An Information Statistics Approach to Zone Design targeting

The Traffic Accidents of Children in Tyne and Wear

Dr. Daras Konstantinos

GIS/IT Department

Eratosthenes S.A., Greece

1. Introduction

Accidental injury in young children is a crucial cause of mortality, morbidity and disability in developed countries. In the UK a considerable proportion of health care resources are allocated to reduce children accidents (Towner et al., 1993). One of the most persisting unintentional injuries is the traffic accidents affecting 1-19 years old children. In the early 1990s the number of child pedestrian fatalities in Britain relative to child population, was considerably larger than the EU average (DfT, 2003). The UK government has set a clear target of 40% reduction of accidents by the end of 2010. In this study, we will investigate the relationship between the traffic accidents of children and the deprivation in the Super Output Areas (SOAs) of Tyne and Wear.

The newly proposed methodology for identifying the most informative aggregation level is tested using traffic accidents of children in Tyne and Wear. Thus, we will provide an automated methodology for selecting the aggregation level in which analysis is going to take place.

Furthermore, the Areas to Zones system (A2Z) will construct two new aggregation sets at the Akaike Information Criterion (AIC) proposed level targeting homogeneous zones in terms of the children traffic accident rates and the Townsend deprivation index respectively. During the zone design process, the A2Z will control the shape compactness of zones in a very weak manner because the aim of this study is to maximise the differences between zones. This way, it will be possible to clearly identify the zones with high rates or scores.

2. Children's' traffic accidents

The majority of studies on childhood injury have concluded that there is higher risk among children from deprived areas. For example, Abdalla et al. (1997a; 1997b) and Abdalla (1997) noted higher rates of casualties in deprived than affluent areas of Lothian region in Scotland. Their research conducted at census output areas level calculating the distances between the output area of residents and the accident locations. Although, their findings appeared to be related with the class of road and social status of output areas, they did not explore the aggregation effects of their study. On the other hand Sharples et al. (1990) came into the similar conclusion using individual data records in Northern Regional Health Authority and the social status of children involved in fatal accidents at ward level.

Additionally, attempts to explain social differences in traffic accidents have suggested a strong correlation between the social class and injury mortality. Christie (1995) stated that the risk of children traffic accidents is strongly class related. She indicated that children in the lowest social classes are over four times more likely to be killed than children in highest social classes. Her results were based on mortality rates and socioeconomic characteristics in the UK at individual or household level. Similar studies suggested that the social gradients in injury mortality exceed those for any other cause of death in young people, and the inequalities between different socioeconomic groups are higher in relation to child pedestrian deaths than all causes of death in children (Jarvis, 1999; Roberts and Power, 1996).

In the UK the majority of studies are conducted using accident and socioeconomic data at individual and household level. Recent findings have shown high injury rates of children in households with single parents (Judge and Benzeval, 1993; Roberts and Pless, 1995) while a strong association to traffic accidents have been seen in poor housing conditions including type, quality, and tenure of housing as well as overcrowding (Alwash and McCarthy, 1988; Runyan et al., 1992). The general conclusion from these studies suggested that deprivation determinants have strong relationships to traffic accidents among children. Generally speaking this may be true but it could be an example of ecological fallacy as the high rates of accidents in deprived areas could be result of risk factors on top of household deprivation (Reading et al., 1999; Haynes et al., 2005).

On every occasion, traffic accidents investigated at area-based dataset, the studies were based on areas created to serve administrative purposes such as local authorities and wards. These arbitrary administrative areas are very heterogeneous in terms of socioeconomic characteristics. As Cockings and Martin (2005) mentioned, when the research aim is to test a hypothesis or to explore spatial patterns of disease then the maximisation of internal homogeneity is required. To address these problems, a few epidemiological studies have constructed manually social homogeneous geographical areas. For example, Propper et al. (2004) suggested the construction of equally populated neighbourhoods of around 500 people aggregating the census 1991 enumeration districts. Moreover, Reading et al (1999) produced 'social areas' by adjacent enumeration districts with similar Townsend Index values and proportions of the population aged 0-4 years. Both examples were based on visual inspection and the researchers' knowledge of the area. Although the use of manual aggregation is possible in small study areas, the construction of social zones becomes unfeasible as the number of areal units increases.

Furthermore, the selection of number of zones is usually based on the knowledge of researcher of the studied area and is prone to human fault. In the current study, the basic geographies are automatically aggregated by means of the A2Z system constructing geographies with similar Townsend Index scores. The AIC method is used to define the most informative aggregation level combining the variance of traffic accidents and the degrees of freedom of each aggregation. Both the A2Z system and the AIC method provide a variety of automated aggregations using advance statistic measures to explore and select the appropriate aggregation level (Daras 2006). In addition, we examine the significance of children traffic accidents at two aggregation levels: super output areas (SOAs), output zones based on Townsend Index scores. While the comparison of social indicators and the exploration of MAUP at each aggregation level provide valuable information concerning the characteristics of dangerous areas.

3. Datasets and statistical methods

The study area covers the county of Tyne and Wear in England as defined at the time of the 2001 Census. The county of Tyne and Wear has total population of roughly one million people with 237,645 children aged 0-17 years old and it is consisted of five local districts: Newcastle upon Tyne, Gateshead, North and South Tyneside and Sunderland (Figure 1). In the table 1, the population in each local district according the 2001 Census tables are listed providing the number of children aged 0-17 years old.

Table 1: The total population and children aged 0-17 years old in Tyne and Wear.

Authority	Children (aged 0-17 yrs)	Total Population
Newcastle upon Tyne	56,540	259,536
Gateshead	41,599	191,151
North Tyneside	40,998	191,659
South Tyneside	34,672	152,785
Sunderland	63,836	280,807
Tyne and Wear	237,645	1,075,938

Source: (UK National Statistics, Census 2001)

The event data were supplied by the Tyne and Wear Police/Local authority providing information on the exact location where the incident took place and limited information on people's social background or home postcode. The dataset involves children 0-17 years old experiencing a traffic accident in the period 1996-2001 in Tyne and Wear County. In this study, we investigate a five year period instead of single year tackling this way the problem of small numbers.

