010 NIMDM Results
educ*	MDM Domain - Education, skills and training	Updated with 2010 NIMDM Results
Prx	MDM Domain - Proximity to services	Updated with 2010 NIMDM Results
env*	MDM Domain - Living environment	Updated with 2010 NIMDM Results
Crn	MDM Domain - Crime and disorder	Updated with 2010 NIMDM Results
idac*	MDM Domain - Income deprivation affecting children	Updated with 2010 NIMDM Results
<u>Road</u>		
Tproxy	A relative measurement of travel activity in an SOA	No change
Jno	Number of junctions in SOA	No change
Rdtot	Total road length in SOA	No change
NCLASS*	Road length by road class (N=M, A, B, C, U class)	No change
<u>Composite</u>		
Rdenha	Road length per hectare	No change

Jdenkm	Junction density per road km	No change
Jdenha	Junction density per hectare	No change
Tproxrd	travel activity per km of road	No change
Tproxj	travel activity per junction	No change
Popj	SOA population per junction	No change
Empresj	Number of employed people resident in SOA per junction	No change
Empwrkj	Number of workplace population in SOA per junction	No change

* Indicates those variables which in the previous study were excluded from the level 2 models on grounds of potential auto-correlation.

2.10 The process of updating the SOA level data resulted in the following six SOA level datasets, one for each set of casualty numbers:

- **SOA11YPedCycC** – 11-year pedestrian+cyclist casualty numbers by SOA of collision
- **SOA11YCarPasC** – 11-year car passenger casualty numbers by SOA of collision
- **SOA3YPedCycC** – 3-year pedestrian+cyclist casualty numbers by SOA of collision
- **SOA3YCarPasC** – 3-year car passenger casualty numbers by SOA of collision
- **SOA3YPedCycH** – 3-year pedestrian+cyclist casualty numbers by SOA of home
- **SOA3YCarPasH** – 3-year car passenger casualty numbers by SOA of home

Aggregated data

2.11 Aggregated versions of the six SOA level datasets were then created by aggregating the casualty and population numbers in the SOA data by the following factors:

- Age group (x4) [0-4, 5-7, 9-11, 12-15]
- Gender (x2)
- Rural/ Urban (x2)
- MDM decile (x10)

2.12 Each Aggregated dataset has 160 pairs of casualty and population numbers (=4x2x2x10) organised in rows. These were used in the casualty rate analysis carried out in task C2 as reported in section 4 below.

3 Task C1

3.1 This task is concerned with the question, can collision location be used as a proxy for location of residence when classifying the casualties into MDM deciles?

3.2 To answer this, the 3Y casualty (PSNI) data were analysed to identify the proportion of casualties with home postcode information and the proportion of casualties which have the collision in the same SOA as the casualties home. The following table details these measures and the number of casualties by casualty type for the 11Y and 3Y casualty datasets.

Table 3.1 Number of child casualties in 11Y and 3Y datasets and level of matching in home and collision SOA

Casualty type	Number of child casualties 11Y	Number of child casualties 3Y	Percentage of casualties in 3Y with home SOA	Percentage of casualties with home SOA = collision SOA
Car driver	267	24	88%	57%
Car passenger	7,333	1,934	93%	17%
Pedestrian	2,815	721	95%	45%
Pedal cyclist	567	211	95%	57%
Motorcycle/ pillion	90	19	95%	22%
All casualty types	11,072	2,909	94%	27%

3.3 The above analysis shows that:

- the 3Y data does not have home SOA detail for all child casualties since April 2007: Overall there is home SOA detail for 94% of casualties;
- over a quarter (27%) of all child casualties with home SOA detail had a collision in the same SOA as their home SOA;
- ignoring the small number of child car driver casualties, matching home and collision SOA is highest for pedal cyclists (57%) and second highest for pedestrians (45%), whilst only 17-22% of other casualty types had matching home and collision SOA; and
- the proportion of car passengers with a matching home and collision SOA is much lower than that for pedestrians, reflecting the fact that car journeys are typically longer distance journeys.

3.4 Given that nearly half (48%) of child pedestrians and cyclists (combined) casualties have a matching home and collision SOA, we can expect there to be similarities in the relationship between casualty rates and home/ collision MDM deciles.

3.5 The following analyses compares the MDM deciles categorised according to home SOA and collision SOA to investigate the differences in the MDM category that arise from the two locations.

Table 3.2 Level of matching in home and collision MDM for child pedestrian + pedal cyclist casualties

Home											Match	Within +/- 1 deciles	Within +/- 2 deciles
Collision	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10			
M1	165	14	2	5	2	5	3	0	4	2	82%	89%	90%
M2	24	79	3	3	6	4	5	5	1	1	60%	81%	83%
M3	5	9	60	8	2	5	6	5	4	3	56%	72%	79%
M4	5	8	2	49	6	2	5	4	6	1	56%	65%	76%
M5	6	8	5	1	45	1	4	3	2	3	58%	60%	72%
M6	1	0	1	2	5	26	5	3	0	2	58%	80%	91%
M7	1	10	10	2	5	5	29	9	4	2	38%	56%	68%
M8	0	1	1	2	1	2	6	35	4	6	60%	78%	91%
M9	0	4	2	6	2	3	4	3	29	5	50%	64%	71%
M10	0	0	2	3	1	2	0	2	5	26	63%	76%	80%
Total											61%	74%	81%

3.6 Nearly two thirds (61%) of pedestrian + cycle casualties have the same MDM category when coded according to home or collision SOA. This is higher than the 31% measure with matching home and collision SOA codes – a result of the fact that a casualty can have a collision in an SOA away from the home SOA, which has the same MDM decile.

Table 3.3 Level of matching in home and collision MDM for child car passenger casualties

Home											Match	Within +/- 1 deciles	Within +/- 2 deciles
Collision	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10			
M1	101	25	5	9	9	2	8	16	8	9	53%	66%	68%
M2	29	39	12	14	11	9	11	9	9	13	25%	51%	60%
M3	8	18	65	19	22	18	13	11	13	9	33%	52%	67%
M4	16	16	31	51	22	17	14	10	12	13	25%	51%	68%
M5	10	19	11	18	96	22	14	11	11	3	45%	63%	75%
M6	20	19	18	21	16	60	21	13	16	8	28%	46%	62%
M7	6	18	13	11	20	25	60	25	12	11	30%	55%	71%
M8	5	6	12	17	9	21	26	51	14	13	29%	52%	72%
M9	7	7	8	11	7	6	20	12	38	25	27%	53%	67%
M10	3	9	4	5	7	5	6	27	19	32	27%	44%	67%
Total											33%	54%	68%

- 3.7 17% car passenger casualties had the collision in their home SOA and a third (33%) had the collision in an SOA with the same MDM decile as the home.

There is strong correlation between MDM decile based on collision location and home. The correlation is particularly strong for pedestrian+cyclist casualties which often occur close to home. The correlation for car passenger is weaker than pedestrian+cyclist reflecting the longer distance nature of these journeys.

- 3.8 Task C2 in the next section investigates this question further, by looking at and contrasting the relationship between casualty rate and home and collision MDM.

4 Task C2

- 4.1 Rather than taking collision MDM as a proxy for home MDM as the starting point, the association between casualty rates and home and collision MDM has been investigated independently of each other, to see if they represent different things.

- 4.2 Casualty rate analyses by MDM for pedestrian+cyclist and car passenger casualties were carried out on the 3Y data which has both collision and home location details. The following matrix of charts contrast:

- Casualty rates for child pedestrian+cyclist versus by **collision** MDM decile, against
- Casualty rates for child pedestrians+cyclist versus by **home** MDM decile

for males, females and males + females and four age groups.

- 4.3 The legends include the correlation coefficients – 1 very highly correlated, 0 no correlation and a significance codes which indicate statistical reliability of the estimates of the correlation coefficient is as follows:

- *** significant at the 99.9% level
- ** significant at the 99% level
- * significant at the 95% level
- . significant at the 90% level
- "." not significant

4.4 The casualty rate results against MDM of collision site (the red bars) show the following:

- there is a strong and statistically significant association between casualty rates for pedestrians + cyclists and the MDM of collision site in the youngest three age categories;
- this association for 12-15 year olds is very weak if any;
- the younger the casualty the stronger the association; and
- for children aged 7 or less, the association for males and females are very similar; for 8 and over the association is weaker for males.

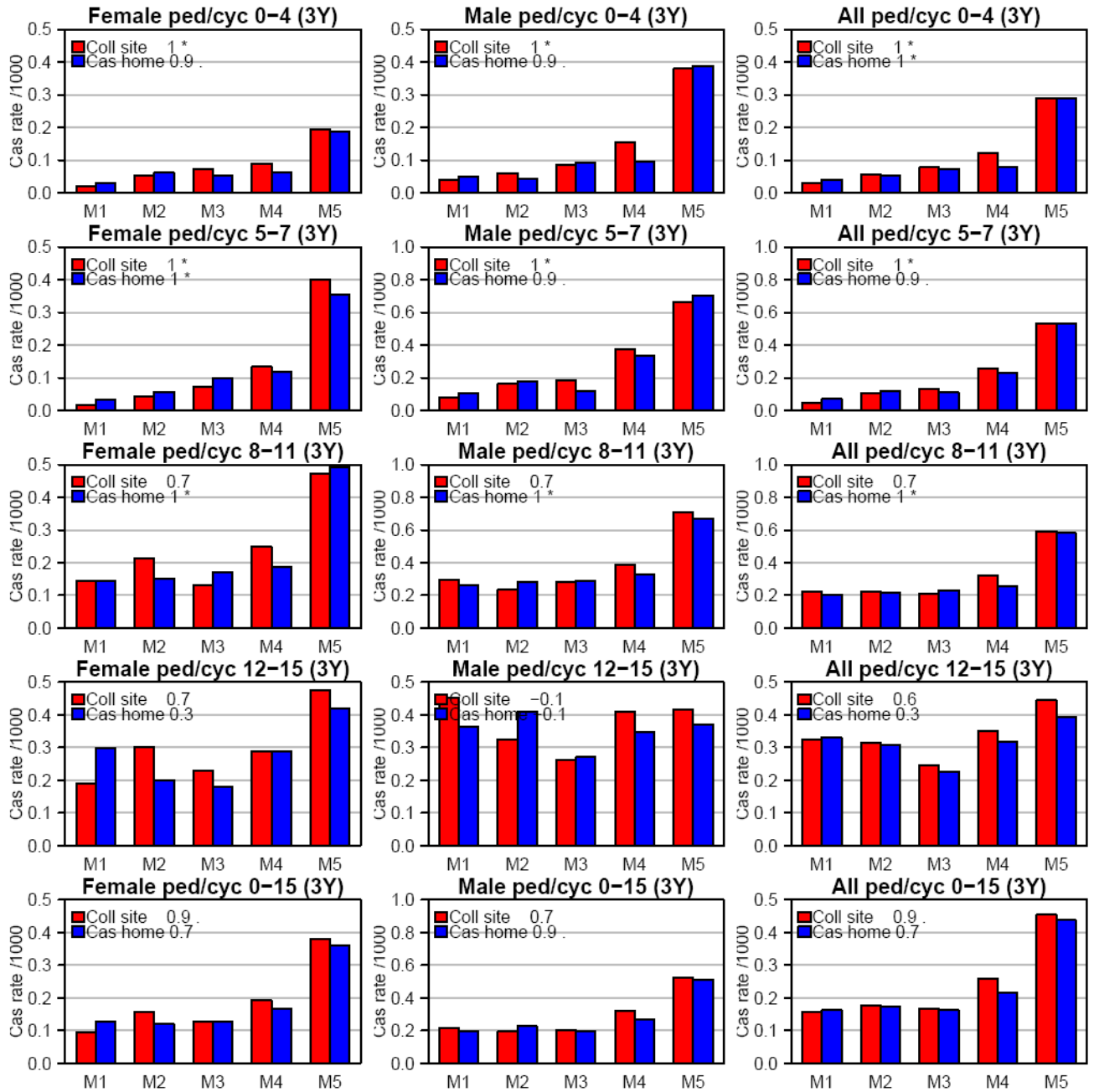
4.5 The casualty rate results against MDM of home (the blue bars) show the following:

- the casualty rate results against MDM of home generally share the same features listed above for MDM of collision; and
- the strength of association and the level of significance is generally less for casualty rates by MDM of home than the corresponding results by MDM of collision site, although there a small number of counterexamples.

