Safety in Numbers for Cyclists in England: Measuring the Effect

George Ursachi and Richard Owen

Road Safety Analysis Limited 2016

Safety in Numbers for Cyclists in England: Measuring the Effect

What is Safety in Numbers (SIN) and why does it happen?

The theory is that in a mixed traffic environment; with cars, motorbikes, other motor vehicles, pedestrians and cyclists all sharing the road, the balance of road users can affect relative injury risk to individual groups. To put it simply, more cyclists on the road can equal a lower casualty rate.

Safety in Numbers, as a phenomenon in cycling, was first identified in 2003 in an academic paper by public health researcher Peter Jacobsen, 'Safety in numbers: more walkers and bicyclists, safer walking and bicycling'[1]. He summarised the findings saying, 'More riders, fewer crashes; fewer riders, more crashes'. Jacobsen thought that 'adaptation in motorists' behaviour' was the most plausible explanation. For example, when there are a lot of cyclists on the road, drivers take more notice of them and adapt their behaviour accordingly. This places the assumption that drivers of motorised vehicles are 'to blame' for injuries to cyclists; a matter not covered in this report.

CyclingUK[2] says that besides the fact that drivers become more aware of cyclists, there are two other possible reasons for SIN: Firstly, that drivers are more likely to be cyclists themselves and are therefore more sympathetic, and secondly that there is greater political will to improve cycling conditions.

Grist.org[3] has also cited Jacobsen's theory: "The bigger SIN story is that those cities /countries that have encouraged bicycling have been rewarded with more trips by bike, and not just a non-linear increase in injuries, but a decrease in injuries."

One question that regularly gets asked in relation to SIN is, 'how many cyclists do we need to achieve a reduction in risk?', a question that is often unanswered with any certainty. This is because isolating the effect of a single variable, in this case more cyclists, from other trends and features over a long period of time is incredibly difficult. Furthermore, comparing different nations with different patterns of use and different standard of road network is also fraught with danger.

Existing evidence

There are many studies, facts and figures that seem to support the concept of SIN. Impressive figures from Copenhagen between 1995 and 2006 (where cycling increased by 44%), show a 60% drop in the number of cyclists killed or seriously injured. Similarly, in The Netherlands between 1980 and 2005, where cycling increased by 45%, cycling fatalities decreased by 58%. These results are notable because they show a real fall in the number of cyclists killed and injured, not just the rate.

CyclingUK state, in their Safety in Numbers report[4], 'research suggests that a doubling of cycling would lead to a reduction in the risk of cycling by around a third'. The same report also presents a

[1] Safety in numbers: more walkers and bicyclists, safer walking and bicycling, P.L. Jacobsen, Injury Prevention, 2003, Issue 9, pages 205-209

[2] http://www.cyclinguk.org/campaign/safety-in-numbers

[3] http://grist.org/cities/2010-10-11-theres-safety-in-numbers-for-cyclists/

[4] http://www.cyclinguk.org/campaign/safety-in-numbers

chart (Figure 1) with cycling levels and cyclists killed across EU countries. It clearly indicates that countries with high cycling levels pose a lower risk to cyclists.

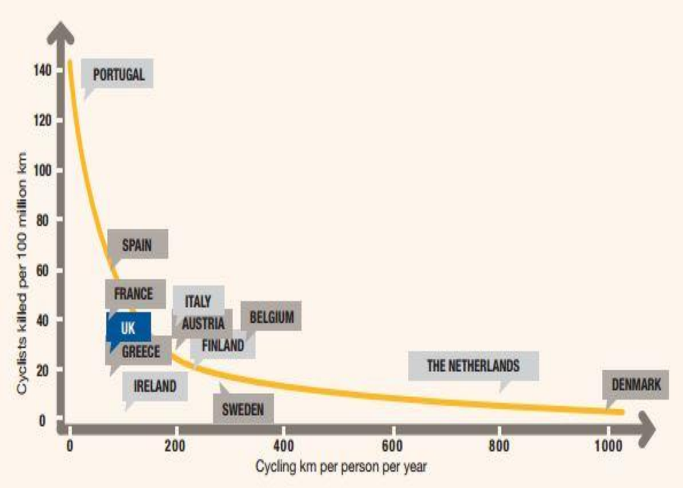

Figure 1. EU countries cycling levels and cycling risk

The issue of comparing different road networks with different traffic types still applies to this data. The question could be posed, 'If we had Danish levels of cyclists on our roads, what would happen?'

In his research, Jacobsen used 3 population level and 2 time-series data sets with a final output showing the relationship between activity and injury (Figure 2). He reached the conclusion that, 'Policies that increase the number of people walking and bicycling appear to be an effective route to improving the safety of people walking and bicycling'.