Altogether 6,111 children were involved in traffic accident during the period between 1996 and 2001, of which 0.6% were fatal incidents, 13.8% were serious accidents with possibly hospital attendance and 85.6% were slight incidents. In the analysis, the event data was the total children incidents without any severity filtering, as the level of analysis was very low and the risk of small number area effect was present. Although, Walsh and Jarvis (1992) have suggested that epidemiological studies of accidents should not use slight severe injuries but those recorded by hospitals, the later datasets are protected by confidentiality regulations with authorisation from the ethic committee of each organisation. As far as this study is concerned, 85.6% of children accidents are slight severe injuries and its possible exclusion limits the dataset to roughly 880 cases in whole Tyne and Wear making the data unsuitable for use. Thus, we included the slight severe injuries into our analysis knowing that the finding will reflect more the less serious than the fatal and severe accidents.

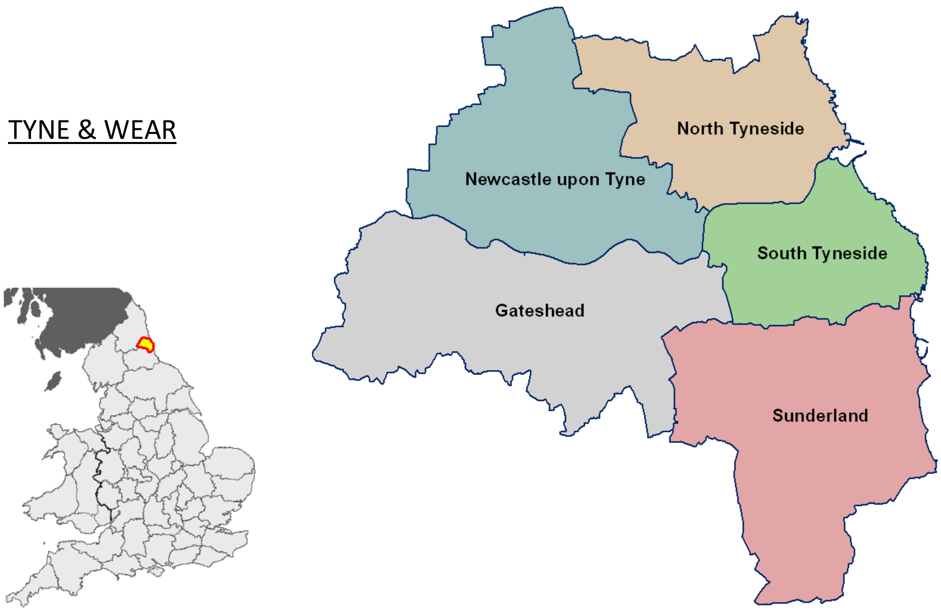

Figure 1: Map of local districts in Tyne and Wear County.

We have used five social determinants to investigate the socioeconomic profile in relation to the traffic accident on children. The social determinants are provided by the 2001 Census datasets (Table 2). The first four indicators: overcrowding, unemployment, owner occupancy and car availability are used to construct the Townsend material deprivation index (Townsend et al., 1988). The Townsend index ranges from negative values, indicating affluent areas to positive values for deprived areas. The strong relation of the Townsend index to the health status of population has been discussed in Morris and Carstairs (1991) as a general measure of deprivation. We selected the 2001 Census determinants because the analysis should be as much coherent as possible with the period of accident events. Therefore, the 1991 census determinants cannot used as they are too dated.

Table 2: The indicators used in the current case study (Census 2001)

Table	Indicators	Numerator Variables	Denominator Variable
KS019	Overcrowding	4	1
KS009	Unemployed	5	2+3+4+5+6
KS018	Owner Occupancy*	2+3	1
KS017	Car availability	2	1
KS020	Lone Parents	11+12	1

* Calculated as follows: 100 – ((KS0180002 + KS0180003/ KS0180001)*100)

Our study area consisted of the 2001 census super output areas (SOAs) whose socioeconomic variables are calculated by summarising the output areas (OAs) information within each SOA. SOAs were chosen as the basic geographical unit because they are one level above output areas (OAs) in the UK for which census data are available. The selection of SOAs instead of OAs was necessary because on each OA there were few accident cases resulting in small number effects. There are 719 SOAs in Tyne and Wear and their population varies between 1,000 and 2,000 people. Descriptive statistics in the SOAs level of Tyne and Wear show an average of 40.8% households not owned by the residents and 41.0% households not owning a car with standard deviations of 24.2 and 17.2 respectively. In addition, an average of 6.0% households is overcrowded reaching a maximum of 33.6% in areas close to Newcastle upon Tyne city centre. Similarly, unemployment in Tyne and Wear scores an average of 8.4% with maximum values of 33.9% next to Tyne River in the Newcastle city centre and in South Tyneside.

Although the population of SOAs is between 1,000 and 2,000 people providing almost equally populated areas, there is an average of 329 children aged 0-17 years old per SOA reaching 726 children in some areas. It is obvious that the children population is not equally distributed within Tyne and Wear County as it varies between 10 and 726 children per SOA. Investigating the cases of children traffic accidents, the distribution of data is ranged from 0 to 49 accidents with additional two outlier SOAs of 72 and 102 accidents. Figure 2 shows the frequencies of children accidents without the two outliers for better graph representation. The distribution of accidents is squished to the left side with a mean value 9 while the standard deviation value is 7.5. Investigating the distribution of accidents, we can identify four groups of SOAs. The two groups ranged from 0 to 9 and from 10 to 17 accidents reflect the majority of SOAs while the last two groups referred to SOAs with many accidents including the outliers. The Figure 3 illustrates the above classification providing information about the SOAs with children traffic accidents in terms of actual numbers. It can be seen that Newcastle's and Gateshead's SOAs provide high numbers of accidents. However, the actual number of accidents can be misleading as two same large numbers of accidents may be referred to different population bases. For example, if there are 30 cases in an urban area with 300 people it will be less serious if those 30 cases recorded in a rural area with 90 people.

