UNION INTERNATIONALE DES CHEMINS DE FER

INTERNATIONALER EISENBAHNVERBAND

INTERNATIONAL UNION OF RAILWAYS

10 th World Level Crossing Symposium
Safety and Trespass prevention

Name:	Dr Chris Elliott
Topic:	Risk management
Sub Theme:	Economics
Title of Paper:	Statistical analysis of risk
Job Title:	Director
Company Name:	Pitchill Consulting Ltd
Country:	UK
Tel:	+44 1483 273793
E-mail:	chris.elliott@pitchill.com
Fax:	+44 1483 273465

Statistical analysis of risk

Dr Chris Elliott FREng

Pitchill Consulting Ltd, Magalee, Moon Hall Road, Ewhurst, Surrey, GU6 7NP, UK

Abstract

For the first time there will be comparable data for railway risks in many countries. This is in part, at least for Europe, as a result of the Railway Safety Directive which requires data on which to base Common Safety Indicators and Common Safety Targets. This paper examines the data that is becoming available and its reliability, then uses it to identify statistical links between possible causes of risk and the consequences. The results do not prove a causal link but can identify the features of LX that are correlated with the level of risk and therefore provide guidance when seeking to improve management of the road/rail interface. It also comments on the challenge of defining Common Safety Indicators and Common Safety Targets

The principle conclusions are:

- data is available but there are omissions, not helped by inappropriate secrecy in some cases
- the best predictors of LX accident rates are the behaviour of road users, suggesting that improving LX safety is primarily an issue for roads authorities
- indicators intended to compare the relative safety performance of countries must be carefully designed to be robust and defensible, especially is used to inform legally binding targets.

Introduction

Accidents at level crossings, between trains and road vehicles, are perceived as a rail issue which must be addressed by actions by the railway company. In principle this is not unreasonable; LX accidents contribute a large fraction of the safety risk of railways. Conversely, LX accidents represent a much smaller fraction of the risk of road travel so the road authorities are reluctant to divert resources to reducing the risk.

There is little point in demanding action by the railways if the cause of the accident lies out of the railways' control. It is hard to prove where the responsibility lies but it is possible to draw inferences from the statistics of those accidents.

The Rail Safety Directive requires national authorities to submit data on the numbers of people killed or seriously injured at LX each year, and the number of LX. There is also a Common Safety Target for safety at LX. However, those statistics alone are not enough to indicate why LX accidents occur.

Approach

This paper seeks to find the factors that are correlated with LX accident rates. Although the existence of a correlation does not necessarily imply a causal link, it is not unreasonable to assume that there is at least a common root cause for the safety of LX and the factors which are correlated with it. This is analogous to the use in medical research of “risk factors”. For example, obesity is a risk factor for heart disease in that obese people are more likely to suffer heart disease. It may be that heart disease is caused by obesity; it may be that both heart disease and obesity have a common root cause. In either case, the existence of the correlation provides researchers with pointers to the aspects of diet, lifestyle, genetic makeup and other factors to explore.

The approach taken here is to seek the combination of factors that is the best predictor of LX risk. The steps are:

1. identify a measure of risk, such as the number of people killed in LX accidents per year
2. identify a combination of risk factors that might be a good predictor (taking care that the combination has itself scales properly; a good test is to imagine two identical countries being merged – the Safety Indicator should not change)
3. calculate Safety Indicators, defined as the measured level of risk in each country divided by a scaling factor which is a combination of factors for that country.
   Formally, the Safety Indicator for country m is referred to as SI_m and is defined by $SI_m = F_m / S_m$ where:

   - F_m the average number of LX fatalities per year in country m

   - S_m the scaling parameter for country m

   For example, one simple safety indicator is (a measure of risk) divided by the number of LX in the country (a scaling factor).
4. calculate the spread in values of SI_m across all countries for which data is available. Three measures of spread are considered: standard deviation of the values, a weighted measure of deviation that gives greater significance to larger countries and the ratio of maximum to minimum SI.

A low value of spread implies a good correlation and hence that the chosen SI is a good predictor of the level of safety risk.