Both collision and home MDM show strong correlations with child pedestrian+cyclist casualty rates. Therefore both physical and societal variables will be explored later in the development of statistical modelling.

4.6 In order to allow modelling against home-based societal variables the 3 year dataset will be used. This is also consistent with the use of the latest data which builds on the success of road safety improvements implemented in the period 1999 to 2007.

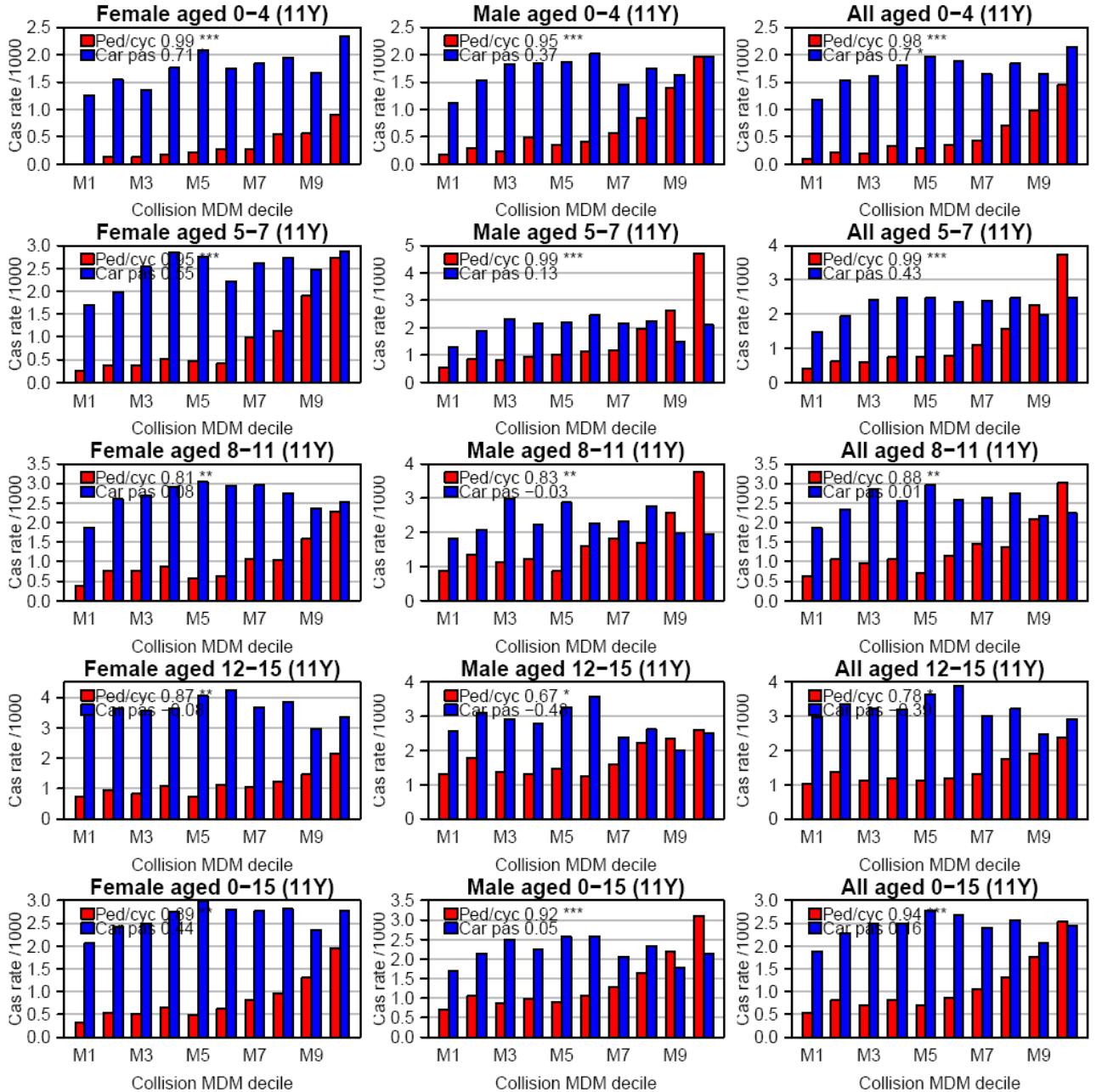
4.7 To address concerns that some of the casualty rates in the preceding analysis were being estimated from rather small sample sizes, the same analysis was repeated for MDM grouped into quintiles as below:



4.8 These results support the findings of the analyses by MDM deciles.

4.9 Next, the relationship between car passenger casualty rates and MDM deciles was investigated. First the 11Y dataset was used to contrast the relationships:

- casualty rates for child **pedestrian+cyclist** versus by collision MDM decile, against;
- casualty rates for child **car passengers** by collision MDM decile.

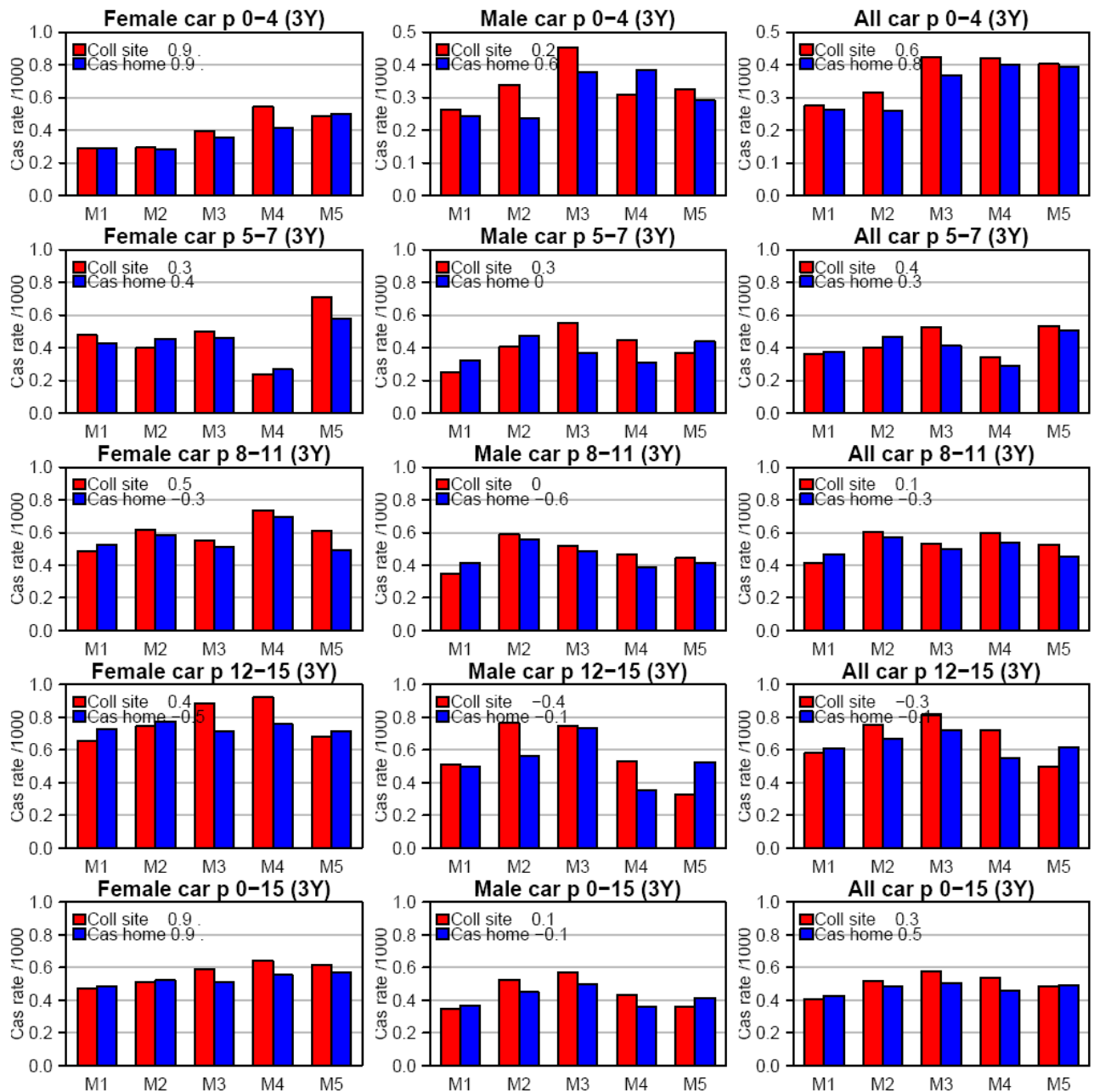


4.10 The casualty rates for pedestrians and cyclists by collision MDM (red bars) show the same characteristics as seen in the corresponding results on the 3Y data outlined in paragraph 4.4. However, there does not appear to be any relationship between casualty rates for car passenger (blue bars).

4.11 The 3Y dataset was then used to look at the relationship between casualty rates for car passengers by home and collision MDM quintile:

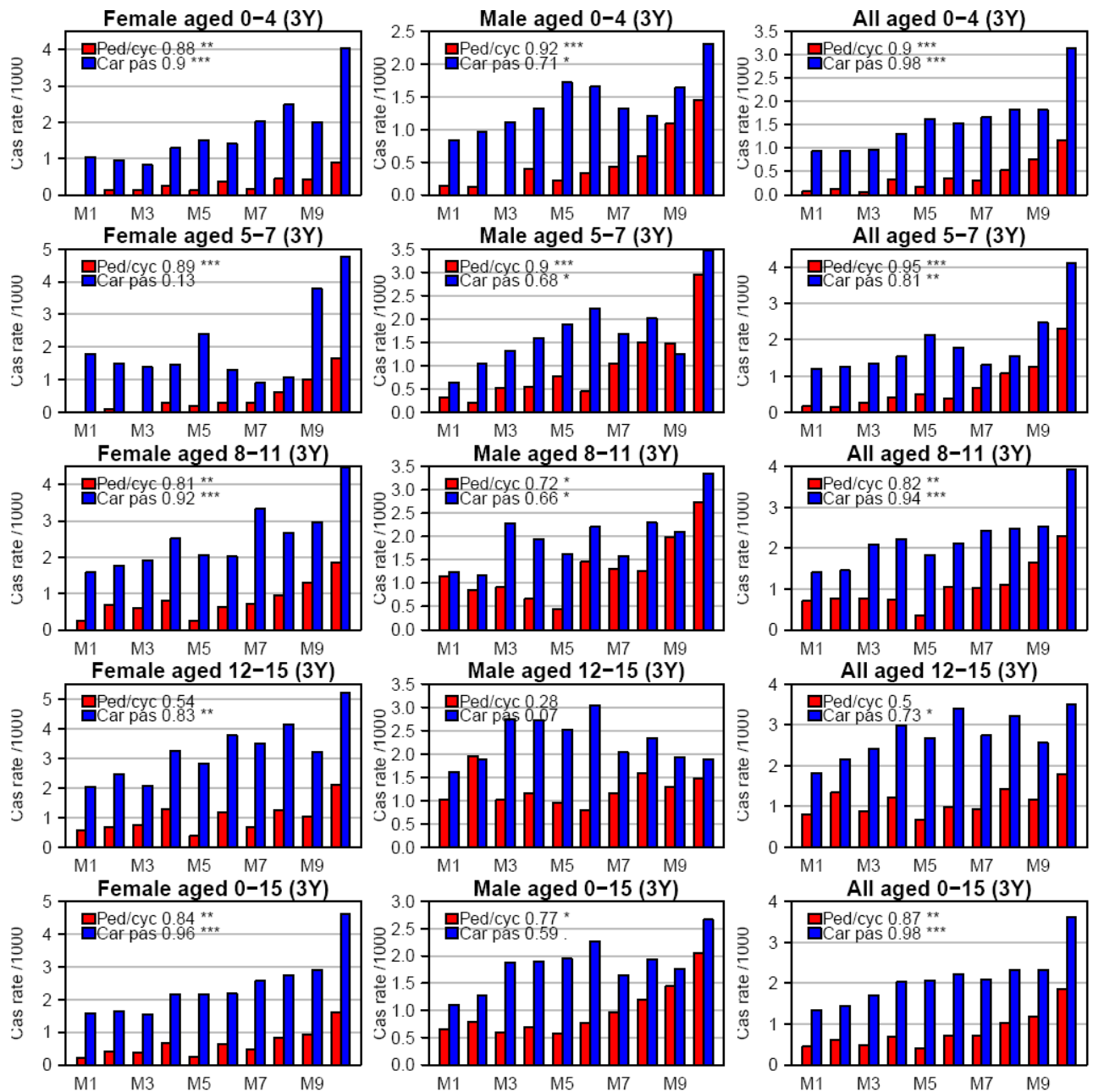
- casualty rates for child car passengers versus by **collision** MDM quintile, against;
- casualty rates for child car passengers versus by **home** MDM quintile.