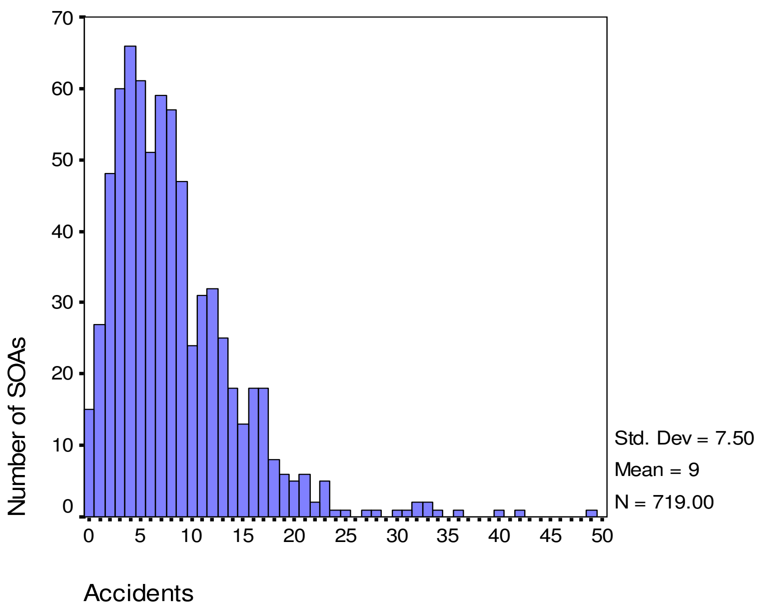

Figure 2: Frequencies of children accidents at SOAs level.

Therefore, we transform the accident cases to accidents rates using the following definition. The children traffic accident rate of an area *z* is defined as:

$$CTAR_z = \sum_{i \in z} y_i \Big/ \sum_{i \in z} B_i \qquad (1)$$

where, y_i is the number of children traffic accidents involving children and B_i is the number of children aged 0-17 years old population for each SOA *i*.

In Figure 4, most of SOAs are low rated while the SOAs at the border of Newcastle and Gateshead score higher with the City of Newcastle to experience the highest rates of 1.06 and 1.14. A careful examination of the study area shows that the specific two areas suffers from many children traffic accidents but the children may not be resident of this problematic area as the data are reporting incident places without any information of home address of child. Therefore, the accident rate indicates how dangerous an area is rather how many children resident in this area involved in traffic accidents.

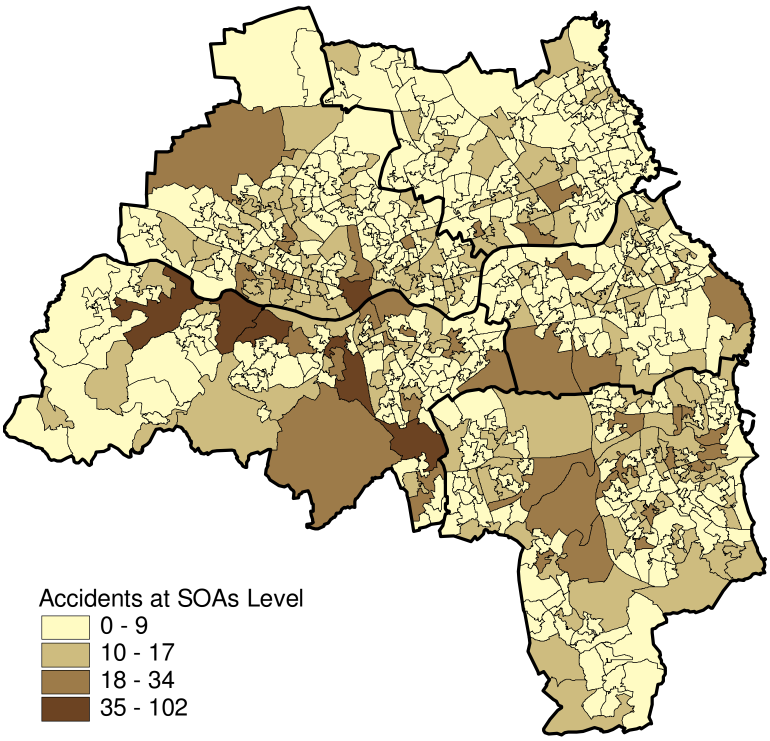

Figure 3: The cases of children accidents at SOAs level.

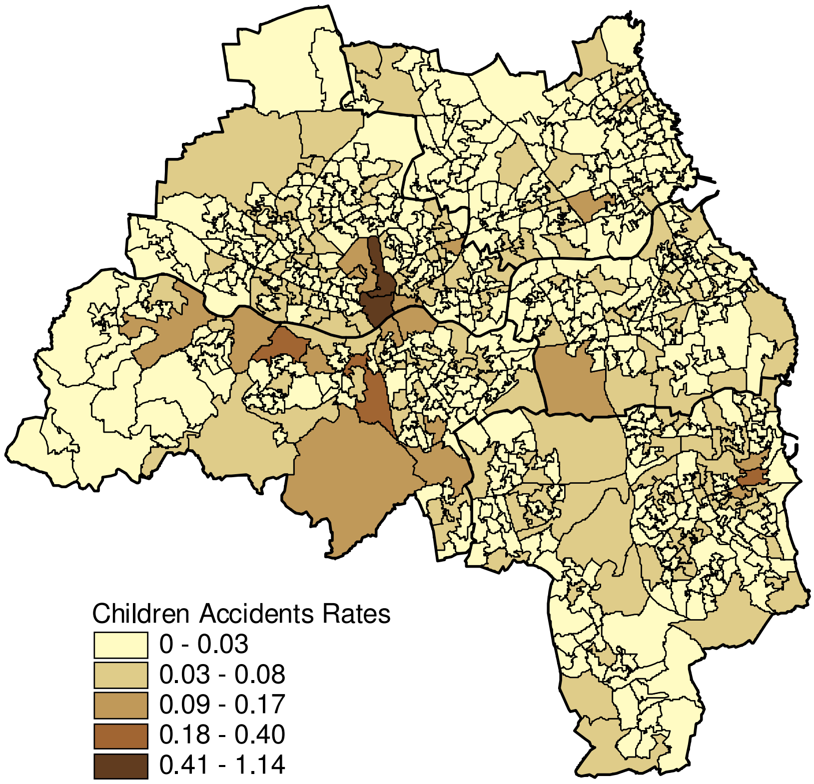

Figure 4: Children traffic accident rates at SOAs level.

The relationship of selected deprivation indicators with the rates of traffic accidents at SOA and ward level are explored presenting the most significant determinants for accidents in Tyne and Wear. Table 3 shows the 'overcrowding households' indicator as the most strongly correlated determinant to children traffic accidents with r = 0.286 and 0.571 at the SOA and ward levels respectively. The unemployment indicator at the SOA level is not significant while the same indicator at ward level becomes significant with r= 0.370. This is an example of the well-known scale effect whereas the correlation coefficient tends to increase as the level of geographical aggregation increases (Gehlke and Biehl, 1934). In addition, a negative correlation appears in the lone parent indicator at SOA level changing to a non-significant determinant at the ward level. Although the literature suggests that lone parent households are a strong determinant of children accidents, in the case of Tyne and Wear it seems that lone parents' households are not correlated with children traffic accidents, at least at the areal level of SOAs and wards.