Availability of data

There is data for the number of people killed each year in LX accidents but no reliable data for the number of serious injuries. There is also no comparable data between countries for the type of victim – road vehicle driver or passenger, pedestrian, train passenger etc. There is data for the number of LX accidents per year.

Data is also needed for the factors that might be used to scale the risk data. Eurostat publishes some of this data and the United Nations Economic Commission for Europe also issues useful statistics.

In practice it was not easy to find comprehensive and reliable data. Different sources were inconsistent; the most authoritative was used but even then private information available to the author suggests even that is not accurate. Data for European countries that are not members of the EU is even more difficult to find.

Where it was possible to obtain an almost complete dataset for a country but one number was unavailable, the missing value was estimated by assuming that the country conformed to the European average. For example, for one country it was not possible to find the number of passenger car km per year. It was assumed that this would be:

$$\frac{\text{(number of cars in that country) x (total passenger km in all other countries)}}{\text{(total number of cars in all other countries)}}$$

Data for 26 European countries was included in the research.

Factors that might contribute to risk

The table below lists some of the factors that might influence the level of risk at level crossings. Some of them cannot be measured directly so the second column lists readily-available data that might be used as a surrogate for the risk factor.

Risk factors	Readily available data
number of LX	Ideally the SI would be broken down according to different types of LX. ERA's working group is developing such a classification but this is still not complete. This note considers only the total number of LX of all types. Symbol: **N**
intensity of use by trains – ideally the frequency with which trains pass each LX	A useful measure of the frequency of use by trains is the number of train km per year divided by the length of track in km. This is a measure of the average number of times that a train passes any point on the network each year. Symbol: **T**
intensity of use by road vehicles – ideally the frequency with which road vehicles pass each LX	Data is readily available for each country for the numbers of cars, buses and commercial vehicles in use, for the numbers of passenger km by car and by bus and the number of vehicle km for freight vehicles. For purposes of this paper, two measures of the intensity of use of the roads are tested: • the number of cars on the roads divided by the total length of all roads in the country: Symbol **R1** • the number of car passenger km divided by the total length of all roads in the country: Symbol **R2** These measures include many assumptions, such as that the number of bus and freight movements is proportional to the number of car movement, and that the number of car km is proportional to the number of car passenger km (ie the average number of people in a car is the same in every country).
level of risk accepted by road users – ideally a measure of the propensity of national road users to breach road safety laws and therefore to abuse level crossing safety.	Much of the risk of LX arises from abuse by road users who fail to obey warnings and signs. It is believed that the willingness to do this varies with national culture. For the purposes of this paper, two measures of the propensity of national road users to breach road safety laws are tested: • number of people killed on the roads per year divided by the population of the country: **P1** • number of people killed on the roads per year divided by the number of car passenger km: **P2**

Table 1: Definitions of data used for scaling factors.

There are many different SI that might be constructed even from that small number of factors. This paper uses 14 different Safety Indicators for illustration; more could be defined:

	Description	**Definition**
SI 1	Number of fatalities per year per LX	$F_m / (N_m)$
SI 2	Correction for train intensity	$F_m / (N_m \times T_m)$
SI 3	Correction for train and road intensity	$F_m / (N_m \times T_m \times R1_m)$
SI 4	As SI 3 with different measure of road intensity	$F_m / (N_m \times T_m \times R2_m)$
SI 5	Correction for train and road intensity and road driver behaviour	$F_m / (N_m \times T_m \times R1_m \times P1_m)$
SI 6	As SI 5 with different measure of road driver behaviour	$F_m / (N_m \times T_m \times R1_m \times P2_m)$
SI 7	As SI 5 with different measure of road intensity	$F_m / (N_m \times T_m \times R2_m \times P1_m)$

SI 8	As SI 6 with different measure of road intensity	F_m / (N_m x T_m x $R2_m$ x $P2_m$)
SI 9	Correction for road driver behaviour	F_m / (N_m x $P1_m$)
SI 10	As SI 9 with different measure of road driver behaviour	F_m / (N_m x $P2_m$)
SI 11	Correction for train intensity and road driver behaviour	F_m / (N_m x T_m x $P1_m$)
SI 12	As SI 11 with different measure of road driver behaviour	F_m / (N_m x T_m x $P2_m$)
SI 13	Correction for road intensity	F_m / (N_m x $R1_m$)
SI 14	Simple ratio of LX fatalities per year to road fatalities per year, no allowance for number of LX or intensities	(No of LX fatalities) (No of road fatalities)

Table 2: Definition of Safety Indicators tested

A similar set can be constructed using the number of accidents, A_m, instead of F_m.