Other research from the USA looks at the link between activity and injuries. The City of Portland Oregon regularly publish results of their traffic counts and in a 2009 report[5] an impressive trend line for indexed bicycle crash rates can be seen. Although the rate is measured using counts across specific cycling bridges, rather than on the entire network, it remains a worthwhile study.

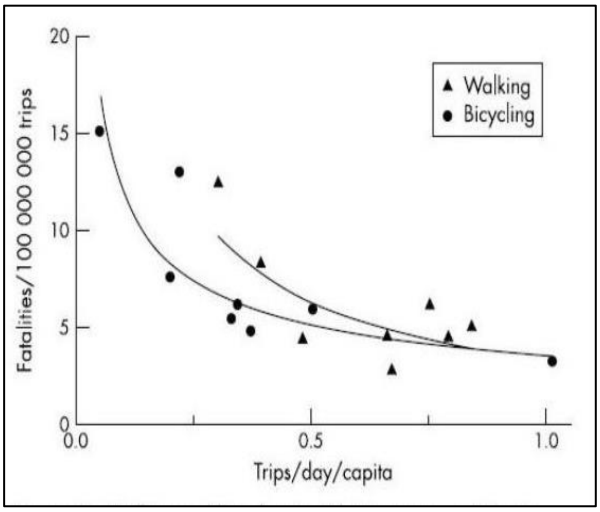

Figure 2. Walking and bicycling in eight European countries in 1998 – From Jacobsen 2003[1]

The same phenomenon seems to be happening in New York City, where an impressive decrease in annual casualties negatively correlated with an increase in ridership (whilst ridership has doubled, casualties have halved)[6].

An increasing amount of research and evidence suggests that SIN can be evidenced, although there are differing methodologies used, and therefore comparisons between studies is difficult. Furthermore, obtaining a 'magic figure', which would predict what happens when cyclists increase by a fixed percentage, is very difficult.

[5] https://www.portlandoregon.gov/transportation/44671
[6] http://www.nyc.gov/html/dot/html/bicyclists/bikestats.shtml

Our Study

Previous analysis of cycling risk can be seen on the PACTS Constituency Dashboard[7]. The dashboard displays cycling risk to constituency residents against local population rates to create an index value allowing comparison between areas. Whilst a residency approach is better than measuring casualties injured in an area and then comparing them to the populations of an area (which results in a mismatched numerator and denominator), it doesn't consider the amount of cycle traffic in the area.

A clear example of this is the casualty figures for Cambridge, which reveal an index of 504 (4 times the national average) when based on population. However, it is well-known that cycling levels in Cambridge are much higher than the national average and therefore exposure needs to be taken into account.

In order to undertake a more robust analysis and compare risk around the country based on the amount of cycling, it is necessary to find a more appropriate denominator.

The Data

For this preliminary analysis, the following data sets were used:

- The last 5 years' (2010 – 2014) average adult cyclists' casualties, based on residence, from MAST Online. Residency is calculated by using the postcode of the casualty.
- The Proportion of residents who cycle (any length) for utility purposes at a given frequency in England, 2013 to 2014. Link: https://www.gov.uk/government/statistical-data-sets/cw030-proportion-of-residents-walking-or-cycling-for-utility-purposes (Table CW0321);
- Population number for each local authority district (LAD), from the GB statistics, for adults (16+);

Here the 'utility cycling percentage' value is used as the denominator to assess cycling risk based on exposure. It is worth noting at this point that it is by no means a perfect metric. In order to assess risk more accurately it would be preferable to know the distance cycled on shared networks per month, but this is not recorded. Therefore, for the purposes of this analysis, it will have to assumed that cycling around the country is a similar mix of on and off road.

Preliminary analysis

Using the 3 sets of data, it is possiblrto measure cycling exposure for each LAD as a product of the proportion from the exposure data, and the population from the statistics. This allowed to the following to be produced:

1. Cycling exposure, expressed as the number of cyclists per 1,000 population;
2. Cycling risk as the proportion of casualties based on cycling exposure;
3. Cycling risk per 1,000 cyclists;
4. Descriptive statistics for both cycling levels per 1,000 population and cycling risk per 1,000 cyclists;
5. Groupings of LADs by high and low cycling levels per 1,000 population (above and below the mean);
6. Groupings of LADs by cycling risk per 1,000 cyclists (above and below the mean).
7. 4-way matrix of LADs by the grouping in (5) and (6). These groups were compared and the risk trend analysed for each;

[7] http://www.pacts.org.uk/dashboard

8. A linear regression for the 4 groupings. This assumes all the other factors remain constant and are incorporated in the main constant of the function.