Table 3: Correlations with children traffic accidents aged 0-17 years old at SOA and Ward level.

	Children Traffic Accidents aged 0-17 yrs old					
	SOA Level			***Ward Level***		
	r	***Sig.***	***N***	***r***	***Sig.***	***N***
No Car	0.155**	0.000	719	0.443**	0.000	113
Overcrowding	0.286**	0.000	719	0.571**	0.000	113
Not owner occupied	0.150**	0.000	719	0.443**	0.000	113
Unemployment	0.066	0.078	719	0.370**	0.000	113
Lone Parent	-0.101**	0.007	719	0.034	0.723	113

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

4. Zone Design characteristics

In this research, the zone design plays a central role as it is used for identifying the most informative aggregation level by means of AIC as well as for constructing homogeneous zones in terms of accident rates and Townsend index. The utilisation of zone design system to investigate the statistical variation of accidents at all possible aggregation levels can be a time consuming process. Therefore, here the method of segments is used providing a quick identification of the suggested aggregation level. In detail, the A2Z system aggregated the 791 SOAs until the zoning solution scores the minimum AIC value (Figure 5). Also, an important asset of zone design, the objective function, set to optimise the zoning systems using the deviance function. For each aggregation process, the zone design executes the optimisation for 10 runs of 300 iterations each or until the objective function does not improved any more.

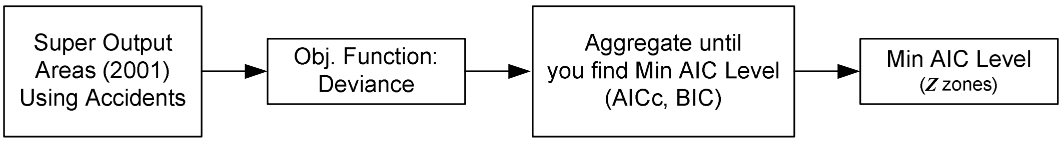

Figure 5: The diagram for identifying the most informative aggregation level.

Then, the number of zones (z) which are statistically appropriate for further investigation according the AIC is used to provide the scale in which an extensive aggregation will take place. As the Figure 6 shows, the areal units of this study (SOAs) grouped twice using the suggested scale level. Both aggregations are based on the k-Homogeneity objective function. In the first grouping, the zone design process targets the maximisation of homogeneity within zones in terms of accident rates, while the second aggregation aims to build homogeneous zones using the Townsend Index. The grouping process for both aggregations is repeated for 100 runs of 300 iterations each investigating an extended range of grouping solutions at the same scale.

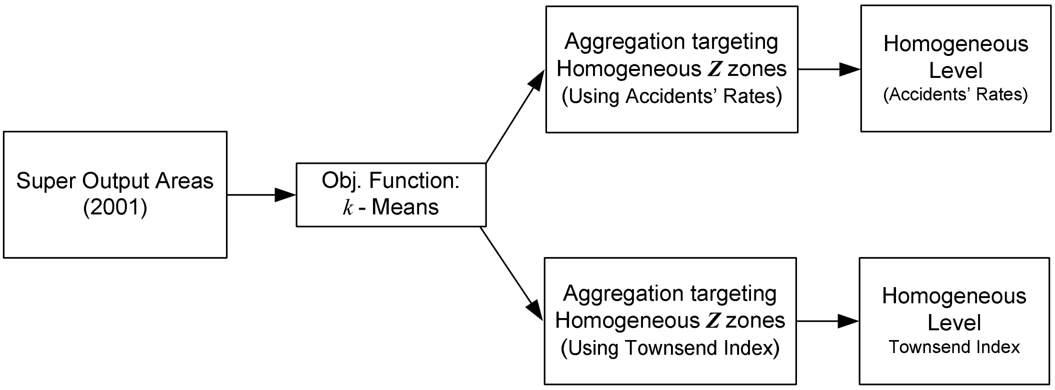

Figure 6: The diagram of aggregation for providing homogeneous zones at predefined scale level.

5. Selecting the most informative aggregation level

As this research strongly supports the idea of constructing geographies with criteria focused on health related characteristics rather using administrative areas constructed to tackle other social issues, it is important to investigate the appropriate aggregation level of this case study, using advanced statistical measures of goodness of fit at each scale. The methodology of selecting the most statistically informative aggregation level by means of AIC criterion as discussed in Chapter 4, is extensively applied in this case study. Every aggregation level is constructed by the A2Z system using the SOAs as basic units and optimising the homogeneity of children traffic accident rates.

The objective function for this case study is the *deviance* or *likelihood ratio* statistic measure. As the deviance measures the variance within the zones, a minimisation of deviance is equivalent to homogeneous constructed zones. The deviance of an aggregation *A* is calculated using the equation 2.

$$D_A = -2(l_A - l_0) \qquad (2)$$

where, l_A is the log-likelihood of aggregation *A* and l_0 is the log-likelihood of the aggregation with zero degrees of freedom. For this case study the aggregation level 0 is equivalent to the SOA level. The calculation of log-likelihood at aggregation A is defined as:

$$l_A = \sum_i (-E_i + O_i \ln E_i) \qquad (3)$$

where, E_i is the estimated number of children traffic accidents in *i* SOA and it is calculated as follows:

$$E_i = CTAR_z \cdot B_i \qquad (4)$$

To calculate the log-likelihood of SOA level (l_0), equation 3 is converted by replacing the estimated children traffic accidents (E_i) with the observed values (O_i):

$$l_0 = \sum_i (-O_i + O_i \ln O_i) \qquad (5)$$

By minimising the deviance at a selected aggregation level, the zone design system produces similar children accident values within each zone. The CC method of shape constraint is used providing a slight compactness rule that prevents extremely complicate shapes. The A2Z system randomly generates *z* zones every time one of the 20 runs starts and improves the aggregation output for 300 iterations per zone.

At first stage, the SOAs were aggregated to a sequence of equally divided intervals of 2, 72, 144, 216, 288, 360, 500 and 719 zones and we measured the chi-square, deviance, Akaike Information Criterion Corrected (AICc), AIC and Bayesian Information Criterion(BIC) at each aggregation level. As the target of this method is to identify the minimum information criterion, we should select one of the three available criteria to use for further exploration. Hurvich et al. (1998) suggested the corrected AIC (*AICc*) is less biased than classical AIC, as the advantage of AICc is concentrated in the parameters calculation. In Table 4, the AICc suggests the need of closer investigation at the segments between the 113 and 216 zones. At the second stage the A2Z system aggregated 123,133,153,163,173,183 and 193 zones repeating the calculations of goodness of fit measures.