4.12 The following chart matrix summarises the results.



4.13 As well as there being no relationship between car passenger casualty rates and collision MDM quintile, there is also no relationship between car passenger casualty rates and home MDM quintile. (Results with MDM categorised in deciles show the same.)

4.14 Finally, the relationship between car passenger casualty rates and collision MDM was investigated further by calculating 'adjusted' car passenger casualty rates - the number of car passenger casualties as a proportion of the number of children with access to a car owning households. The number of children within households with access to a car was calculated as: the number of children * the proportion of children (aged 0 – 15) living in households with access to a car/van¹. The following chart presents 'adjusted' car passenger casualty rates results contrasted with casualty rates for pedestrians+cyclists.



4.15 The results for the adjusted casualty rates for car passengers (blue bars) (as a proportion of children resident in car owning households) show:

¹ The proportion was calculated using 2001 Census data at Super Output Area and supplied by DOE officers.

- there is a strong and statistically significant association between adjusted female car passenger casualty rates and the MDM of collision site in all but one of the age groups;
- there is a rather weak association between adjusted male car passenger casualty rates and the MDM of collision site, although it is statistically significant in all but one of age the groups;
- the relationship between adjusted car passenger casualty rates and MDM is similar to that for pedestrian + cyclists: the relationship for the adjusted car passenger casualty rates appear to be more variable, particularly for males despite there being more than double the number of car passenger casualties in the data;
- casualty rates for car passengers are considerably higher than those for pedestrians + cyclists; and
- adjusted casualty rates for car passengers are higher than unadjusted casualty rates because the population of children within households with access to a car is substantially less than the total number of children. This is particularly the case for the most deprived households where car ownership is at its lowest.

There is a strong and statistically significant relationships between MDM and child casualty rates for pedestrian + cycle. There is a weaker and more variable relationship between child car passenger casualties and MDM once adjusted to reflect household car availability but no relationship without this adjustment.

Conclusions and discussion

- 4.16 The analysis shows that there is a strong and significant relationship between casualty rates of child pedestrian+cyclist and MDM decile at collision site. The relationship between casualty rates of child pedestrian+cyclist and home MDM decile is similar but slightly weaker. The relationships for collision and home MDM are expected to be similar given that the Home MDM = collision MDM for 61% of child pedestrian+cyclist casualties. What is less clear is why the relationship would weaken slightly when based on home MDM.
- 4.17 This perhaps could be taken as evidence to suggest that the physical environment, represented by collision MDM, is a slightly stronger factor in causing collisions than societal factors, represented by home MDM.
- 4.18 There is no relationship between casualty rates of child car passengers and MDM decile of collision site or home. However, there is a relationship between adjusted car passenger casualty rates which are based on the estimate of the number of children within households with access to cars. The relationship is weaker and more variable than that for pedestrians+cyclists. Rather than there being a gradual increase in adjusted car passenger casualty rates with increasing MDM of collision site, for many age groups there are instances where casualty rate decreases gradually over three increments in MDM decile.
- 4.19 Except for males aged 12-15, adjusted car passenger casualty rate is highest in the most deprived MDM decile and often by a significant margin. If the most deprived MDM decile were

excluded from the analysis of car passengers it is clear that there would be very little evidence of a strong relationship in any age sex category. The same is true, but to a lesser extent, for the relationship between pedestrians+cyclists casualty rates and MDM decile of collision site for casualties in the 12-15 age group.

- 4.20 We recognise that authorities have key targets for numbers of Serious and Fatal and that there are likely to be differences in the findings if analysis differentiated between Severe vs Slight casualty types. Sample sizes are likely to be the main limitation of an exercise to develop models for serious/ fatal casualties and would need careful consideration if this area of research was pursued in the future.

5 Task C3

- 5.1 The purpose of this task is to develop a statistical model to identify factors associated with child pedestrian+cyclist and child car passenger casualties. The models were developed in a four stage process as follows:

- **Stage 1** - Identify inter-correlated variables and eliminate an initial set of variables from subsequent stages of the model development process. Not all inter-correlated variables are necessarily identified at this stage: further variables may be identified as inter-correlated in the subsequent stages.
- **Stage 2** - Fit high-level ('level 1') models to identify relationships between casualty numbers and key variables of deprivation, age, sex and rural/ urban classification.
- **Stage 3** - Fit detailed ('level 2') models which measure the influence of individual variables to identify which variables are statistically significant and should be retained and those which should be eliminated from the final model
- **Stage 4** - Fit a final ('level 3') model which based only on those variables which are statistically significant.

- 5.2 The remainder of this section reports on each of the four stage of model development in turn.

Stage 1 – Identify and remove inter-correlated variables

- 5.3 With the exception of updated casualty and population numbers, the SOA level data used in the model development is essentially the same data used in the previous study in January 2010. In the previous study the following variables were eliminated during stage 1 of the model development due to being inter-correlated with MDM:

- noacar - Percentage of households with no car
- tenren - Housing tenure: percentage of households rented
- tenown - Housing tenure: percentage of households owned by resident
- tenN - Housing tenure (1=Owned, 2=mortgaged, 3=Shared own, 4=NIHE let,5=HA let, 6=Private let, 7=Other)
- inc - MDM Domain - Income
- emp - MDM Domain - Employment

- health - MDM Domain - Health deprivation and disability
- educ - MDM Domain - Education, skills and training
- env - MDM Domain - Living environment.

- 5.4 In addition, all road length by road class variables (ACCLASS, BCLASS, CCLASS & NCLASS) were excluded because of inter-correlation with total road length.
- 5.5 Rather than simply eliminate these variables again, a decision was made to not exclude any variables in the SOA level data (see Table 2.1) from the subsequent stages of model development, irrespective of whether they were inter-correlated. Instead, inter-correlation would be identified during the model development stages by examining model coefficients: If inter-correlation between variables is adversely affecting the model the model coefficients for those variables will usually be excessively large.

Stage 2 – Level 1 models

- 5.6 The casualty data were modelled using the negative binomial regression for of a generalised linear model (GLM) which is appropriate for modelling count data when the count data are over-dispersed. This is appropriate for modelling the NI casualty data where the distribution of the casualty totals by SOA show variances in excess of the mean (i.e. over-dispersed).
- 5.7 Negative binomial regression models can be used to model both counts and rates. Examples of each would be:
- the year by year number of collisions at a particular junction; and
 - the number of casualties in collisions in a given time period occurring in SOAs.
- 5.8 The number of collisions at a junction is purely a count outcome and can be successfully modelled using standard regression techniques. However to model casualty numbers which vary according to the size of the population in the groups or areas the data has been being aggregated, it is necessary for the model to also take account of population.
- 5.9 This is achieved by specifying the population as an offset variable in the negative binomial model. The model structure used in the model development was a negative binomial model with $\log(\text{population})$ defined as an offset variable. As was the case with the analysis carried out in Workstream B and tasks C1 and C2 of Workstream C, the R package was used to conduct the model development.
- 5.10 A simple level 1 model of casualty numbers using a MDM decile as the single explanatory variable was fitted to all six aggregate datasets to test whether the basic relationship of increasing casualty numbers with increasing MDM existed and how significant any relationships were between pedestrian+cyclist and car passenger child casualty numbers and collision and home MDM. The following list gives a description of the six aggregate datasets tested:
- **A11YPedCycC** – 11-year pedestrian+cyclist casualty numbers, MDM decile of collision SOA
 - **A3YPedCycC** – 3-year pedestrian+cyclist casualty numbers, MDM decile of collision SOA
 - **A3YPedCycH** – 3-year pedestrian+cyclist casualty numbers, MDM decile of home SOA

- **A11YCarPasC** – 11-year car passenger casualty numbers, MDM decile of collision SOA
- **A3YCarPasC** – 3-year car passenger casualty numbers, MDM decile of collision SOA
- **A3YCarPasH** – 3-year car passenger casualty numbers, MDM decile of home SOA

5.11 Although it was concluded in task C2 that model development should focus on the latest 3 years of data, level 1 models were fitted to the 11Y data to provide a point of reference since 11Y results will be very similar to the 10Y results reported in Workstream B and the previous study. This provided a simple gauge for how well the model performed on 3-year data which has not been modelled this way before.

5.12 The following table presents the model coefficients and significance codes (as defined in para 4.3) for models with casualty numbers as the dependent variable and MDM decile as the single explanatory variable.

Table 5.1 Model coefficients and significance codes for casualty numbers modelled on MDM decile for six Aggregated datasets

Casualty: MDM location: Dataset:	Ped+Cyc Collision 11Y	Ped+Cyc Collision 3Y	Ped+Cyc Home 3Y	CarPass Collision 11Y	CarPass Collision 3Y	CarPass Home 3Y
<u>Model term</u>						
(Intercept)	1.2 ***	-0.3 *	-0.1	2.4 ***	0.8 ***	0.9 ***
mdmdecM2	0.3 **	0.3 .	0.0	0.1 .	0.1	0.0
mdmdecM3	0.3 *	0.2	0.2	0.3 ***	0.3 ***	0.2 .
mdmdecM4	0.4 ***	0.5 **	0.2	0.3 ***	0.4 ***	0.2 **
mdmdecM5	0.3 *	0.1	0.0	0.5 ***	0.5 ***	0.2 *
mdmdecM6	0.5 ***	0.5 ***	0.3 .	0.4 ***	0.5 ***	0.3 ***
mdmdecM7	0.8 ***	0.6 ***	0.4 **	0.3 ***	0.4 ***	0.1
mdmdecM8	0.9 ***	0.9 ***	0.5 ***	0.3 ***	0.4 ***	0.1
mdmdecM9	1.3 ***	1.2 ***	1.0 ***	0.3 ***	0.4 ***	0.2 **
mdmdecM10	2.3 ***	2.2 ***	2.0 ***	1.1 ***	1.1 ***	1.0 ***

5.13 The model coefficients, indirectly representing casualty numbers associated with MDM deciles M1-M10, display many of the features seen in the casualty rate analysis in task C2. Generally the model coefficients increase with increasing MDM decile or at least peak in the most deprived decile. The characteristics of the relationship between 3-year pedestrian+cycle and car passenger casualty numbers and collision and home MDM decile mirror those seen in the casualty rate analysis earlier.