Results

NOTE: The analysis described here is preliminary, not due to incompletion but because there are other considerations that have not been included which would fully explore the relationship between the number of cyclists and risk of injury. The sample sizes included within this analysis are more than sufficient for the results to stand on their own, but there is also potential for a wider debate into the topic.

The preliminary findings look very encouraging and are in line with the expectations based on previous evidence. They also provide a good starting point for further development of the analysis. There is a clear relationship between risk and cycling levels with the distribution of points in Figure 3 being similar to those noted in previous studies and referenced in Figures 1 & 2.

Within the chart, the separate linear regressions for each of the following 4 categories of LAD are displayed:

1. High Risk – High Level   Blue
2. High Risk – Low Level   Red
3. Low Risk – High Level   Orange
4. Low Risk – Low Level   Yellow

For each category, it has been assumed that all other factors remain constant. Simple linear regression functions were then calculated to determine whether the influence was statistically strong enough.

Group	Linear Regression	Statistically significant
HL - HR	12.592 – 0.22*Cycling Level (CL)	No
HL - LR	5.289 – 0.03*CL	No
LL - HR	31.099 – 0.759*CL	Yes
LL - LR	9.210 – 0.097*CL	Yes
Total Population	12.331 – 0.73*CL	Yes

In all cases, the relationship between cycling levels and cycling risk is negative (meaning they influence each other in opposite direction) and in 3 tests the results are statistically significant at the 95% confidence interval.

For more accurate and powerful prediction functions, the number of factors analysed should be increased and a time series method used. More details about function and statistical significance can be found in Appendix 1.

