Testing for statistically significant changes in road accident and casualty numbers

Introduction

Using data for London between April 2004 and March 2007, Spiegelhalter and Barnett (2009) found that homicides in London have a predictable level of fluctuation over time, which allowed them to apply statistical techniques for analysis. This methodology was used by the Home Office to show that the number of homicide incidents occurring per day in England and Wales closely follows a Poisson distribution. Statistical tests were then carried out to determine whether the numbers of homicide incidents in any two years are statistically significantly different from one another (see here). This methodology has been applied to show that the number of fatal road accidents per day in Great Britain follows a Poisson distribution and to test whether changes in fatal road accidents between years are statistically significant (see section 1). In addition, it is important to indicate whether the latest road casualty figures represent a statistically significant deviation from prior observed numbers and trends. The methodology used to test for such changes can be found in section 2. Further work will be undertaken to improve on the methodology outlined here as well as considering different approaches.

1) Statistical interpretation of year on year changes in fatal road accidents

The number of road fatalities recorded by the police fluctuates from year to year and there is interest in knowing the extent to which these fluctuations represent an indication of a real underlying trend as opposed to random year-to-year variation. It should be noted that the discussion in this section is related to the statistical properties of the data, and it should not be interpreted as saying that small changes in the numbers of road fatalities are not of concern.

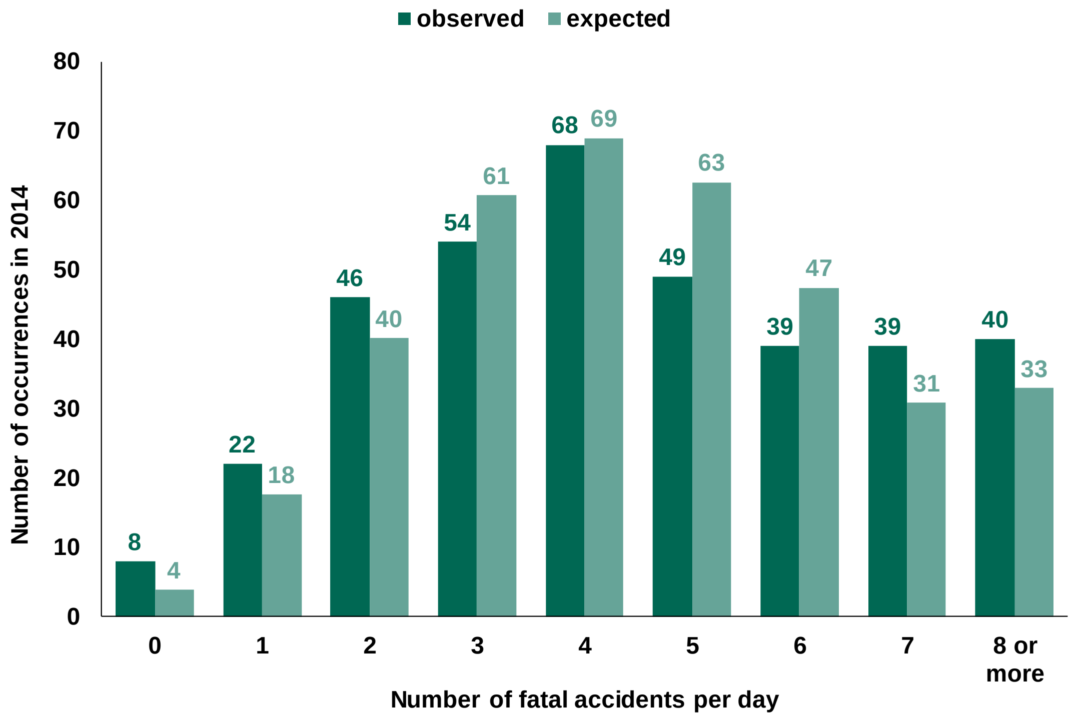
This section presents analyses of fatal road accidents, defined as an accident in which at least one person is killed. Fatal road accidents can have more than one fatality so the numbers in this section therefore differ from the total number of reported fatalities. The focus of this analysis is on accidents as it is not possible to statistically model multiple fatalities that relate to one case and were recorded on the same day. For example, on 22nd December 2014 there was a road accident in Scotland in which six people were killed. In this analysis, this road accident was treated as one incident.

The probabilities of fatal road accidents happening can be modelled by a Poisson distribution[1]. This can be used to test whether any change in the numbers of fatal road accidents per year is statistically significant, or what can be termed within the range of expected 'natural variation' of the data.

