ELSEVIER

Journal of Hazardous Materials 111 (2004) 105–110

Journal of Hazardous Materials

www.elsevier.com/locate/jhazmat

Accident versus near miss causation: a critical review of the literature, an empirical test in the UK railway domain, and their implications for other sectors

Linda Wright [a,*], Tjerk van der Schaaf [b]

[a] *Prorail, De Inktpot, Moreelsepark 3, 3511 EP Utrecht, Postbus 2038, Utrecht 3500 GA, The Netherlands*
[b] *Eindhoven University of Technology, Eindhoven, The Netherlands*

Available online 14 May 2004

Abstract

An essential assumption for the usefulness of basing accident prevention measures on minor incidents is the common cause hypothesis: that causal pathways of near misses are similar to those of actual accidents (such as injuries and damages). The idea of a common cause hypothesis was originally proposed by Heinrich in his seminal book "*Industrial Accident Prevention*" [McGraw-Hill, New York]. In this paper, it is argued that the hypothesis of similarity of causes for major and minor accidents has become confounded with the interdependence of the ratio relationship between severity and frequency. This confounded view of the hypothesis has led to invalid tests of the hypothesis and erroneous conclusions. The evidence from various studies is examined and it is concluded that the hypothesis has not been properly understood or tested. Consequently, such a proper test was carried out using data from the UK railways which were analysed using the confidential incident reporting and analysis system (CIRAS) 21 cause taxonomy. The results provide qualified support for the common cause hypothesis with only three out of the 21 types of causes having significantly different proportions for the three consequence levels investigated: 'injury & fatality', 'damage' and 'near miss'.

Keywords: CIRAS; Formal inquiry; SPAD

1. Introduction

Decision making about investing in safety improvements is usually based upon the relative importance of root causes in accidents and failures. However, such decisions can only be reached reliably by referring to statistics from large databases. As accidents themselves are (fortunately) too few in number to aid such decision making processes the use of near misses to dramatically increase the number of data in databases is one way to counteract this problem. This use of near misses as causal predictors of later, more serious, accidents is based upon the assumption that these near misses and accidents have the same relative causal patterns (the so-called common cause hypothesis). Such a causal relationship is also a vital argument to motivate employees to contribute to near miss reporting schemes on a voluntary basis.

When the common cause hypothesis is discussed there is inevitably discussion of the ratio data studies performed by Heinrich [1], Bird [2] and Salminen et al. [3]. This very starting point is central to the way thinking about the common cause hypothesis has become focussed. Heinrich's original triangle was not intended to convince the reader of the commonality of causes between different accident outcomes, but to illustrate the fact that prevention need not wait until an accident occurred, and that prevention should not only be aimed at the most severe consequences but also to events at the lower levels of triangle. In this endeavour, Heinrich was successful. The ratio triangles or icebergs are used profusely in industry today. However, Heinrich did not base the common cause hypothesis upon the ratio relationship between major accidents, minor accidents and no injury accidents, although the proposed ratio relationship seemed (to him) to substantiate the idea of a common causal pathway. Today, the common cause hypothesis has come to imply a ratio relationship of consequences (and not of causes). How did then this confusion arise, and where did the common cause hypothesis spring from?.

The validity (or refuting) of the common cause hypothesis has major implications for accident prevention and analysis. If the different levels of severity really do have

* Corresponding author. Tel.: +31-30-235-35-94; fax: +31-30-235-89-85.
E-mail addresses: linda.wright@prorail.nl (L. Wright), t.w.v.d.schaaf@tm.tue.nl (T. van der Schaaf).

0304-3894/$ – see front matter
doi:10.1016/j.jhazmat.2004.02.049

completely different patterns of causes, then industry has been concentrating on levels of severity (near misses, small failures) which may have little impact on the frequency of accidents which cause the greatest injuries. On the other hand, if common causal pathways can be demonstrated then a concerted effort is required to collect appropriate data (i.e. via voluntary near miss reporting schemes) and to ensure that causal analysis techniques become more widespread.