5.14 The level of significance across the two 3-year pedestrian+cyclist datasets is generally weak in the five least deprived deciles, especially where casualties numbers have been aggregated by home MDM. This complements the earlier finding that the relationship between casualty numbers and home MDM decile is weaker than that with collision MDM.

5.15 The models for 3-year car passenger casualties by collision MDM and home MDM contrast markedly. The model of collision MDM has highly significant model coefficients across all deciles with the exception of decile 2 only, even though the model coefficients do not gradually increase as MDM increases. The level of significance of the coefficients in the model of home MDM is

much weaker, particularly for deciles 2, 3, 7 and 8. The model coefficients for both of these models also reinforce the casualty rate analysis findings that there is little evidence of a relationship of increasing number of car passengers casualties with increasing level of deprivation of home.

- 5.16 In the context of model development, the results allow a number of conclusions to be drawn.
- There is sufficient sample to develop relatively robust level 1 models using 3-year data – suggests that the development of the more detailed level 2 models could potentially use 3 year data allowing the analysis to explore factors associated with road safety that reflect the outcomes of the road safety initiatives carried out prior to 2005.
 - The identification of a relationship between car passenger casualties rates and MDM, once an allowance is made for car ownership is consistent with the general conclusions regarding pedestrian casualty rates. However for child passengers casualties there was not a clear trend across the deciles but rather a strong relationship with decile 1 the most deprived SOAs. Without a clear relationship between level of casualties and MDM, models are unlikely to identify any evidence of relationships between deprivation and child casualties, fundamental to this study. However, the subject is worthy of further statistical study and car passenger safety should be included in any education plan.
 - Similarly, models developed for casualty numbers by MDM of home will in general be weaker than those based on MDM of collision location.
 - The model of the 3-year pedestrian+cyclist casualties by MDM of collision compares most favourably with the equivalent 11-year model. This data appears to offer the most potential as a platform for re-examining factors associated with road safety whilst reflecting the outcomes of the road safety initiatives carried out prior to 2005.
- 5.17 Therefore, subsequent model development focuses on models of 3-year data of pedestrian+cyclist casualties using MDM of collision site.
- 5.18 A series of four level 1 models were fitted, adding each of the explanatory variables (MDM, age, sex and rural/ urban) in the 3-year pedestrian+cyclist aggregated data in turn. A further two models were fitted to test for interaction between MDM & rural/urban and age & sex.
- 5.19 The following list details the six model formulas. The dependent variable (to the left of the '~' symbol) is the number of 'casualties'; the explanatory variables (to the right of the '~' symbol) are explanatory variables 'mdmdec' (MDM decile), 'age', 'sex' and 'urdg' (Urban/ Rural indicator).
- Model 1: casualties ~ mdmdec
 - Model 2: casualties ~ mdmdec + age
 - Model 3: casualties ~ mdmdec + age + sex
 - Model 4: casualties ~ mdmdec + age + sex + urdg
 - Model 5: casualties ~ mdmdec + age + sex + urdg + mdmdec*urdg
 - Model 6: casualties ~ mdmdec + age + sex + urdg + mdmdec*urdg + age*sex
- 5.20 The following Table 6.2 presents the results for each model.

Table 5.2 Model coefficients and significance codes for models of 3-year pedestrian+cyclist casualty numbers

Model term	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	-0.3 *	-1.1 ***	-1.4 ***	-2.2 ***	-3.2 ***	-3.2 ***
mdmdecM2	0.3 .	0.3 .	0.3 .	0.3 *	1.6 **	1.6 **
mdmdecM3	0.2	0.2	0.2	0.3 .	1.4 **	1.4 **
mdmdecM4	0.5 **	0.5 **	0.5 **	0.6 ***	1.9 ***	1.9 ***
mdmdecM5	0.1	0.1	0.1	0.2	1.8 ***	1.8 ***
mdmdecM6	0.5 ***	0.5 ***	0.5 ***	0.6 ***	2.1 ***	2.1 ***
mdmdecM7	0.6 ***	0.6 ***	0.6 ***	0.7 ***	1.4 **	1.4 **
mdmdecM8	0.9 ***	0.9 ***	0.9 ***	0.9 ***	1.9 ***	1.9 ***
mdmdecM9	1.2 ***	1.2 ***	1.2 ***	1.2 ***	0.4	0.4
mdmdecM10	2.2 ***	2.2 ***	2.2 ***	1.9 ***	1.7 ***	1.7 ***
ageA2		0.7 ***	0.7 ***	0.7 ***	0.7 ***	0.5 **
ageA3		1.1 ***	1.1 ***	1.1 ***	1.1 ***	1.1 ***
ageA4		1.2 ***	1.2 ***	1.2 ***	1.2 ***	1.4 ***
sexM			0.5 ***	0.5 ***	0.5 ***	0.6 ***
urdgU				1.1 ***	2.3 ***	2.3 ***
mdmdecM2:urdgU					-1.4 **	-1.4 **
mdmdecM3:urdgU					-1.3 *	-1.3 *
mdmdecM4:urdgU					-1.6 **	-1.6 **
mdmdecM5:urdgU					-2.1 ***	-2.1 ***
mdmdecM6:urdgU					-1.9 ***	-1.9 ***
mdmdecM7:urdgU					-0.8	-0.8
mdmdecM8:urdgU					-1.1 *	-1.1 *
mdmdecM9:urdgU					0.8	0.8
mdmdecM10:urdgU					-	-
ageA2:sexM						0.3
ageA3:sexM						-0.1
ageA4:sexM						-0.3

5.21 The results illustrate a number of relationships many of which have been already been seen in the casualty rate analysis.

5.22 **Model 1** – The model coefficients for Model 1 are the same as those for reported in Table 6.1 under column ‘Ped+cyc, Collision 3Y’. The model term ‘Intercept’ is the base-line category and represents the MDM decile M1.

5.23 The key findings are:

- increasing deprivation predicting higher casualty rates.

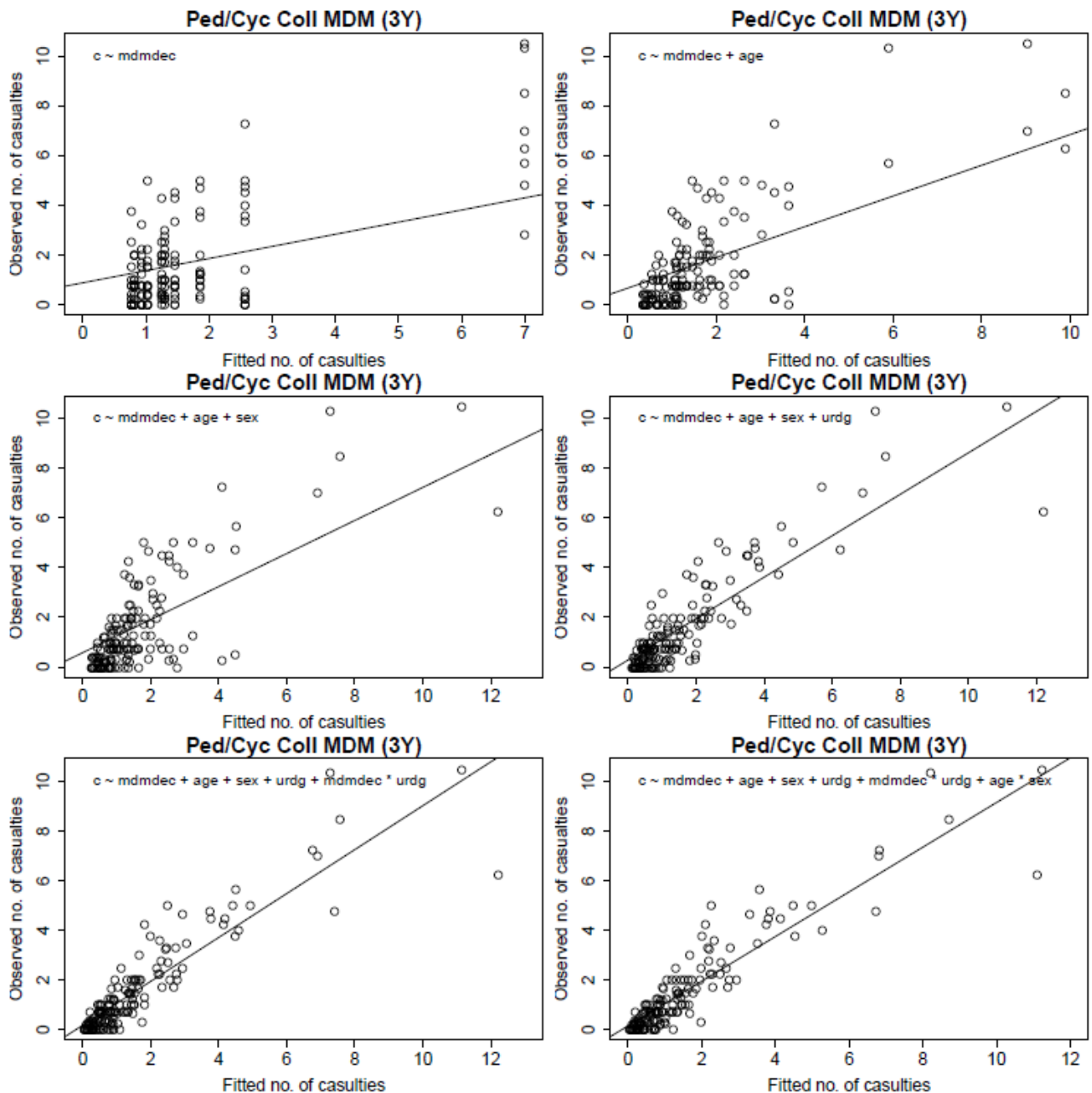
5.24 **Model 2** – The results for Model 2 has coefficients for MDM deciles M2-M10 and age groups A2-A4 [5-7, 9-11, 12-15] and the Intercept which represents the modelled casualties age group A1 [0-4] in MDM decile M1. The log of the modelled number of casualties for a specific combination of age and MDM decile can be determined as Intercept + mdmdecMn + ageAn. For example the modelled number of casualties aged 5-7 in MDM decile 8 = -1.1 + 0.9 + 0.7 = 0.5. The coefficients ageA2, ageA3 and ageA4, 0.7, 1.1 and 1.2 show that, irrespective of the level of

deprivation, the number of pedestrian+cyclist casualties increase the older the age group. The significance codes for age A2, ageA3 and ageA4 show that the association between casualty numbers and age is a highly statistically significant.

- 5.25 It is also worth noting that the fact that the mdmdec coefficients M2-M9 in model 2 are the same as those in model 1 show that the effects of MDM and age on casualty numbers are independent of each other (the effects are 'additive'); i.e. no there is no interaction (or 'multiplicative' effects) between MDM and age.
- 5.26 The key findings are:
- casualty rates varying by age group.
- 5.27 **Model 3** – The results for Model 3 show that the number of male casualties is higher than the number of female casualties (as also seen in the casualty rate analysis in task C2) and that the effect of gender is highly statistically significant. The model coefficients for mdmdec and age have not changed as a result of including sex as a term in the model meaning that the effect of sex is operating independently of mdmdec and age.
- 5.28 The key finding is:
- males giving rise to higher casualty rates than females.
- 5.29 **Model 4** – The results for Model 4 show casualty numbers in urban areas ('urdgU') are significantly higher than those in rural areas and that there is a small amount of interaction between urdg and mdmdec which is likely to reflect the fact that there are more highly deprived areas in urban locations than rural locations.
- 5.30 The key finding is:
- the urban classification giving rise to higher casualty rates than the rural classification.
- 5.31 **Model 5** – The results for Model 5 provide a test of the interaction between mdmdec and urdg hinted in the results of Model 4. The significance codes show that there is significant interaction between rural/ urban and deprivation level. The interpretation of the model coefficients is not obvious but generally it is saying that taking the relationship of generally increasing numbers of casualties as deprivation level increases as the base-line, the rate of increase in urban areas is less than that in rural areas particularly in the mid-range MDM deciles 5 to 7.
- 5.32 The key finding is:
- the effect of the URDG classification is uneven with rural casualty rates for MDM deciles 9 and 10 being less than for deciles 2 to 8. The high estimate value for the MDM decile 10/urban interaction suggests that casualty rates for this category are significantly higher than for other MDM decile values.
- 5.33 **Model 6** – The results for Model 6 show that there is no interaction between age and sex in the way they affect casualty numbers.
- 5.34 The key findings are:
- The significance codes for the interaction terms show that there are no significant differences between the estimates calculated.