Expected fatal road accidents per day

In 2014, the police recorded 1,658 independent fatal road accidents in Great Britain. Chart 1 illustrates that the observed number of fatal road accidents on a daily basis matches the expected number under the Poisson distribution. For example, from knowing there was an average of 4.5 fatal road accidents a day in 2014, we would predict over the time period of 365 days that there would be 69 days on which there would be exactly four independent fatal road accidents. This is close to the observed number of 68 fatal road accidents, indicating that the occurrence of these apparent 'clusters' is not as surprising as one might anticipate. A statistical test (χ^2) shows no significant difference between the expected and observed figures. Thus, the observed figures are in fact Poisson distributed. This allows for calculation of the number of days on which it would be expected that no fatal road accident or one fatal road accident occurs and so on.

Chart 1: Observed and expected number of fatal road accidents recorded on a day: GB, 2014



---

[1]The Poisson distribution expresses the probability of the number of events occurring in a given period of time if these events occur with a known average rate and independently from each other.

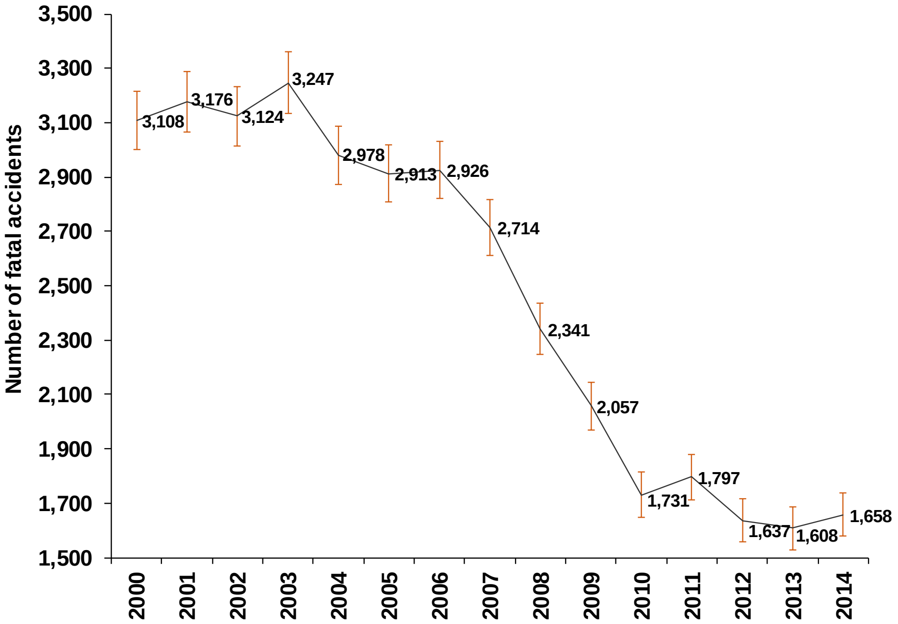
Trend analysis

Knowing that fatal road accidents are statistically distributed allows the application of statistical techniques in order to assess longer-term trends. Chart 2 shows the number of fatal road accidents since 2000. For each year, the observed count is given along with a 95% confidence interval on the Poisson error. The interval represents the range of values one would expect to measure 95% of the time if the underlying risk of a road fatality remained unchanged. These confidence intervals can be used as a rough approximation to determine whether the numbers of fatal road accidents in any two years are statistically significantly different from one another. If the confidence intervals do not overlap, one can infer there has been a significant change in the underlying risk of a road fatality. However, it is possible for the confidence intervals to overlap and for there still to be a statistically significant change in the number of fatal road accidents. In order to measure the significance more accurately, it is appropriate to use a statistical test.

As has been shown, the Poisson distribution can be applied to the number of fatal road accidents per year and the number of these incidents is large enough to approximate the normal distribution. Therefore, a statistical test (in this case a Z-test) can be used to determine if the counts in each year are statistically significantly different from one another at the 95% confidence level i.e. whether there has been a true change in the underlying risk.

Based on this statistical test, the number of fatal road accidents recorded in 2014 (1,658) was statistically significantly lower at the 5% level to the number recorded in every year shown in Chart 2 with the exception of 2010, 2012 and 2013. This means the risk of becoming a road fatality was, in fact, lower for 2014 compared with those years. At the 5% level, we cannot say that the number of fatal accidents in 2014 was statistically significantly higher than in 2013. Therefore, the 4% increase in road fatalities in 2014 compared to 2013 is likely to have arisen due to natural statistical variation in the data.

Chart 2: Fatal road accidents: GB, 2000-2014



Predicting the number of fatal accidents by a given date

If we assume there to be no major change in the average number of fatal road accidents per day between 2013 and 2014 then the number of fatal road accidents by a given date in 2014 can be predicted using the average number of fatal accidents per day in 2013 (1,608/365=4.4). The expected number of fatal accidents by a given date in 2014 is calculated by multiplying the average number of fatal accidents per day in 2013 (4.4) by the number of days passed in the year e.g. by the 20/11/14 324 days have passed so we would have expected 4.4*324 = 1,427 fatal road accidents by this date. Prediction limits have also been calculated and we would have been 95 per cent sure that the number of fatal road accidents by 20/11/14 would be between 1,352 and 1,503.