In practice it is convenient to adjust each SI so that the average across Europe is 1. The values for each country then represent the amount by which the risk in that country varies from the European average. This allows the different SI definitions to be compared by inspection.

Results

The calculated Safety Indicators are shown in Annex 1.

1. The results using the number of accidents (A_m) were similar to those for number of fatalities (F_m) but showed consistently greater spread. This probably reflects greater variation in definitions between countries. The accident-based results have been omitted; the conclusions would be the same if they were included but are less clear because of that greater scatter.

2. The countries cannot be identified from the data; they are in a random order. This is necessary because some of the data was provided in confidence to the author.

3. Three different measures of spread are shown: the standard deviation of the SI; the ratio of maximum to minimum value of the SI and a root mean square measure that is weighted according to the size of the country. A single overall spread is calculated by multiplying these three together.

4. The best predictor of safety risk is SI 14 – the number of road fatalities in the country per year. The next best is SI 10, which is (number of level crossings) x number of people killed on roads per year per passenger km. The third best is SI 1, which is the number of level crossings.

5. The apparent relative safety performance of countries depends critically on which SI definition is adopted. For example:
   - 17 of the 26 countries are above the European average according to at least one definition of SI and below the European average according to another
   - 7 different European countries have the best safety performance, depending on which definition of SI is adopted
   - 6 different European countries have the worst safety performance, depending on which definition is of SI adopted

Conclusions

1. On the availability of data:

1.1 Reliable and consistent data is hard to find, especially for countries outside the EU. This may improve when the first set of Common Safety Indicator (CSI) data is published by ERA, although the problems of inconsistent definitions might remain.

1.2 There is a wholly inappropriate concern about secrecy. Safety improvements come when safety performance is exposed to public scrutiny; the railway sector in many countries hides its safety performance. This too may improve with the publication of CSIs.

1.3 Even with those concerns, there is a wealth of data available and properly designed CSIs could be used.

2. On the predictors of LX accident rates:

2.1 The results indicate clearly that the dominant factor that determines LX accident rate is the behaviour of road users.

2.2 The number of LX in the country is also relevant but there is little evidence that the intensity of use of the railway is significant.

2.3 Given these conclusions and the observation that most of the victims of LX accidents are road users, there is powerful support for the argument that LX safety is primarily a roads issue and not a rail issue. Action to improve safety may require cooperation between rail and road authorities but these results provide support for an argument that the initiative and funding should come from the roads authorities.

3. On the calculation of Common Safety Indicators:

3.1 There are many different ways of calculating a Safety Indicator that allow comparison between countries. This paper has illustrated this with 14 different definitions, all of which are plausible but which reach very different relative performance indications.

3.2 If indicators of this type are to be used as the basis of legally binding safety targets, there must be a robust and defensible reason for selecting the data used to measure safety (or lack of safety) and the data used to scale that data to allow for the scale of rail operations in that country.

Acknowledgements and Disclaimer

Some of the original preparatory work for this paper was carried out under contracts funded jointly by UIC Safety Platform and the UK Rail Safety and Standards Board.

The views expressed in this paper are solely those of the author and do not necessarily represent the views of any other organisation or person with whom the author has a professional involvement.