5.35 The following matrix of charts illustrate the extent to which the six models are representing the casualty data by plotting the observed number of casualties against the number predicted by the model. Models 1 and 2 are presented in the pair of charts in the top row matrix, models 3 and 4 in the next row and so on.

5.36 The charts show that model's ability to predict casualty numbers improves with the addition of each new explanatory variable in the first four models(excluding the interaction terms). The inclusion of the $mdmdec*urdg$ term improves the model fit yet further whilst the inclusion of the $age*sex$ term does not appear to add any real improvement.



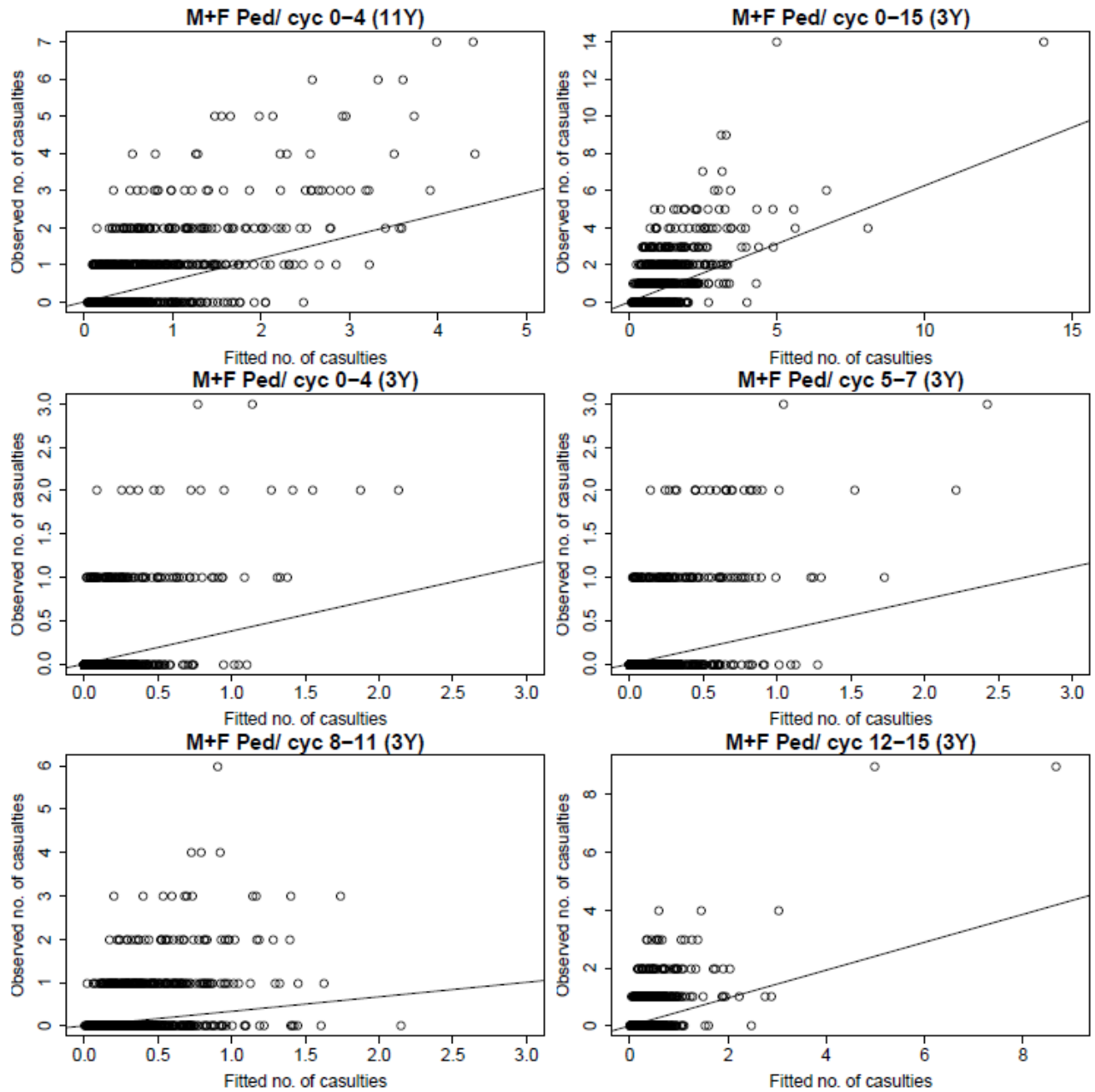
5.37 This Level 1 model analysis demonstrates that there are clear and highly significant relationships between casualty rate and the age and sex of casualty and the rural/ urban classification and level of deprivation of the collision location. This is a key finding in the context

of this study since this confirms that the factors associated with child casualties observed between January 1999 and December 2009 are still important factors in casualties observed much more recently in April 2007 and March 2010.

Stage 3 – Level 2 models

- 5.38 In the high level (Level 1) models it was established that there are clear and highly significant relationships between casualty rate and the age and sex of casualty and the level of deprivation and URDG classification of the collision location. In this more detailed stage of model development level 2 and initial set of models were fitted to the 3-year SOA level data for each casualty age group, casualty sex to see how reliable level 2 models were when fitted to a relatively small dataset compared to the 10Y models developed in the previous study. The model tested all variables in the SOA level data as detailed in table 2.1. The same model specification was fitted to each age and sex combination to see how the explanatory variables representing environmental and other factors behave in the models across each of the age and sex stratifications.
- 5.39 Given that the number of casualties in the 3-year dataset is approximately four times less than in the 11-year data, it was decided not to stratify the data by URDG classification and instead include this as a term in the model.
- 5.40 The first model revealed a problem with the LGD (Local Government District) variable. The LGD variable specifies 26 areas each which produced its own model term (minus one base-line area which is incorporated in the Intercept term). Several of the area had excessively high model coefficients suggesting that it was inter-correlated with some other component in the model. It was unclear exactly what LGD was inter-correlated: one explanation was that it was auto-correlated with the data itself which is aggregated by SOA since LGD areas are essentially aggregations of SOAs. Consequently LGD was removed. The decision to do so was supported by the fact that whilst there is interest in understanding which areas were associated with high or low casualty rates, the spatial relationships in the data will be explored in finer detail in task C4 where the model results are used to identify the best and worse performing SOAs.
- 5.41 A series of 5 further models were then fitted which tested all the variables excluding those variables which had been identified as being inter-correlated with another variable in the preceding model. The following list details the variables excluded at each stage in this series of models beginning, for completeness, with the first model discussed above:
- lgd - Local Government District – inter-correlated to SOAs; lgd is redundant since spatial dimension is integral to the SOA level dataset.
 - tenown and ten1-7 - Housing tenure: percentage of households owned by resident – inter-correlated with tenren; all the tenure variables are essentially different categorisations of the same information and only one is needed.
 - rdtot – Total road length in SOA - inter-correlated with M, A, B, C, D & U CLASS road lengths; chose to remove rdtot since this is a composite measure based on the classified road lengths and any relationships between the specific road categories would be more informative.

- inc, emp & idac – MDM income, employment idac domains – inter-correlated with MDM; All four variables are highly inter-correlated with each other so it was necessary to exclude a minimum of three. The model has several other terms related to employment. Despite being a composite variable MDM was considered to be more informative than income and there was merit in retaining MDM as the central measure to the study.
 - rdenha – Road density per hectare – inter-correlated with area and road length variables; rdenha removed (rather than area or road length) because it is a composite of area and road length.
 - MCLASS – Road length of motorways – inter-correlated with SOAs (only a relatively small number of SOAs are served by motorways)
- 5.42 Having identified all the inter-correlated variables a seventh model was fitted with all the inter-correlated variables removed to test all the remaining variables. The seventh model was fitted to each combination of age group and sex to see how the model performed across the strata. With the inter-correlated variables now excluded it was possible to a critical look at the model outputs and assess the model performance.
- 5.43 The first observation was that a large number of the models were being flagged with warnings indicating that the models were not converging properly. This problem is associated with the low sample sizes in individual SOAs and is a consequence of modelling 3-year data rather than 10-year data used in the previous study. This problem manifested itself in the results as large inconsistencies, across the age and sex categories, in the signs of the model coefficients. The varying signs represent unstable positive and inverse relationships between casualty rates and the explanatory factor, which could be found even when attention was confined to just the significant variables. There was also a large degree of variability in the level of significance across age and gender categories.
- 5.44 Collectively these observations suggested that level 2 models were not performing as they should when fitted to 3-year data. To get a better idea of how well the models were performing, a matrix of charts was produced to illustrate the extent to which the models are representing the casualty data by plotting the observed number of casualties against the number predicted by the model as presented below.
- 5.45 The matrix contains the following six charts:
- Male+Female pedestrian+cyclist casualties aged 0-4 in the 11-year data to provide a point of reference
 - Male+Female pedestrian+cyclist casualties aged 0-15 in the 3-year data
 - Male+Female pedestrian+cyclist casualties aged 0-4 in the 3-year data
 - Male+Female pedestrian+cyclist casualties aged 5-7 in the 3-year data
 - Male+Female pedestrian+cyclist casualties aged 8-11 in the 3-year data
 - Male+Female pedestrian+cyclist casualties aged 12-15 in the 3-year data



5.46 The plots show that even for the 11 year data, the model fit for males+female pedestrian+cyclist casualties aged 0-4 is very approximate. The plot illustrates how the model has fitted the observed number of casualties per SOA which range from 0 to 7 (i.e. eight distinct levels) using explanatory variables for each of the 890 SOAs. It shows that for SOAs with zero observed number of casualties the modelled number of casualties ranges from 0 through to approximately 2.5 – quite a large degree of variation. The bottom four charts show how sparse the male+female casualty numbers are for individual age categories and very little evidence of a any correlation i.e. with points positioned about the diagonal regression line.

5.47 The plot of males+female pedestrian+cyclist casualties aged 0-15 shows the observed and modelled casualties are beginning to align about the regression line and showing a basic level of correlation between the observed and modelled casualty numbers. Based on the a observation made on the the model outputs and the plots above it was decided that the next stages of the

model development would focus solely on models of the 3-year males+female pedestrian+cyclist casualties aged 0-15.

- 5.48 Whilst there is a potential to loose insight that might be gained from having individual models models for specific age groups, developing a model for all children as a single group will still produce a single model specification as was the outcome in the previous child deprivation and road safety study.
- 5.49 The level 2 model outputs of the 3-year males+female pedestrian+cyclist casualties aged 0-15 were inspected to identify the which variables were significant and were to be put forward to stage 4 of model development to determine a final Level 3 model specification. The following table summarises the outcome of stage 3 of the model development. Variables which were significant at the 95% level were classified as significant whilst variables which were significant at the 99% level or above were classified as highly significant. All other variables were classified as not significant.