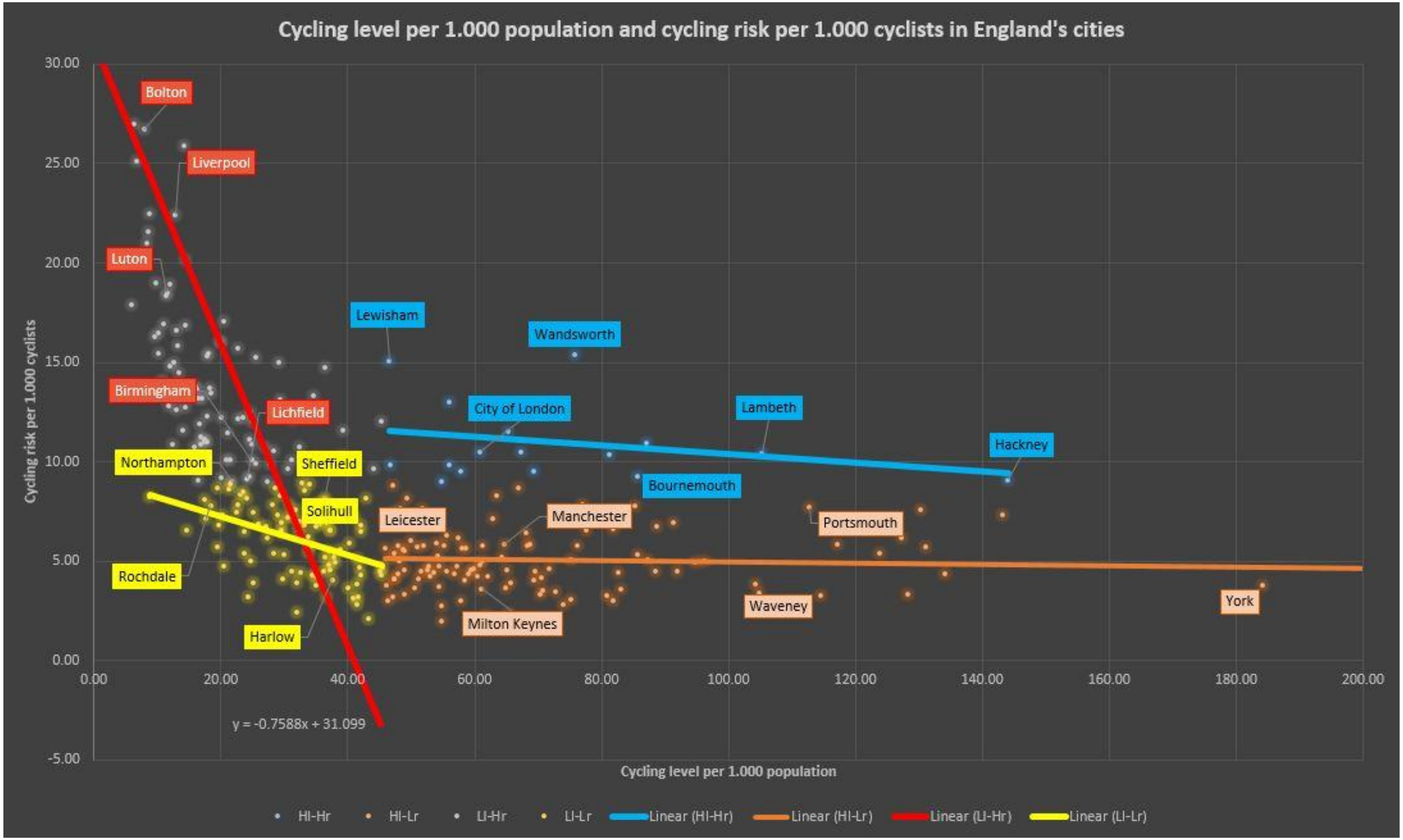

Figure 3. Cycling level and cycling risk for England's cities with examples

What does this mean?

Due to the existing levels of risk in English towns and cities it is not simple to produce a model that says if you double the number of cyclists, absolute risk will change by X%. In towns and cities of high risk and low levels, the potential for risk reductions is greatest; although the absolute risk to 'new' cyclists will be much higher than in areas of low risk and high rates.

For areas with existing high cycling levels, the effect is moderate but still shows a slight negative correlation (suggesting that further increasing cycling levels will have an additional benefit to risk rates). If these findings are correlated with the other health and economic benefits cycling creates or enables, then it could be concluded that improving cycling level is beneficial to all cities, at all levels.

Weaknesses with the methodology

There are many things to consider when comparing different parts of England. The cycling infrastructure and environment varies greatly, especially between rural and urban areas. There could even be arguments for different cultures, both of cyclists and drivers in different parts of the country, which would lead to different attitudes for cyclist safety. This applies to an even greater degree with international studies. Segmenting the LADs in England by rurality, traffic density or some other characteristic may enhance future analyses. Having time series data would help understand how each area reacts to changes in cycling levels as well as in cycling infrastructure.

If comparable international data was available at the city level, it would enable other meaningful comparisons to be made.

When comparing England's LADs, the following possible differences and their effects should be investigated:

- Road or lane width and whether or not cyclists are traveling on those lanes or have separate lanes
- Speed limits
- Visibility – especially on country or urban roads
- Road safety culture and attitudes to cyclists
- Length of segregated cycle paths, and on-road cycle lanes per 1,000 km of road or per 1,000 km of cycling trips
- Highway condition
- Segregation of cycle lanes/paths

Next Steps

It is already possible to re-run the analysis using the most recent data and to construct a time-series. Grouping LADs according to a similarity criteria would allow other factors to be reviewed and would provide a greater understanding of local networks.

The creation of an online tool to assist planners and the public understand local risk and safety benefits would also be beneficial. Such a tool would show what may happen if cycling numbers increased and would explore in more detail the demographics of those who are collision-involved. Highlighting the relative level of safety in many parts of the country could lead to a reduction in the fear of cycling, but it would also highlight the current problems cyclists in some areas face.

Ensuring that local policy makers fully understand the relationship between exposure and casualty rates is essential as it will encourage an investment in the promotion of cycling as well as the implementation of safety measures.

How to improve absolute and relative safety for cyclists

Setting simultaneous safety and modal shift objectives without incorporating clear safety improvements will almost certainly result in more casualties. This is because the number of casualties is a function of the rate of collisions as the number of people cycling or as the distance of kilometre cycled. Increasing exposure will only result in a higher number of casualties and is demonstrated in the model below (Figure 4). Therefore objectives should initially be set for reducing the collision occurrence rate by introducing other measures known to reduce injuries. The two most effective strategies are stopping collisions occurring altogether by separating modes, or reducing the chance of injury should a collision occur.