Annex: Table of Safety Indicators and calculated spreads

Country	SI	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1		0.4	0.8	1.8	2.5	2.2	1.9	3.0	2.7	0.5	0.4	1.0	0.9	0.9	0.2
2		0.6	0.6	0.7	0.6	0.9	1.6	0.8	1.4	0.7	1.3	0.7	1.3	0.7	0.7
3		2.4	6.5	4.2	4.4	4.2	8.1	4.5	8.6	2.4	4.6	6.6	12.7	1.5	2.0
4		1.9	2.3	3.5	4.3	2.7	3.2	3.3	4.0	1.5	1.8	1.8	2.1	3.0	1.0
5		0.8	1.1	0.6	0.7	0.6	0.4	0.6	0.4	0.8	0.5	1.1	0.7	0.4	1.1
6		4.8	16.6	46.5	46.2	80.2	19.9	79.7	19.8	8.3	2.1	28.6	7.1	13.5	1.5
7		5.0	9.7	39.4	40.2	32.8	36.6	33.5	37.4	4.2	4.7	8.1	9.0	20.4	2.5
8		2.2	1.5	1.8	1.6	1.7	2.7	1.5	2.4	2.1	3.4	1.4	2.3	2.6	1.3
9		1.8	2.1	2.6	3.1	2.0	2.0	2.4	2.4	1.4	1.4	1.6	1.6	2.2	3.6
10		0.6	0.4	0.2	0.2	0.3	0.5	0.3	0.5	0.8	1.3	0.5	0.9	0.3	0.7
11		2.7	3.6	1.1	1.3	1.1	1.3	1.3	1.5	2.7	3.3	3.5	4.3	0.8	0.6
12		0.5	0.3	0.2	0.1	0.3	0.5	0.3	0.5	1.0	1.7	0.6	1.0	0.3	0.4
13		2.4	5.9	54.1	7.8	38.0	34.8	5.5	5.0	1.7	1.6	4.2	3.8	22.1	0.7
14		1.3	2.2	2.6	3.7	1.8	1.3	2.6	1.8	0.9	0.7	1.6	1.1	1.5	1.1
15		0.4	0.9	2.9	4.0	1.3	1.0	1.9	1.4	0.2	0.1	0.4	0.3	1.3	0.2
16		0.2	0.4	0.7	0.6	1.4	2.3	1.2	1.9	0.4	0.7	0.8	1.4	0.4	1.1
17		0.6	0.4	0.3	0.2	0.4	0.7	0.4	0.7	0.9	1.7	0.6	1.0	0.4	2.3
18		0.3	0.1	0.2	0.1	0.3	0.5	0.2	0.4	0.4	0.7	0.2	0.4	0.4	0.5
19		5.0	11.1	47.9	28.3	22.9	26.2	13.5	15.5	2.4	2.7	5.3	6.0	21.6	1.3
20		1.3	0.5	0.4	0.5	0.3	0.3	0.4	0.4	1.3	1.2	0.5	0.5	0.9	1.1
21		0.9	0.4	0.2	0.2	0.5	0.7	0.5	0.7	1.9	2.5	0.8	1.0	0.6	2.1
22		0.5	1.1	1.8	1.5	2.6	4.6	2.2	3.8	0.7	1.3	1.6	2.8	0.8	2.0
23		1.1	1.7	3.5	4.6	2.9	2.0	3.8	2.6	0.9	0.6	1.4	0.9	2.4	1.7
24		0.2	0.3	0.3	0.3	0.6	1.0	0.6	0.9	0.4	0.6	0.5	0.8	0.2	1.4
25		1.0	0.8	0.6	0.6	0.6	1.1	0.6	1.1	1.1	2.0	0.8	1.5	0.7	0.4
26		1.2	0.8	0.8	0.8	0.8	1.2	0.8	1.2	1.2	1.7	0.8	1.2	1.1	2.6
Spread: STD		1.45	4.04	16.9	12.3	17.8	10.6	16.4	8.14	1.64	1.21	5.61	3	6.93	0.84
Spread: max/min		27.2	112	328	331	276	112	368	104	40.4	31.4	124	39.6	94.2	19.4
Spread: weighted		0.37	0.54	0.96	0.92	0.6	0.41	0.58	0.4	0.29	0.26	0.37	0.3	0.58	0.29
Overall spread		**14.4**	**244**	**5320**	**3741**	**2967**	**487**	**3505**	**342**	**19.2**	**9.99**	**259**	**35.2**	**381**	**4.8**