Table 5.3 Summary of the outcomes of the stage 3 model development

Category	Variable name	Description	Outcome
Area	sun	Annual sunshine hours	Highly significant
Demo	empres	Number of employed people resident in SOA	Highly significant
Demo	noqualspc	Percentage of population with no qualification	Highly significant
Demo	scpl14	School places: age 5-15	Highly significant
Demo	fsm_ppp	Percentage free school meals: primary and post primary	Significant
Deprivation	health	MDM Domain – Health deprivation and disability	Highly significant
Deprivation	prx	MDM Domain - Proximity to services	Highly significant
Deprivation	env	MDM Domain - Living environment	Highly significant
Deprivation	crm	MDM Domain - Crime and disorder	Highly significant
Road	tproxy	A relative measurement of travel activity in an SOA	Highly significant
Road	jno	Number of junctions in SOA	Highly significant
Road	UCLASS	Road length – U class	Highly significant
Composite	jdenkm	Junction density per road km	Highly significant
Composite	jdenha	Junction density per hectare	Highly significant
Composite	tproxrd	travel activity per km of road	Highly significant
Area	area	Square area (ha)	Not significant
Area	popden2	Population density per hectare	Not significant
Area	urbrur	Urban-Rural Definition Group (URDG) Classification	Not significant
Area	rain	Annual rainfall (mm)	Not significant
Demographic	nocar	Percentage of households with no car	Not significant
Demographic	tenren	Housing tenure: percentage of households rented	Not significant
Demographic	empwrk	Number of workplace population: the number of persons aged 16-74 in employment in the SOA	Not significant
Deprivation	mdm	MDM score	Not significant
Deprivation	educ	MDM Domain - Education, skills and training	Not significant
Road	ACLASS	Road length – A class	Not significant

Road	BCLASS	Road length – B class	Not significant
Road	CCLASS	Road length – C class	Not significant
Composite	tproxj	travel activity per junction	Not significant
Composite	popj	SOA population per junction	Not significant
Composite	empresj	Number of employed people resident in SOA per junction	Not significant
Composite	empwrkj	Number of workplace population in SOA per junction	Not significant
Area	lgd	Local Government District	Removed - inter-correlated
Demographic	tenown	Housing tenure: percentage of households owned by resident	Removed - inter-correlated
Demographic	ten1-7	Housing tenure: percentage of households in individual categories of tenure	Removed - inter-correlated
Deprivation	inc	MDM Domain – Income	Removed - inter-correlated
Deprivation	emp	MDM Domain – Employment	Removed - inter-correlated
Deprivation	idac	MDM Domain - Income deprivation affecting children	Removed - inter-correlated
Road	rdtot	Total road length in SOA	Removed - inter-correlated
Road	MCLASS	Road length – M class	Removed - inter-correlated
Composite	rdenha	Road length per hectare	Removed - inter-correlated

5.50 Probably the most surprising outcome of stage 3 model development finding that MDM score is no longer significant variable. The reason for this is the presence of variables for the individual MDM domains – health, education, proximity to services, the environment and crime which are five of the seven components of the composite multiple deprivation measure.

5.51 Detailed below are further observations made on the level 2 model:

- the MDM Crime domain score is a highly significant positively correlated variable – the more crime related to the SOA the greater the casualty rate;
- the MDM Proximity to Services domain score is a highly significant inversely correlated variable – the less deprived the SOA is in terms of proximity to services the greater the casualty rate;
- the MDM Health deprivation and disability domain score is a highly significant positively correlated variable – the more health deprivation related to the SOA the greater the casualty rate;
- the MDM Living Environment domain score is a highly significant inversely correlated variable – the less deprived the SOA is in terms of the living environment the greater the casualty rate;
- population density is found to be a significant positively correlated variable for half the models examined – the greater the density, the greater the casualty rate;
- road length of unclassified roads is a significant positively correlated variable – the longer the length of unclassified roads, the greater the casualty rate;

- the traffic proxy variable which provides a relative measure of travel activity in the SOA is a highly significant positively correlated variable – the more activity there is in the SOA, the greater the casualty rate;
- the number of school places is a highly significant positively correlated variable – the more school places, the greater the casualty rate;
- the proportion of free school meals is a significant inversely correlated variable – the smaller the take up of free school meals provision in the SOA the greater the casualty rate;
- sunshine hours is seen to be a highly significant generally inversely correlated variable – the greater the number of sunshine hours, the lower the casualty rate;
- the proportion of people with no qualifications variable is a highly significant positively correlated variable – the larger the proportion of people with no education the greater the casualty rate;
- the number of employed residents variable is a highly significant positively correlated variable – the larger the number of employed residents the greater the casualty rate. This is counter to the notion that more employed, the less deprived an area, the smaller the casualty rate but when employees are considered as a travel proxy (due to commuting) this is plausible; and
- the number of junctions is a highly significant inversely correlated variable whilst the number of junctions per area and km of road are highly significant positively correlated variables. The variables are closely related and their effects should be considered in combination with each other. The model coefficients suggest that the overall, average effect of these variables is positive – the more junctions the higher the casualty rate.

Stage 4 – Level 3 model

5.52 Following on from the Level 2 modelling, the final model included 15 variables. Each variable is significant at the 99% level with the exception of the proportion of free school meals variable which is significant at the 95% level. The list of the variables, model coefficients and significance codes for the final model is:

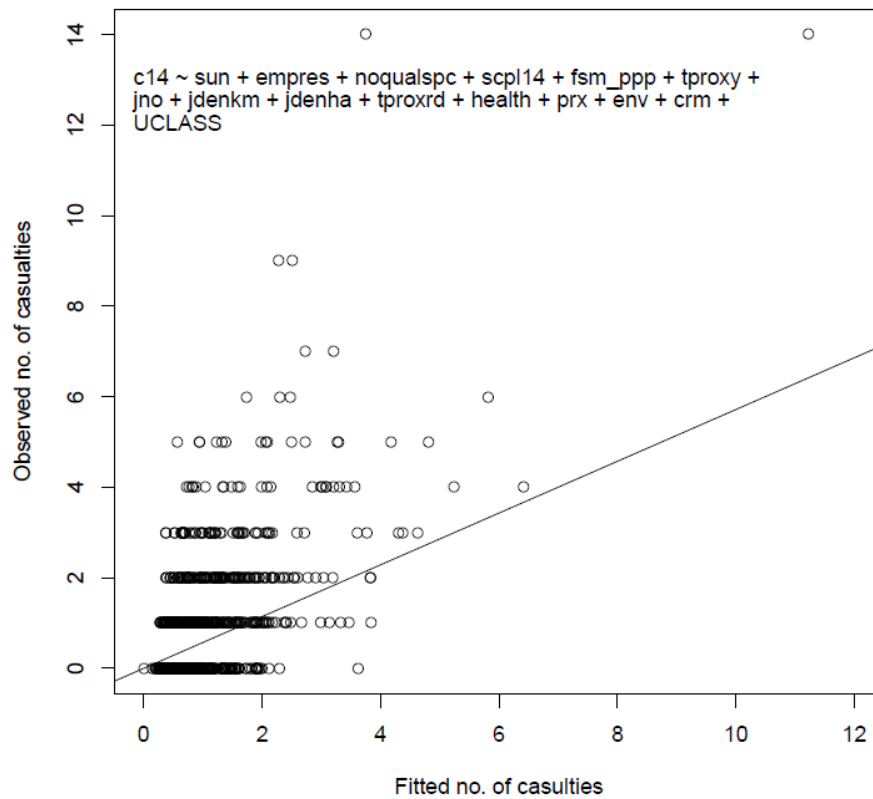
Table 5.4 Results of the final, level 3 model of April 2007 to March 2010 pedestrian+cyclist casualties

Variable name	Description	Coefficient	Significance code
Intercept		-0.7689	***
Sun	Sunshine hours	-0.0003	**
empres	Number employed	0.0004	***
noqualspc	Percentage no qualifications	0.0116	***
Scpl14	Number of school places	0.0004	***
fsm_ppp	Percentage free school meals	-0.0023	*
tproxy	Travel activity	0.0006	***
Jno	Number of junctions	-0.0050	***

jdenkm	Junctions per road km	0.0369	**
jdenha	Junctions per hectare	0.1814	***
tproxrd	Travel activity per km of road	-0.0016	**
health	MDM - Health	0.3010	***
Prx	MDM - Proximity to services	-0.0960	***
Env	MDM - Living environment	-0.0096	***
Crm	MDM - Crime and disorder	0.0112	***
UCLASS	Road length	0.0045	*

- 5.53 As identified in paragraph 5.51, the sign of the coefficient for some of the modelled variables are negative suggesting there is an inverse relationship between casualty rate and the variables when intuitively the relationship is expected to be positive, and vice versa. Occasionally this can correct our understanding of the relationship between the two, by prompting a reassessment of what appeared first to be common sense. More often, however, counter intuitive results can be explained through the process of interpreting the model results as a whole rather than individual factors in isolation.
- 5.54 Individual model coefficients in models with more than variable represent the contribution of the variable in the presence of all other variables in the model. Individual model coefficients represent the change in the number of casualties for each unit increase in the variable relative to the baseline category where all other variables are zero. When there is correlation between two (similar) explanatory variables in the model, the model will be fitted with one variable dominating the other (in the presence of all other variables in the model) and the variable being dominated compensates by countering the effect of the other to achieve a balance.
- 5.55 A good example of this effect was observed in the jno, jdenkm, jdenha variables which are clearly related but different measures of the number of junctions. jdenha is dominating jdenkm and jno variables (0.181 versus 0.039, -0.005 respectively): jno in isolation suggests that the number of junctions has an inverse relationship with the number of casualty. However the three variables are related: if the number of junctions increases in real life, the values in each variable would also increase. Therefore, the three modelled variables need to be considered in combination. The overall effect of an increase in the number of junctions can be approximated as $0.181 + 0.039 + -0.005 = 0.213$ - a positive relationship. In reality, this example is not as clear cut as this, these variables may also interact with other variables so that the overall effect of the number of junctions is likely to be slightly different to illustration of the overall effect of the three junction variables (i.e. 0.213).
- 5.56 The following chart presents a plot of the observed casualty numbers against modelled casualty numbers for the final level 3 model.

M+F Ped/ cyc 0-15 (3Y)



- 5.57 Even at the relatively high level of aggregation of the 3-year data to a single group of all children (male+female, 0-15) and without stratification by URDG classification, there is a large amount of variation in the modelled casualty numbers for any particular level of observed casualties. The variation is much greater than that seen in the equivalent level 1 model fitted to the aggregated data. This is because, the process of aggregating produces average values which removes a large amount of the variation seen at the SOA level.
- 5.58 To understand how well the final 3 year model of pedestrian+cyclist casualties was performing, stages 3 and 4 of the model development process was repeated using the 11-year casualty dataset. This meant that the 11-year model development started with the full list of explanatory variables as presented in Table 5.3, minus the variables identified in stage 2 of the 3-year model development as correlated. The same set of explanatory variables are correlated in both the 3-year and 11-year data sets since the only difference between the two datasets are variables containing the number of casualties (the dependent variables).
- 5.59 The following table compares the final (level 3) model results for the 3-year and 11-year data. The results for the final 3-year model are the same as those presented in Table 5.4.