The table below compares the number of fatal road accidents observed in 2014 to what would have been predicted based on the average number occurring per day in 2013. By the end of March 2014 there were 356 fatal road accidents. This was well below the central prediction based on the 2013 average and just outside the 95 per cent confidence interval. The number of fatal road accidents that occurred by the end of September was 1,178 which was below the central prediction but within the confidence interval. There were 480 fatal road accidents in the last quarter of 2014 which led to the number of fatal road accidents observed by the end of 2014 being above the central prediction. However it is within the 95 per cent confidence interval, so despite being higher than 2013 the 1,658 fatal road accidents in 2014 was not really unusual.

Actual and predicted fatal road accidents: GB, 2014

period ending	observed	central prediction	lower	upper
31/03/14	356	396	357	436
30/06/14	759	797	741	854
30/09/14	1,178	1,203	1,133	1,272
31/12/14	1,658	1,608	1,528	1,688

2) Statistical interpretation of changes in the road casualty trend

To prevent over-ambitious interpretations and reactions to latest outcomes, it is important to indicate whether the latest road casualty figures represent a statistically significant deviation from prior observed numbers and trends. Time series models that take into account previous trends have been used to forecast road fatalities, seriously injured and slightly injured casualties in each quarter of 2014. The forecasts include confidence intervals which enable a judgement to be made as to whether the new incoming data for 2014 is a significant deviation from previous trend.

The X-13ARIMA-SEATS seasonal adjustment package is used in official statistics in the United Kingdom (UK) to seasonally adjust statistical series. The regARIMA part of the X-13ARIMA-SEATS program enables time series to be modelled to extend the series forwards (adding forecasts) and backwards (adding backcasts). This functionality has been utilised to extend the 2000 to 2013 quarterly road fatalities, seriously injured and slightly injured casualty series to 2014 and predict the casualties in each quarter.

Road fatalities

The automatic selection of the order of the ARIMA model in X-13ARIMA-SEATS was an (0,1,1)(0,1,1) ARIMA model with log transformation. This chosen model was used to generate the forecasts below.

2014 road fatality forecasts: GB

	Central forecast	Lower	Upper
Q1 2014	374	324	433
Q2 2014	408	345	481
Q3 2014	445	370	535
Q4 2014	438	358	535
2014	**1,665**	**1,397**	**1,984**

It is important to remember that these forecasts give an indication of what we would expect to happen to fatalities in 2014 **if and only if the fatality trend continues in the same way as over the period 2000 to 2013**. These forecasts can be used as a baseline to identify changes when compared to the final 2014 fatality figures.

2014 final road fatality figures

The final quarterly figures for 2014 can be compared to the forecast confidence intervals and if the final figure lies outside then we can say that there has been a significant deviation from the trend observed in fatalities over 2000 to 2013.

2014 final road fatality figures

Q1 2014	376
Q2 2014	428
Q3 2014	457
Q4 2014	514
2014 total	**1,775**

The Q1 2014 fatalities are very close to the forecasted Q1 2014 value. Both the Q2 2014 and Q3 2014 fatalities are slightly higher than the forecasted values, but reasonably close. The fourth quarter is much higher than the forecasted value, which shows that this was a higher final quarter than observed in recent years. However, it doesn't lie outside of the confidence interval of the forecasted value, so we cannot say that this quarter represents a significant deviation from the prior observed trend over 2000 to 2013.

Seriously injured casualties

The automatic selection of the order of the ARIMA model in X-13ARIMA-SEATS was an (0,1,1)(0,1,1) ARIMA model with no log transformation. This chosen model was used to generate the forecasts below.

2014 seriously injured casualty forecasts: GB

	Central forecast	Lower	Upper
Q1 2014	4,642	4,057	5,228
Q2 2014	5,413	4,773	6,053
Q3 2014	5,848	5,157	6,540
Q4 2014	5,470	4,731	6,209
2014	**21,374**	**18,718**	**24,029**

It is important to remember that these forecasts give an indication of what we would expect to happen to seriously injured casualties in 2014 **if and only if the seriously injured trend continues in the same way as over the period 2000 to 2013**. These forecasts can be used as a baseline to identify changes when compared to the final 2014 seriously injured figures.

2014 final seriously injured casualty figures

Q1 2014	5,168
Q2 2014	5,805
Q3 2014	6,035
Q4 2014	5,799
2014 total	**22,807**

Serious injuries in all quarters are higher than the central forecasted values and are towards the upper end of the confidence intervals. However, no quarter lies outside of the confidence interval of the forecasted value, so we cannot say that there has been a significant deviation from the prior observed trend over 2000 to 2013 in any quarter.