In Figure 7, the values of three criteria (AIC, AICc and BIC) are graphed suggesting the aggregation level of 183 zones as the most appropriate and informative level as all of them reach their minimum value. In depth aggregation at the segment between 173 and 193 could provide the optimum number of zones, but as it is mentioned at previous chapters the zone design system explores a number of possible solutions optimising certain criteria. This nature of zone design is responsible for the deviation between the informative measures and the theoretical curve. In this case study, the 183 zones are representing the aggregation level in which the significance of selected social indicators is evaluated. Additionally, the AICc measure seems to increase the differences between aggregated levels as the parameters are calculated differently in each criterion. In practice, Figure 7 illustrates the performance of the three criteria showing a rapid increase of AICc values after the 183 zones in comparison to AIC and BIC criteria.

At the bottom of Table 4, the measures of goodness of fit at ward level imply a high variation of children accident values in comparison with the 113 aggregated zones of A2Z system. Both chi-square (4318.41) and deviance (3000.92) measures show that the ward level is statistically equivalent to the homogeneous zones between 2 and 72 aggregation levels. As it is obvious, the aggregation levels of 2 to 72 zones are high scaled geographies with strong social heterogeneity within and between zones. Comparing the ward level with 113 homogeneous zones, the table illustrates strong differences of variance supporting the comment that the ward level may not be suitable for analysis of health related phenomena such as the children traffic accidents. On the other hand, the suggested level of 183 zones scores far better at AIC, AICc and BIC criteria with -16438.85, -16312.97 and -16282.07 values, while at ward level the values are -14174.27, -14131.68 and -14077.46 respectively. Worth noted here is the SOAs level where the degree of freedom is zero representing the full model (l_0) as mentioned earlier in this section. The AIC, AICc, and BIC values are extremely high as the parameters of the full model reach the maximum value (719). Consequently, the two issues of selecting the appropriate aggregation level and constructing homogeneous zones based on social determinants have been addressed here using the statistical information criteria in a zone design context.

Table 4: Measuring the goodness of fit at various automated aggregation levels.

Number of Zones	Chi-Square	Deviance	df	AIC	AICC	BIC	1st Stage	2nd Stage
2	8462.93	3511.96	717	-13885.22	-13885.21	-13883.51	■	
72	1402.95	1236.16	647	-16021.02	-16004.75	-15959.34	■	
113	1175.32	1066.22	606	-16108.96	-16066.37	-16012.15	■	
123	1026.89	966.19	596	-16188.99	-16137.72	-16083.61		■
133	1010.13	869.79	586	-16265.39	-16204.46	-16151.45		■
144	855.06	786.20	575	-16326.98	-16254.23	-16203.62	■	
153	770.58	729.95	566	-16365.24	-16281.83	-16234.16		■
163	775.16	718.18	556	-16357.00	-16260.67	-16217.35		■
173	664.01	647.14	546	-16408.04	-16297.57	-16259.83		■
183	**619.36**	**596.33**	**536**	**-16438.85**	**-16312.97**	**-16282.07**		■
193	606.79	588.15	526	-16427.04	-16284.40	-16261.69		■
216	611.03	571.05	503	-16398.13	-16211.39	-16213.08	■	
288	538.91	519.56	421	-16305.62	-15918.50	-16058.89	■	
360	482.16	474.10	359	-16207.08	-15481.05	-15898.66	■	
500	438.95	438.99	219	-15962.19	-13664.03	-15533.83	■	
719	0.00	0.00	0	*FM*	*FM*	*FM*	■	
Ward Level (113)	4318.41	3000.92	606	-14174.27	-14131.68	-14077.46		

FM is the full model where the number of zones is same to areal units.

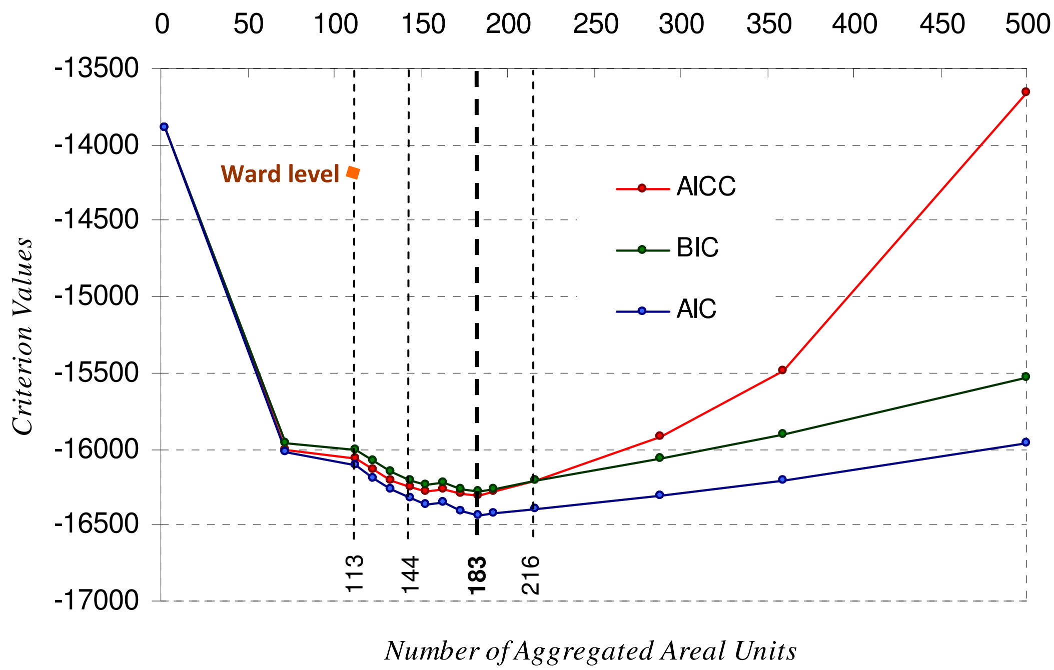

Figure 7: The performance of AICc, AIC and BIC measures as the number of zones increases

6. Results

From the above statistical measures, the level of 183 zones has been selected for further analysis involving the construction of 183 homogeneous zones using the Accident rates (Z_{ACC}) and Townsend Index scores (Z_{TI}). In Figure 8, the 183 homogenous zones constructed by Children Traffic Accidents rates (Z_{ACC}) are showing an increase of traffic accidents in areas with high density road network. As expected, the density of road network is determined by the population density and it is obvious that urban areas such as the Newcastle City are experiencing more traffic accidents than rural areas. However, there are two zones in Gateshead and one in Sunderland with high rates between 0.12 and 0.23. Concentrating in the Newcastle area, there is one zone with 1.08 accident rate. At first glance, it is seems to be a mistake as there are more incidents than the children population. A careful examination of the study area shows that the specific area suffers from many children traffic accidents but the children may not be resident of this problematic area as the data are reporting incident places without any information of home address of child. However, an explanation for the improper value of the rate could be the accident record file. This case study is based on the five years accident record and cases of children having more than one accident are possible (especially for slight severe accident). Therefore, the accident rate indicates how dangerous an area is rather how many children, residents of this area, involved in traffic accidents.