2. The confounding of the ratio model versus the common cause model

How did then these two separate models become so interdependent? Heinrich [1] discusses the relevance of the triangle model as providing some evidence for similarity of cause of frequency and severity (i.e. reduction of events at the bottom of the triangle should lead to a reduction in the number of events at the top of the triangle). However, the common cause hypothesis itself emerged from data analysis rather then from a deduction based on the ratio model, although Heinrich did not further discuss the common cause hypothesis in subsequent editions. The ratio model as evidence for a common causal pathway has become embedded in the literature and is not questioned. The rebuttal of the common cause hypothesis by Petersen [4,5] based on the ratio relationship not holding when serious injuries are reduced via prevention efforts, has further confounded the interdependence of the two models. The ratio of accidents and incidents occurring says nothing about the underlying causal factors—not to mention the barriers in place (e.g. [6]) or the error recovery processes that took place (e.g. [7])—so that it is difficult to understand how the ratio relationship has become so entwined with causation.

3. Literature review

The table below summarises the limited literature in the area.

4. Conclusions from the literature review

It is apparent that frequency, severity and causal mechanism have become inextricably linked (Heinrich, op cit; Petersen, op cit). It appears that researchers have not differentiated between the causes of severity and frequency and the causes of accidents and incidents. Thus, if a ratio is established and the data follow the pattern of the ratio found by Heinrich or Bird, it is suggested that the similar cause hypothesis is validated. Where the ratio is invalidated i.e. severe incidents do not occur at the expected frequency when compared with minor or no injury incidents the similar cause hypothesis is discounted. These positions fail to take into account the fact that the ratio model (whether validated or not) has no bearing on the similar cause hypothesis. A valid test of the common cause hypothesis should be based solely on causal patterns and not ratio data. Such a test should be determined by using data that has been analysed for causal factors and not be based simply on frequencies of accident severity. Causality has no bearing on the ratio relationship propounded by the iceberg model and vice versa.

5. A proper test defined

There are three possible ways in which the common cause hypothesis can be tested:

- comparing the actual occurrence of causal codes based on a dichotomy of causal codes being either present or absent;
- comparing the actual frequency of causal codes contributing to the different incident outcomes; and
- comparing the relative proportions of causal codes contributing to the various incident outcomes.

Testing the common cause hypothesis using causal codes as either present or absent is the weakest method of testing the hypothesis. Each time a causal code occurs more than once it still only counts as having occurred once. Hence the results conceal the actual or relative frequency of occurrence of causes. Comparing the actual frequency of causes that contribute to the severity levels may actually be inappropriate in general and lead to confounding results. This is the case where the data arising from the different consequence levels is collected and investigated in different ways. The incidents used for this study were investigated in three different ways: the more serious incidents were investigated via formal inquiries (where a panel of experts discuss the incident and interview all the staff involved), which consist of a greater depth of investigation than either signals passed at danger (SPAD) investigations or voluntary CIRAS reports. Hence it is unclear whether differences in the absolute numbers of causes assigned are due to the type of incident (e.g. serious events have more contributory causes than near misses) or to the investigation procedures. By using proportions to test the common cause hypothesis, the problems mentioned above are thus avoided.

6. Data collection

The three different investigation methods used by the UK railway industry and providing the data for this study are described below.

6.1. Formal inquiries

Following a major accident or an incident with the potential for a major loss, an internal formal inquiry may be performed. In the case of a formal inquiry, the company