Table 5.5 Final, level 3 model results for the 3-year and 11-year pedestrian+cyclist casualties

Variable	Description	3-year final model	11-year final model
(Intercept)		-0.7689 ***	1.7560 ***
area	Area of SOA		0.0001 ***
urbrurU	Urban/ Rural classification		-0.0876 *
rain	Rain fall (mm)		-0.0002 ***
sun	Sunshine hours	-0.0003 **	-0.0006 ***
tenren	Proportion rented tenure		-0.0077 ***
empres	Number employed	0.0004 ***	0.0008 ***
noqualspc	Percentage no qualifications	0.0116 ***	0.0043 **
scpl14	Number of school places	0.0004 ***	0.0002 ***
fsm_ppp	Percentage free school meals	-0.0023 *	0.0033 ***
tproxy	Travel activity	0.0006 ***	
jno	Number of junctions	-0.0050 ***	
jdenkm	Junctions per road km	0.0369 **	
jdenha	Junctions per hectare	0.1814 ***	
tproxrd	Travel activity per km of road	-0.0016 **	
popj	Population per junction		0.0080 ***
empresj	Employees resident per junction		-0.0257 ***
empwrkj	Employees workplace per junction		0.0055 ***
health	MDM – Health	0.3010 ***	0.2396 ***
educ	MDM - Education		0.0072 ***
prx	MDM - Proximity to services	-0.0960 ***	-0.1431 ***
env	MDM - Living environment	-0.0096 ***	
crm	MDM - Crime and disorder	0.0112 ***	0.0113 ***
CCLASS	Road length of C-class roads		-0.0124 ***
UCLASS	Road length of unclassified roads	0.0045 *	-0.0046 ***

5.60 At the most basic level of comparison, there is a substantial overlap in the explanatory variables determined in the development of models for the 3-year and 11-year pedestrian+cyclist casualties. Nine of the 24 (excluding the intercept) appear in both models. Other similarities include:

- the variables sun, empres, noqualspc and scpl14 have the same sign and are of similar scale in both models. In the previous study based on 10-year data, sun was also found to be significant, be negatively correlated and be similar in scale;
- both models have four of five MDM domains and neither has the composite MDM score measure as a significant variable. The signs of the model coefficients are the same and similar in scale for those MDM domains appearing in both models.

5.61 There are also a number of differences:

- the intercept is considerably less for the 3-year model than for the 11-year model. This simply reflects the difference in the volume of casualties being modelled in the 3-year data compared to the 11-year data;

- urbrur and rain are significant in the 11-year data but not in the 3-year data. The previous study based on 10-year data also found these factors to be significant;
- fsm_ppp was found to be significant in models for both the 3 year and 11-year data but not in the 10 year data. The 3-year model suggests that there is a negative correlation between free school meals provision and casualty rates whereas the 11-year model suggests the reverse;
- there are a block of eight variables: tproxy, jno, jdenkm, jdenha, tproxrd, popj, empresj and empwrkj which are interrelated as follows:
 - tproxy and tproxrd are derived in part from empres and empwrkj which appear in the model as equivalent relative measure empresj and empwrkj
 - popj, empresj and empwrkj are derived in part from jno which also feature in the model as jdenkm and jdenha

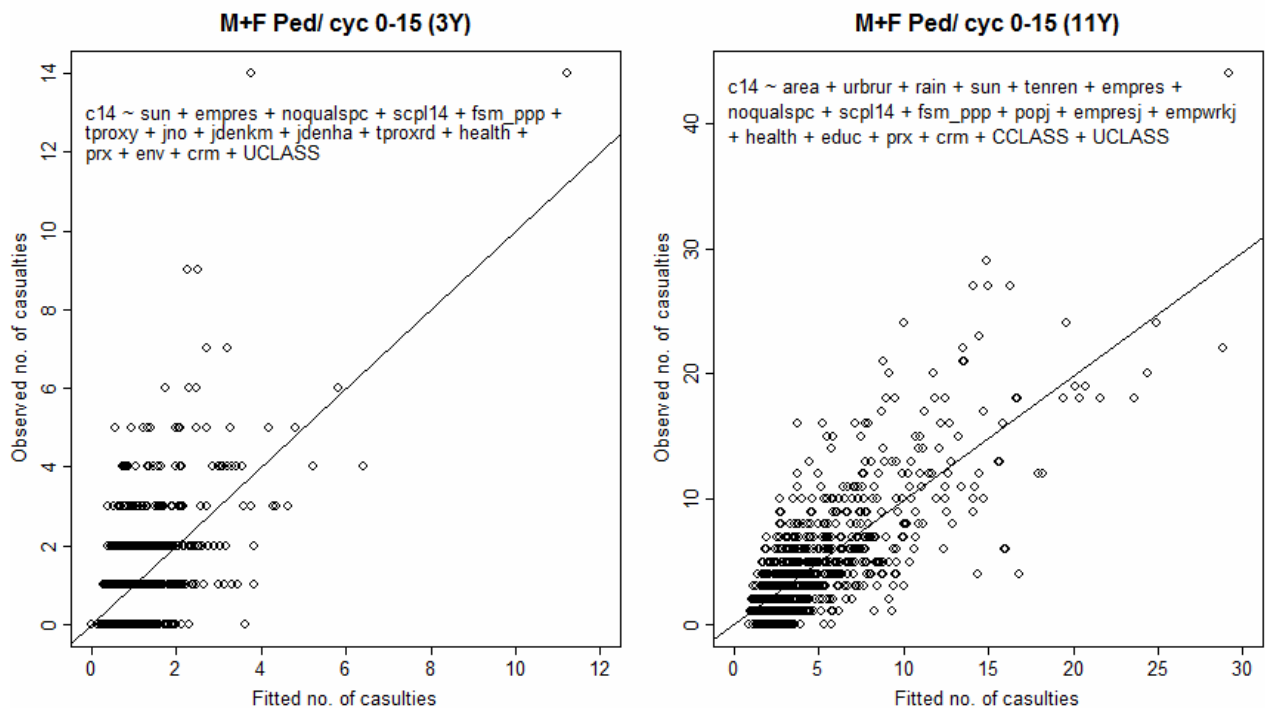
Two alternative formulations of the 11-year model were tested which demonstrated that it was possible to:

- i. replace popj, empresj and empwrkj with tproxy and tproxrd, and
- ii. replace popj, empresj and empwrkj with jno, jdenkm and jdenha

with very little degradation in the model performance - all variables in these models remained significant except jno in model test ii. This illustrated the fact that, whilst there are differences in the two models with regard to the block of eight variables, the sub blocks of variables can be interchanged to improve the match between the 3-year and 11-year model specifications.

- Finally, the 3 year model suggests that there is positive correlation between length of unclassified roads in the SOA whereas the 11 year data suggests the contrary.

- 5.62 The last aspect of the 3-year model tested was to compare the model fit of the 3-year and 11-year models. The following chart contrasts the ability of the models to predict the number of casualties by comparing the observed and modelled estimates of the number of casualties for the two models.
- 5.63 The charts show that the model fitted to the 11-year data performs considerably better than the 3-year model. This is largely due to the fact that there is nearly four times as much casualty data in the 11-year data than in the 3-year data rather than a result of arriving at different model specifications.



6 Task C4

- 6.1 For each SOA the level 3 models provide a predicted value or fitted value for the number casualties in the area. The difference between the fitted values and the observed number of casualties can be used as a measure of road safety performance for the area. Where the difference is a positive value it implies that the area has more casualties than might be expected and the higher the value the poorer the performance. Conversely, where the difference is a negative value it implies that the area has fewer casualties than might be expected and the higher the value the better the performance.
- 6.2 The following two tables list the 30 poorest performing SOAs and 30 best performing SOAs based on the model of 3-year data. A review of the poorest performing SOAs may help to identify and inform the development remedial action or initiatives to child casualties whilst a review of the best performing SOAs may highlight some examples of good practice which could be transferred to other areas of Northern Ireland. Because the poorest and best performing SOA have been identified from the model fitted to 3-year data, this provides an opportunity to make

a more up-to-date assessment of areas since the extent of demographic and infrastructure changes over this period is generally going to be relatively small.

Table 6.1 30 best performing SOAs between April 2007 and March 2010

SOA code	SOA name	Observed	Modelled	Difference
95PP09W1	Devenish	0	3.6	-3.6
95LL13S2	Drumgor_2	1	3.8	-2.8
95SS13S1	Hilden_1	1	3.5	-2.5
95GG47S1	Water Works_1	4	6.4	-2.4
95ZZ10W1	North	1	3.3	-2.3
95VV28S2	St Patrick's_2	0	2.3	-2.3
95GG19S1	Crumlin_1_Belfast	1	3.1	-2.1
95HH15W1	Victoria_Carrickfergus	0	2.1	-2.1
95OO18W1	Killymeal	0	2.0	-2.0
95DD11W1	Dunclug	1	3.0	-2.0
95GG39S3	Shaftesbury_3	0	2.0	-2.0
95SS29S2	Twinbrook_2	0	1.9	-1.9
95GG07S1	Beechmount_1	0	1.9	-1.9
95GG24S2	Fortwilliam_2	0	1.9	-1.9
95JJ06W1	Churchland	0	1.9	-1.9
95FF07W1	Edenderry	0	1.9	-1.9
95GG12S4	Botanic_4	0	1.9	-1.9
95MM23W1	Rosemount	0	1.9	-1.9
95MM28W1	The Diamond	2	3.8	-1.8
95MM27S1	Strand_1_Derry	2	3.8	-1.8
95GG50S3	Woodstock_3	0	1.8	-1.8
95YY09W1	Drumragh	0	1.8	-1.8
95EE13W1	Route	0	1.7	-1.7
95VV15S1	Drumgullion_1	0	1.7	-1.7
95RR10W1	Greystone_Limavady	1	2.7	-1.7
95GG21S3	Falls_3	3	4.6	-1.6
95YY15S2	Lisanelly_2	0	1.6	-1.6
95GG51S1	Woodvale_1	0	1.6	-1.6
95CC22W1	The Mall	0	1.6	-1.6
95GG43S1	Sydenham_1	0	1.6	-1.6

Table 6.2 30 worst performing SOAs between April 2007 and March 2010

SOA code	SOA name	Observed	Modelled	Difference
95XX05W1	Bangor Castle	14	3.7	10.3
95MM05W1	Brandywell	9	2.3	6.7
95GG44S1	The Mount_1	9	2.5	6.5
95DD02S2	Ahoghill_2	5	0.6	4.4
95GG04S3	Ballymacarrett_3	7	2.7	4.3
95PP17W1	Lisnaskea	6	1.7	4.3
95PP06S2	Castlecoole_2	5	0.9	4.1
95KK08W1	Moneymore	5	0.9	4.1
95HH09W1	Killycrot	7	3.2	3.8
95SS06S1	Collin Glen_1	5	1.2	3.8
95GG44S2	The Mount_2	6	2.3	3.7
95GG25S3	Glen Road_3	5	1.3	3.7
95NN17W1	Murlough	5	1.4	3.6
95DD08W1	Castle Demesne	6	2.5	3.5
95UU14W1	Knocklaid	4	0.7	3.3
95BB13S2	Gregstown_2	4	0.8	3.2
95BB19S1	Movilla_1	4	0.8	3.2
95GG06S2	Ballysillan_2	4	0.8	3.2
95YY13S1	Gortrush_1	4	0.9	3.1
95RR05S2	Enagh_2_Limavady	5	2.0	3.0
95WW19S1	Jordanstown_1	4	1.0	3.0
95KK16W1	Tullagh	5	2.1	2.9
95GG35S3	New Lodge_3	5	2.1	2.9
95GG39S1	Shaftesbury_1	14	11.2	2.8
95GG26S2	Glencairn_2	4	1.3	2.7
95JJ22W1	Waterside	4	1.4	2.6
95SS15S2	Hillsborough_2	3	0.4	2.6
95NN19S2	Saintfield_2	3	0.4	2.6
95GG17S2	Cliftonville_2	4	1.5	2.5
95GG48S2	Whiterock_2	5	2.5	2.5

- 6.3 The analysis was repeated on the 11-year data to identify the extent to which the list of the 30 best and worst performing SOA changes between the 3-year and 11-year periods. The following two tables list for the 11-Year data the best and worse performing SOAs respectively.