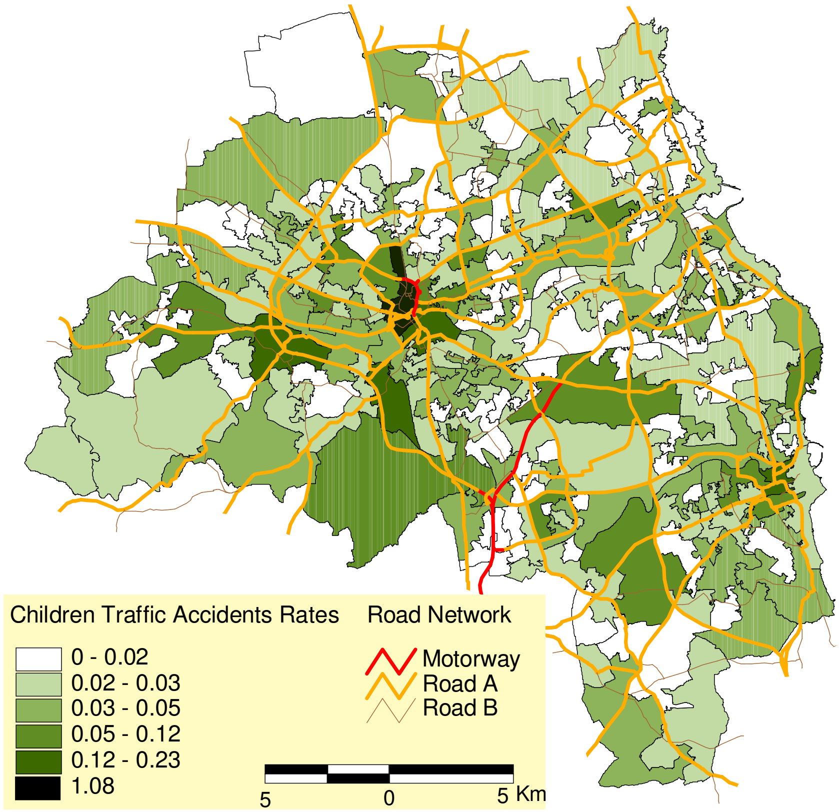

Figure 8: Map of Children Traffic Accidents in percentages using 183 homogenous zones by accident rates (Z_{ACC}) in Tyne and Wear.

The maps in Figure 9 show the Townsend Index scores in Tyne and Wear at SOA level. Also, the boundaries of the 183 homogeneous zones by children traffic accident rates are presented. Some direct findings show high deprivation values in Newcastle area especially at the areas adjacent to Tyne River and in Sunderland area. As expected, the 183 homogeneous zones by Townsend Index scores (Z_{TI}) is capturing the variations of deprived areas better than Z_{ACC} zoning solution. However, Z_{ACC} output zoning illustrates an almost equally homogenous aggregation level suggesting strong relationships between the children accidents and the deprived areas.

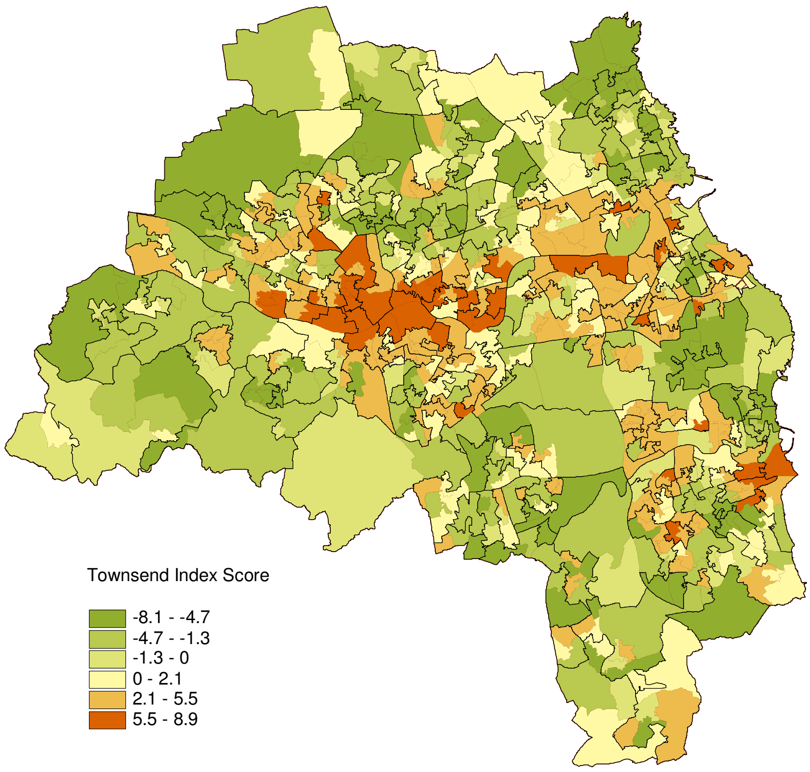

Figure 9: Map of Townsend Index scores at SOA level with 183 zones constructed by similar Children Traffic Accident rates.