Slightly injured casualties

The automatic selection of the order of the ARIMA model in X-13ARIMA-SEATS was an (0,1,1)(0,1,1) ARIMA model with no log transformation. This chosen model was used to generate the forecasts below.

2014 slightly injured casualty forecasts: GB

	Central forecast	Lower	Upper
Q1 2014	34,132	31,125	37,139
Q2 2014	36,787	33,766	39,808
Q3 2014	39,789	36,754	42,823
Q4 2014	40,954	37,905	44,002
2014	**151,661**	**139,550**	**163,773**

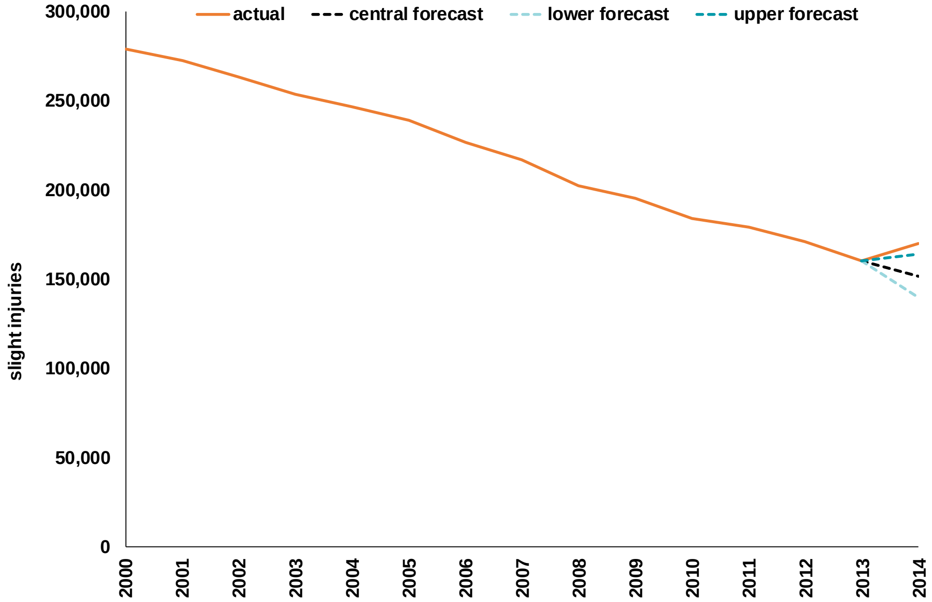
It is important to remember that these forecasts give an indication of what we would expect to happen to slightly injured casualties in 2014 **if and only if the slightly injured trend continues in the same way as over the period 2000 to 2013**. These forecasts can be used as a baseline to identify changes when compared to the final 2014 slightly injured figures.

2014 final slightly injured casualty figures

Q1 2014	40,349
Q2 2014	41,790
Q3 2014	42,972
Q4 2014	44,784
2014 total	**169,895**

The final 2014 slightly injured casualty figures all lie outside of the forecasted confidence intervals. Therefore, there has been a statistically significant deviation from the 2000 to 2013 trend in slightly injured casualties in all quarters of 2014. The central forecast for slight injuries in 2014 is 151,611 with a lower bound of 139,550 and an upper bound of 163,773. The actual slightly injured figure of 169,895 is well above the upper bound. The statistically significant deviation from the 2000 to 2013 slightly injured casualties trend is shown in the chart below. Slightly injured casualties decreased each year over 2000 to 2013, before the 2014 increase.

Slightly injured casualties: GB, 2000-2014



Total casualties

Summing the above forecasts enables forecasts for total casualties to be obtained. These forecasts give an indication of what we would expect to happen to total casualties in 2014 **if and only if the trend continues in the same way as over the period 2000 to 2013**.

2014 total casualty forecasts: GB

	Central forecast	Lower	Upper
Q1 2014	39,149	35,506	42,800
Q2 2014	42,608	38,884	46,343
Q3 2014	46,081	42,280	49,897
Q4 2014	46,861	42,994	50,746
2014	**174,700**	**159,665**	**189,786**

The final 2014 total casualty figures all lie outside of the forecasted confidence intervals, apart from Q3 2014. The central forecast for total casualties in 2014 is 174,700 with a lower bound of 159,665 and an upper bound of 189,786. The actual total casualty figure of 194,477 is above the upper bound.

2014 final total casualty figures

Q1 2014	45,893
Q2 2014	48,023
Q3 2014	49,464
Q4 2014	51,097
2014 total	**194,477**

The statistically significant deviation from the 2000 to 2013 total casualties trend is shown in the chart below. Total casualties decreased each year over 2000 to 2013, before the 2014 increase.

Total casualties: GB, 2000-2014