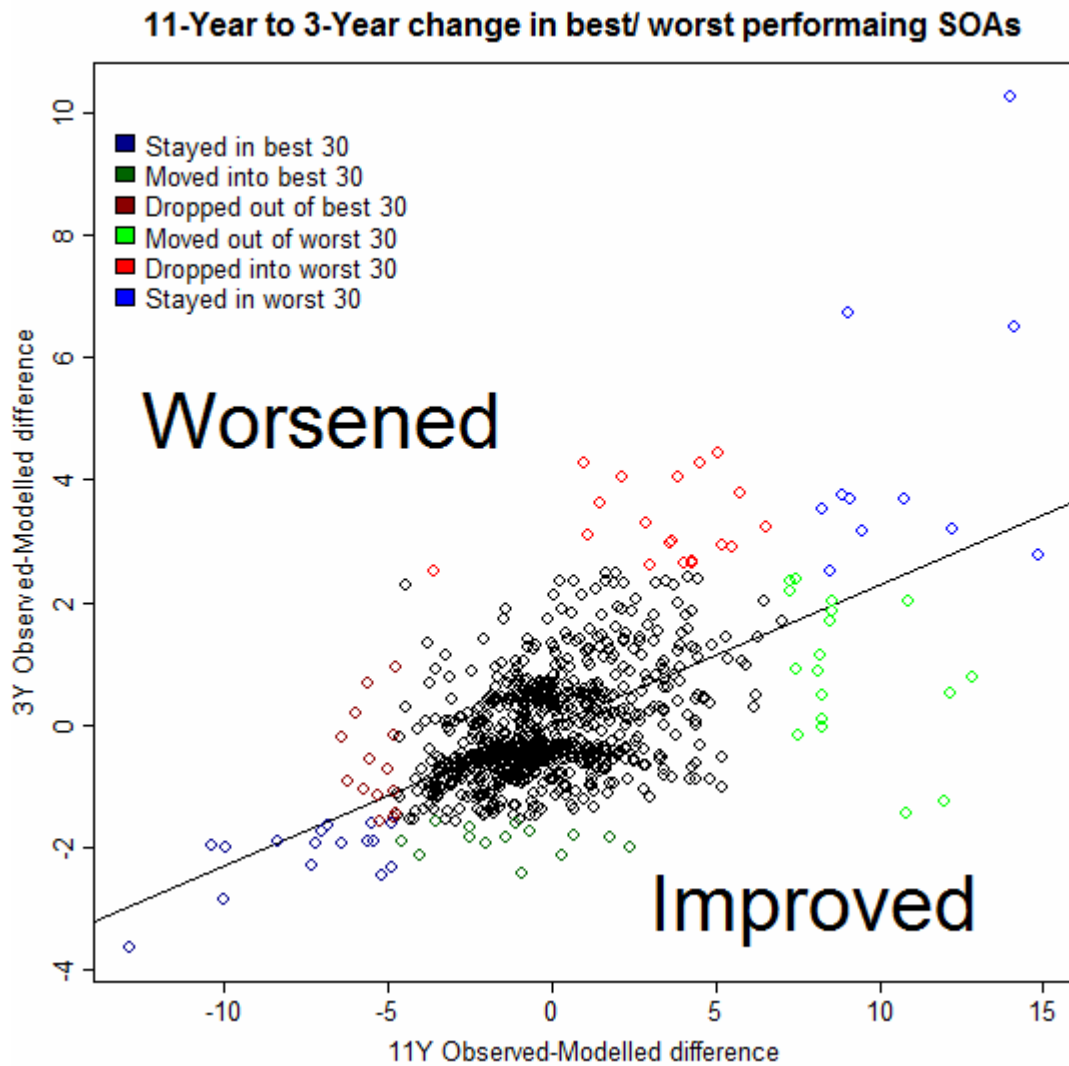
Table 6.3 30 best performing SOAs between January 1999 and March 2010

SOA code	SOA name	Observed	Modelled	Difference
95PP09W1	Devenish	4	16.84129	-12.8413
95GG39S3	Shaftesbury_3	4	14.36493	-10.3649
95LL13S2	Drumgor_2	6	15.99881	-9.99881
95DD11W1	Dunclug	6	15.90232	-9.90232
95FF07W1	Edenderry	1	9.342682	-8.34268
95VV28S2	St Patrick's_2	1	8.274691	-7.27469
95GG24S2	Fortwilliam_2	2	9.183821	-7.18382
95VV15S1	Drumgullion_1	3	10.02266	-7.02266
95GG21S3	Falls_3	22	28.80354	-6.80354
95SS29S2	Twinbrook_2	3	9.37475	-6.37475
95SS16S2	Kilwee_2	6	12.37455	-6.37455
95GG40S2	Shankill_2	12	18.18398	-6.18398
95RR03W1	Coolessan	12	17.93224	-5.93224
95NN07S1	Cathedral_1	0	5.722805	-5.72281
95MM23W1	Rosemount	4	9.588961	-5.58896
95GG21S1	Falls_1	18	23.57649	-5.57649
95LL09S2	Court_2	4	9.521832	-5.52183
95GG51S1	Woodvale_1	3	8.479974	-5.47997
95GG12S4	Botanic_4	5	10.41405	-5.41405
95QQ03W1	Ballyloran	0	5.302066	-5.30207
95CC01W1	Abbey Park	1	6.213757	-5.21376
95SS13S1	Hilden_1	9	14.15798	-5.15798
95GG11S3	Bloomfield_3_Belfast	2	7.001033	-5.00103
95CC22W1	The Mall	1	5.873567	-4.87357
95ZZ10W1	North	8	12.86072	-4.86072
95VV27W1	St Mary's	4	8.799507	-4.79951
95LL14S1	Drumnamoe_1	1	5.789467	-4.78947
95GG12S3	Botanic_3	1	5.773635	-4.77364
95NN04W1	Ballymote	10	14.7649	-4.7649
95GG28S3	Highfield_3	4	8.748066	-4.74807

Table 6.4 30 worst performing SOAs between January 1999 and March 2010

SOA code	SOA name	Observed	Modelled	Difference
95GG39S1	Shaftesbury_1	44	29.14839	14.85161
95GG44S1	The Mount_1	29	14.85906	14.14094
95XX05W1	Bangor Castle	24	9.996577	14.00342
95GG18S2	Clonard_2	27	14.13566	12.86434
95BB19S1	Movilla_1	16	3.766708	12.23329
95GG01S2	Andersonstown_2	21	8.838796	12.1612
95BB06W1	Central_Ards	27	15.02765	11.97235
95GG29S2	Island_2	20	9.116116	10.88388
95MM12S3	Crevagh_3	16	5.202408	10.79759
95GG44S2	The Mount_2	27	16.26817	10.73183
95GG06S2	Ballysillan_2	15	5.53485	9.46515
95GG25S3	Glen Road_3	15	5.858433	9.141567
95MM05W1	Brandywell	18	8.941534	9.058466
95SS06S1	Collin Glen_1	16	7.127594	8.872406
95II05S2	Carryduff East_2	13	4.459046	8.540954
95GG47S3	Water Works_3	23	14.46351	8.536485
95GG02S3	Ardoyne_3	22	13.46941	8.530591
95GG17S2	Cliftonville_2	18	9.520005	8.479995
95LL02S2	Annagh_2	16	7.73594	8.26406
95NN07S2	Cathedral_2	17	8.748482	8.251518
95DD08W1	Castle Demesne	20	11.75667	8.243332
95GG05S1	Ballynafeigh_1	14	5.759012	8.240988
95WW11S2	Collinbridge_2	12	3.773509	8.226491
95II13W1	Graham's Bridge	16	7.892902	8.107098
95SS06S3	Collin Glen_3	21	13.48914	7.510856
95GG31S3	Ladybrook_3	15	7.542568	7.457432
95GG25S1	Glen Road_1	21	13.55754	7.442457
95ZZ14S2	South_2	11	3.722747	7.277253
95MM12S1	Crevagh_1	10	2.747785	7.252215
95KK04W1	Gortalowry	12	4.94861	7.05139

- 6.1 There are substantial differences in the lists of best and worse performing SOAs:
- only sixteen of the 30 best performing SOAs in the 3-Year period are in the best 30 for the 11-Year data;
 - only eleven of the 30 worst performing SOAs in the 3-Year period are in the worst 30 for the 11-Year data.
- 6.2 The differences can be attributable to differences in the models developed for the 3-year and 11-year data and differences in the data itself.
- 6.3 The following chart illustrates the changes in the rankings of 890 SOA between the 11-year and 3-year models/ data.



- 6.4 The main reason why there are substantial differences is the ranking of SOAs in the 3-year model / data is done over a relatively small range of observed-modelled differences in casualty numbers. For instance, a relatively small change from 1 to 2 in the observed number of casualties in an SOA in the 3-year data among those where the points on the above chart are densely packed could (depending upon the extent to which the model could reflect the additional casualty) result in change in the ranking by 100 or more places: There are over 100 SOAs which have a difference in the range 1.0 to 2.0.
- 6.5 In contrast the 11-year model and data cover a much wider range of casualties per SOA meaning each ranking position in turn represents a much larger difference in the observed and modelled number of casualties. Whilst there is clearly merit in investigating SOAs whose performance ranks in the top or bottom 30, the above analysis demonstrates that relative measures such as rank are best considered in parallel with measures of scale (the number of casualties).
- 6.6 The analysis above can also be expected to partly represent real changes that have occurred since April 2007. However there are a number of points to bear in mind when comparing performance in this way, from these models / data:

- the time periods of the data overlap (post April 2007 is reflected in both models);
- the model specifications are different (some differences are attributable to the models themselves); and
- the above analysis compares two relative measures so a drop in the rank of an SOA does not necessarily mean its performance has dropped – the casualty rate may have improved (reduced) but the casualty rate in the majority of other SOAs may have improved to a greater extent leading to a drop in rank.

Comparisons with Previous Research

6.7 Comparisons were also made with the results of the previous research by checking the top and worst performing SOAs using the 11 year model against the previous 10 year model. In general terms there was good agreement of the worst performing SOAs:

- 22 of the 30 worst performing SOAs were common; and
- 26 of the 30 worst performing SOAs in the 11 year model appeared in the 50 worst performing SOAs in the previous 10 year model.

6.8 The best performing SOAs showed only a fair level of agreement:

- 9 of the 30 best performing SOAs were common; and
- 16 of the 30 best performing SOAs in the 11 year model appeared in the 50 best performing SOAs in the previous 10 year model.

6.9 It should be noted that the list of best performing SOAs is more liable to change with the addition of further years of data than the worst performing SOAs. This is because new casualties over the additional year may substantially increase the casualty rate of a best performing SOA with a previously low casualty rate. By contrast the casualty rate of a worst performing SOA will not change so substantially with the addition of zero casualties over the final year.

Implications for Project

6.10 As noted in paragraph 6.9, the list of best performing SOAs is liable to change with the addition of further years of data and therefore is a relatively 'unstable' list. A desktop study was made of the best six performing SOAs and the following noted:

- only 3 were in the 10% most deprived SOAs using the MDM, the other 3 were in the 10-20% and 20 – 30% most deprived bands;
- the best performing 3 had distinctive road / street layouts which would tend to reduce pedestrian casualties:
 - Devenish, in Enniskillen, has limited through routes due to the local river systems, so leading to low traffic volumes and speeds;

- similarly Shaftesbury 3, adjacent to City Hospital has a local road layout affected by the railway tracks and by the hospital site which leads to the formation of cul de sacs;
 - Drumgor 2, in Craigavon, has a Radburn Layout, with residential cul de sacs served by wider sweeping distributor roads with limited pedestrian access.
- All of the SOAs had some degree of traffic calming – generally speed humps.
- 6.11 It was noted also that the distinctive road / street layouts, which seemed to contribute to the reduced casualty rate, had been identified in the focus group research to some degree. It was therefore concluded that there was limited benefit in studying the list of best performing SOAs in greater detail for the purposes of this research.
- 6.12 However, the approach of identifying ‘best performing’ and ‘worst performing’ SOAs and SOAs with highest and lowest casualty rates is clearly instructive in identifying areas with special characteristics. The use of the 3-year and 11-year datasets adds a further dynamic dimension. It is recommended therefore that in future the approach should be considered alongside the use of standard casualty records in order to identify areas for intervention.