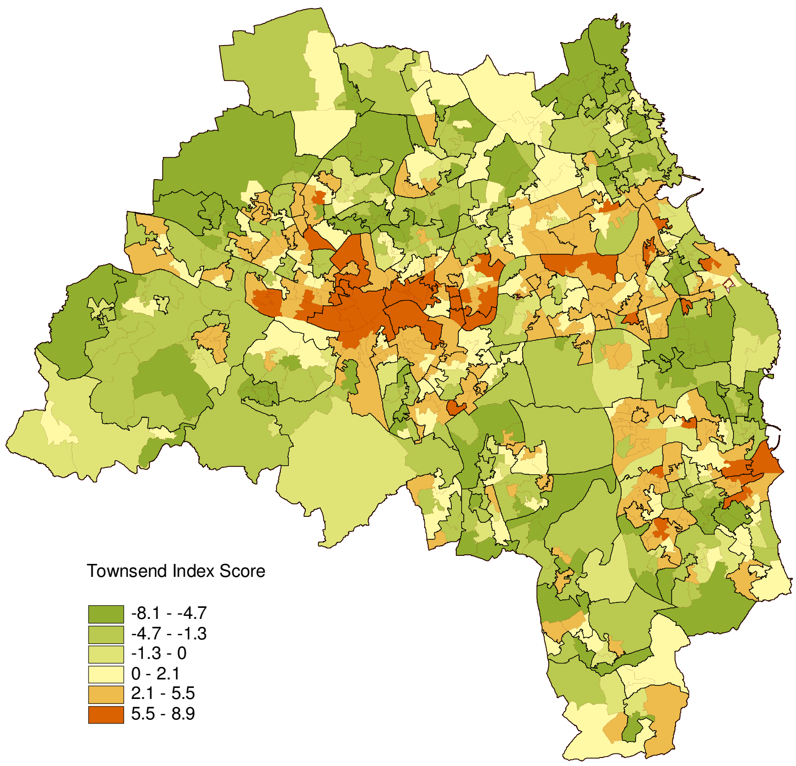

Figure 10: Map of Townsend Index scores at SOA level with 183 zones constructed by similar Townsend Index scores.

To statistically explore if both aggregations Z_{ACC} and Z_{TI} are strongly related the type of relationship between children traffic accident rates, indicators of material deprivation and Lone parent household determinant is investigated using exploratory analysis. In Table 4, the visual findings from the maps are statistically confirmed at both aggregated zones as the significance of children accidents with the determinants result in similar correlation coefficients. The 'overcrowding' determinant remained is now more strongly correlated at the Z_{TI} than Z_{ACC} zones with correlations of 0.474 and 0.425 respectively. The 'no car' and 'not owner occupied' determinants are significant at the 0.01 level but they are not as strongly related with children accidents as the 'overcrowding' indicator. Furthermore, the 'unemployment' determinant is not significant at any aggregation while the 'lone parent' indicator gives negative correlation suggesting that the single parent households are not related to children accidents. According to Table 5, the Z_{TI} zones provide better geographies than Z_{ACC} zones as there was noticeable improvement of correlations between children accidents and determinants, especially for the significant determinants at the 0.01 level.

Table 5: Correlations with children traffic accidents aged 0-17 years old at 183 zones level.

	Children Traffic Accidents aged 0-17 yrs old					
	183 zones (Z_{ACC})			*183 zones (Z_{TI})*		
	r	*Sig.*	*N*	*r*	*Sig.*	*N*
No Car	0.202**	0.006	183	0.234**	0.001	183
Overcrowding	0.425**	0.000	183	0.474**	0.000	183
Not owner occupied	0.209**	0.005	183	0.239**	0.001	183
Unemployment	0.081	0.275	183	0.098	0.186	183
Lone Parent	-0.175*	0.018	183	-0.085	0.252	183

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

For both suggested zoning solutions, their correlations are weaker than ward level as listed in Table 5. This is result of the increase from 113 wards to new 183 zones. As Gehlke and Biehl (1934) mentioned the correlation coefficient tends to increase as the level of geographical aggregation increases, therefore the differences between the correlation at wards and suggested zones levels reflect the 70 additional zones in the Z_{TI} and Z_{ACC} zoning solutions. The Z_{TI} zones obtain more unbiased estimates of correlation because the aggregation process was based on explanatory variables and as Blalock (1964) mentioned the aggregation of target variables often causes severe biases of correlations.

Furthermore, the Figure 11 illustrates the children traffic accident rates at SOA level with the boundaries of 113 wards in Tyne and Wear. It can be seen that the wards are very heterogeneous in terms of accident rates and this is reflected in the measures of variance discussed earlier. In addition, local clusters of SOAs with high accident rates are separated by ward boundaries resulting in heterogeneous wards. Therefore, we strongly encourage the use of informative statistics for investigating the scale level, while the zone design system provides homogeneous zones with specific characteristics.

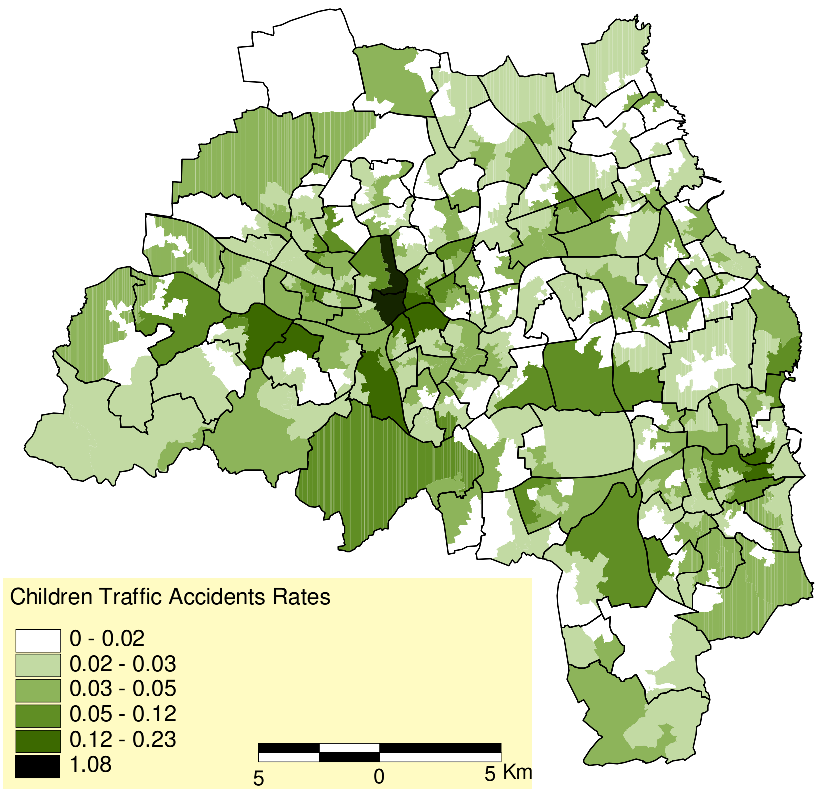

Figure 11: Map of children traffic accidents rates at SOA level with the boundaries of 113 wards in Tyne and Wear.