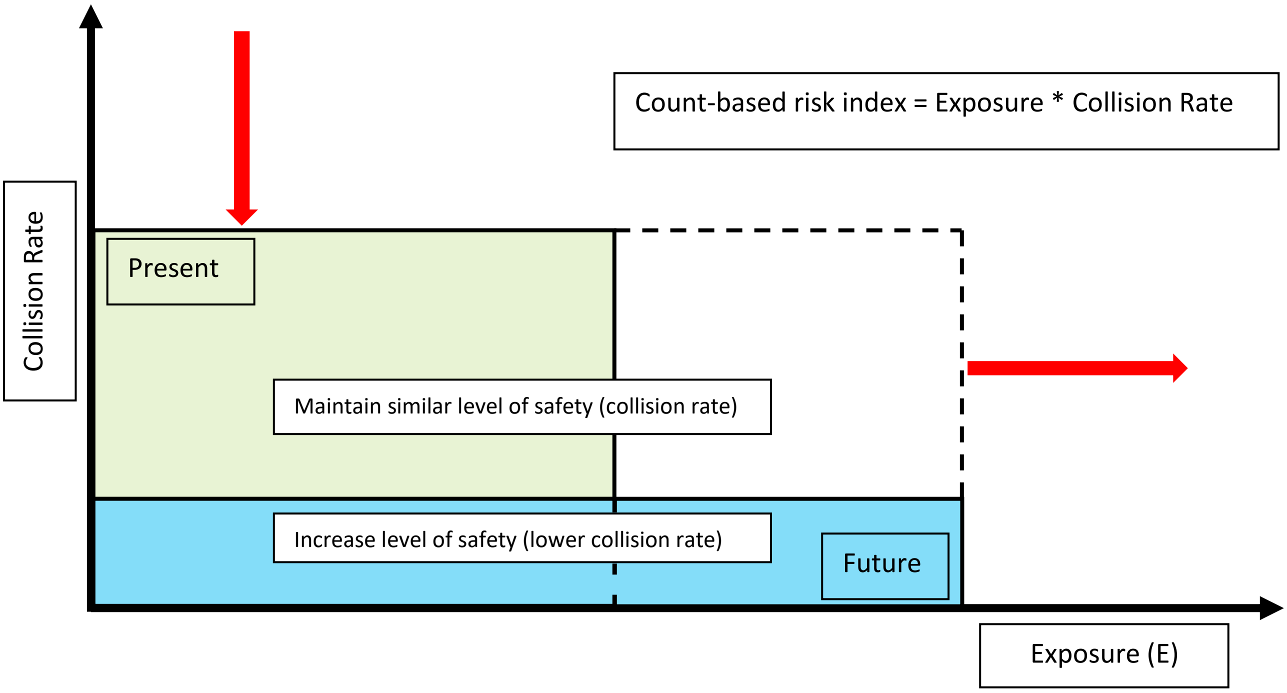

Fig 4. The paradox of trying to decrease count-based risk index while increasing exposure

When there is already a good cycling infrastructure and a reasonably high cycling level, increasing the cycling level is likely to produce further decreases in cycling casualties, but until then, campaigns should focus mainly in decreasing the casualty rate and improving the safety and the awareness of cycling and cycling related issues among cyclists and other traffic participants.

Appendix 1. Regression functions and groups information

Total population

Size: 319 cities

Function: **Cycling Risk = 12.331 – 0.73*Cycling Level**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.331	1.009		12.217	.000
	Level of cycling (cyclists over 1000 persons)	-.073	.017	-.236	-4.319	.000

a. Dependent Variable: Risk per 1000 cyclists

Hl-Hr group

Size: 17 cities with Cycling Level above the national mean (the mean of the 319 cases above) and with Cycling Risk above the national mean

Function: **Cycling Risk = 12.592 – 0.22*Cycling Level**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	12.592	1.550		8.122	.000
	Level of cycling (cyclists over 1000 persons)	-.022	.020	-.269	-1.083	.296

a. Dependent Variable: Risk per 1000 cyclists

Hl-Lr group

Size: 114 cities with Cycling Level above the national mean (the mean of the 319 cases above) and with Cycling Risk below the national mean

Function: **Cycling Risk = 5.289 – 0.03*Cycling Level**

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	5.289	.269		19.653	.000
	Level of cycling (cyclists over 1000 persons)	-.003	.003	-.100	-1.063	.290

a. Dependent Variable: Risk per 1000 cyclists

Ll-Hr group

Size: 95 cities with Cycling Level below the national mean (the mean of the 319 cases above) and with Cycling Risk above the national mean

Function: **Cycling Risk = 31.099 – 0.759*Cycling Level**

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	31.099	4.470		6.958	.000
	Level of cycling (cyclists over 1000 persons)	-.759	.205	-.358	-3.697	.000

a. Dependent Variable: Risk per 1000 cyclists

Ll-Lr group

Size: 93 cities with Cycling Level below the national mean (the mean of the 319 cases above) and with Cycling Risk below the national mean

Function: **Cycling Risk = 9.210 – 0.097*Cycling Level**

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	9.210	.685		13.439	.000
	Level of cycling (cyclists over 1000 persons)	-.097	.021	-.437	-4.638	.000

a. Dependent Variable: Risk per 1000 cyclists