7. Discussion and Conclusions

We have shown an increase of correlation coefficient as the scale changes from SOA to ward in the area of Tyne and Wear. According to Gehlke and Biehl (1934) the correlation coefficient increases as the aggregation level increases and this is clearly presented in Table 3. Despite the changes of scale and aggregation, overcrowding, households without car and owner occupancy determinants are all significant at the 0.01 level in relation to children accidents. An interesting finding is the negative correlation of the lone parent household determinant at Z_{ACC} level, while in ward and Z_{TI} level it becomes insignificant. This findings is in contrast with the relevant literature, but it reflects the nature of data that has been used in this study as most of the literature studies use hospital admission data (Reading et al. 1999) or longitudinal datasets (Daras et al., 2005; Haynes et al., 2005) at the household level. Hospital admission data usually consists of information about the home address of children and the severity of the incident, but does not provide the location of an accident (Stewart-Brown et al., 1986). This is a limitation of the available datasets in the UK, as there is linkage between incident and hospital admission data.

In this case study, a methodology for selecting the appropriate number of zones was proposed using advanced statistical measures such as AIC, AICc and BIC (Daras 2006). In detail, we suggested the minimisation of AICc or similar criterion targeting the appropriate aggregation level (scale effect) and the Deviance or chi-square goodness of fit measures to explore the quality of different aggregations at the same level (aggregation effect). Applying the methodology in Tyne and Wear, we concluded that the most statistically informative level is the one consisting of 183 zones. The construction of homogeneous zones by Townsend Index scores and its comparison with the homogeneous zones by Children Traffic Accident rates exhibit significantly strong relationship of accidents with deprived areas.

References

Abdalla, I. M. 1997, *Statistical investigation and modelling of the relationships between road accidents and social characteristics,* PhD Thesis, Napier University.

Abdalla, I. M., Barker, D. and Raeside, R. 1997a, "Road accident characteristics and socioeconomic deprivation", *Traffic Engineering and Control*, **38**, pp. 672-676.

Abdalla, I. M., Raeside, R., Barker, D. and McGuigan, D. R. D. 1997b, "An investigation into the relationships between area social characteristics and road accident casualties", *Accident Analysis and Prevention*, **29**, 5, pp. 583-593.

Alwash, R. and McCarthy, M. 1988, "Accidents in the home among children under 5: ethnic differences or social disadvantage?", *British Medical Journal*, **296**, pp. 1450-1453.

Christie, N. 1995, *Social, economic and environmental factors in child pedestrian accidents: A research review*, Transport Research Laboratory, Wokingham, Project Report 116.

Cockings, S. and Martin, D. 2005, "Zone design for environment and health studies using pre-aggregated data", *Social Science & Medicine*, **60**, 12, pp. 2729-2742.

Daras, K. 2006, "An information statistics approach to zone design in the geography of health outcomes and provision", *In Ph.D. Thesis*, University of Newcastle upon Tyne.

Daras, K., Haynes, R., Jones, A. and Reading, R. 2005, "Neighbourhood zone design for a study of injuries to preschool children", RGS-IBG Annual International Conference, London, p. 35.

DfT 2003, *Child Road Safety: Archieving the 2010 Target*, Department of Transport, London.

Gehlke, C. E. and Biehl, K. 1934, "Certain effects of grouping upon the size of the correlation coefficient in census tract material", *Journal of the American Statistical Association*, **supplement 29**, pp. 169-170.

Haynes, R., Daras, K., Reading, R. and Jones, A. 2007, "Modifiable neighbourhood units, zone design and residents' perceptions", *Health & Place,***13**, 4, pp. 812-825.

Hurvich, C. M., Simonoff, J. S. and Tsai, C. L. 1998, "Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion", *Journal of the Royal Statistical Society, Series B*, **60**, pp. 271-293.

Jarvis, S. 1999, "Why do we need action on injury in the UK?", *Public Health Forum*, **3**, pp. 1-5.

Judge, K. and Benzeval, M. 1993, "Health inequalities: new concerns about the children of single mothers", *British Medical Journal*, **306**, pp. 677-680.

Morris, R. and Carstairs, V. 1991, "Which deprivation? A comparison of selected deprivation indexes", *Journal of Public Health Medicine*, **13**, pp. 318-325.

Propper, C., Jones, K., Bolster, A., Burgess, S., Johnston, R. and Sarker, R. 2004, *Local neighbourhood and mental health: evidence from the UK* Manchester, ESRC Ref: H333250042.

Reading, R., Langford, I. H., Haynes, R. and Lovett, A. 1999, "Accidents to preschool children: comparing family and neighbourhood risk factors", *Social Science & Medicine*, **48**, pp. 321-330.

Roberts, I. and Pless, B. 1995, "Social policy as a cause of childhood accidents: the children of lone mothers", *British Medical Journal*, **311**, pp. 925-928.

Roberts, I. and Power, C. 1996, "Does the decline in child injury mortality vary by social class? A comparison of class specific mortality in 1981 and 1991", *British Medical Journal*, **313**, pp. 784-786.

Runyan, C. W., Bangdiwala, S. I., Linzer, M. A., Sacks, J. J. and Butts, J. 1992, "Risk factors for fatal residential fires", *New England Journal of Medicine*, **327**, pp. 859-863.

Sharples, P. M., Storey, A., Aynsley-Green, A. and Eyre, J. A. 1990, "Causes of fatal childhood accidents involving head injury in Northern region, 1979-86", *British Medical Journal*, **301**, pp. 1193-1197.

Stewart-Brown, S., Peters, T. J., Golding, J. and Bijur, P. 1986, "Case definition in childhood accident studies: a vital factor in determining results", *International Journal of Epidemiology*, **15**, pp. 352-359.

Towner, E., Dowswell, T. and Jarvis, S. 1993, "The Effectiveness of Health Promotion Interventions: a Literature Review," in *Reducing Childhood Accidents*, Health Education Authority, London.

Townsend, P., Phillimore, P. and Beattie, A. 1988, *Health and deprivation: inequality and the North* Croom Helm, London.

Walsh, S. and Jarvis, S. 1992, "Measuring the frequency of "severe" accidental injury in childhood", *Journal of Epidemiology and Community Health*, **46**, pp. 26-32.