Reference	Type of data used	Confounded view of the iceberg model?	Conclusions
[8]	Frequency data for major and minor accidents in manufacturing and construction industries	Yes. Confuses ratio of minor to major incidents as being the same as causal mechanisms of major and minor incidents	As ratios not in agreement with original iceberg theory, concluded that different causal mechanisms present between major and minor accidents
[9]	Frequency data Comparison of the type of activity taking place prior to accident occurring e.g. during manufacture	Yes. Basic misunderstanding of what constitutes causality. Confusion over activity being performed prior to incident and causes of incident	Supports similar cause hypothesis, as similar tasks were undertaken in the various categories prior to incidents occurring
[10]	Frequency data comparing the occurrence of lost time accidents and loss of containment	Yes. Confuses ratio data for causal data	As ratios not in agreement with original iceberg theory, concluded that different causal mechanisms present between major and minor
[11]	Data classified according to taxonomy of causes. Only for fatalities, all industrial types	No	Similar causes for all fatalities
[12]	Number of accident events assigned to non-lost time and lost time accidents as assigned by victim. Accident events are not described—unable to determine if appropriate causal data	Yes	As differences observed between the number of accident events assigned to the consequence (lost time or non-lost time) concluded different causes
[3]	Finnish accident research model of 14 factors applied to 20 fatalities and 79 serious accidents	No, although to cover all bases, the paper also examines accident type (e.g. struck by object) and part of body injured	Results support Petersen's different causation hypothesis more than identical causation hypothesis, based on Kolmogorov–Smirnov test comparing distributions i.e. the number of causes assigned to the different levels of severity
[13]	Causal factors according to the classification of MERS-TM: technical, human and organisational factors for near misses and actual events	No	Authors' state this data supports the common cause hypothesis—but only under certain severity conditions. Conservative significance level chosen
[14]	Causal factors according to confidential incident reporting and analysis system (CIRAS): technical, proximal, intermediate and distal for near misses and unsafe acts	No	Results based on preliminary analysis of data and comparison graphically. Differences noted between technical and organisational causes between the near misses and unsafe acts

seeks to learn where both individuals and systems can be improved. Thus, the investigation takes place at both the level of human error and organisational causes. Formal inquiries are performed by a panel of highly expert railway managers (usually four or five plus a union representative observer) who act as investigators and interviewers during the process. The investigations are intended to be comprehensive, with the aim of determining technical, human and organisational causes. Staff involved, including witnesses, are interviewed regarding their part in the incident and summarised in the final report, alongwith the conclusions of the panel. Further technical evidence such as speed calculations, brake tests, damage reports and re-enactments of the incident are also presented, as is scientific evidence relating to rail contamination.

6.2. *Signal passed at danger investigations*

SPAD investigations are performed following all signals passed at danger (i.e. red) without authority. They are often less comprehensive than formal inquiries and are investigated by fewer people. SPADs are usually detected automatically by the signaller, but on occasion are reported by the driver. Automatic detection is not the case in a minority of areas, which are not fitted with the appropriate technology (e.g. in depots). SPAD investigations are usually performed by a local manager, following an initial discussion between the driver who has passed the signal and the signaller who has detected it or has received the initial report from the driver. A SPAD investigation usually consists of a written report by the driver involved –no more than a dozen lines- and by any relevant member of staff who was present (e.g. conductor, guard or accompanying driver). Following the written report, the driver is interviewed by the manager and the findings are written in a brief report. These investigations are not comprehensive and usually stop at the level of determining the human or technical causes. Organisational causes are often not discussed or investigated. SPAD investigations are rarely accompanied by technical reports such as brake tests or rail contamination tests, unless the driver has complained about rail or train characteristics.

6.3. *CIRAS near miss report investigations*

CIRAS, the confidential incident reporting and analysis system, is the UK railway national system for the reporting and analysis of railway near misses. CIRAS reports are made voluntarily by railway staff such as drivers regarding near miss incidents and other safety issues. Reports are made on a form or by telephone initially detailing the incident or issue which a driver wishes to report. Following the initial report, CIRAS staff perform a critical incident interview [16] with each driver. These interviews include details about the 'what', 'when', 'where', 'why' and 'how' of the incident. The CIRAS reports used in this study all came from the same company as the formal inquiries and SPAD investigations. An incident is defined here as a specific example of a failure of personnel or equipment to operate as planned or a specific example of general public behaviour that has safety implications.

The table below shows the data source and level of severity of the incidents used.

Severity level	Data source			Total
	Formal inquiry	SPAD investigation	CIRAS report	
Fatality/ injury	17	0	0	17
Damage	18	7	0	25
Near miss	11	81	106	198
Total	46	88	106	240

7. CIRAS analysis

The data were analysed according to the University of Strathclyde CIRAS human factors model which is hierarchical (see [17] for a full description of the system). According to this model individual causal codes are subsumed under one of four top-level categories: 'technical', 'proximal', 'intermediate' and 'distal' which we called the 'macro' codes. These macro codes each comprise an exclusive set of individual causal codes, which we termed 'micro' codes. Thus the common cause hypothesis can be tested on two levels: the more general level of macro codes, and the specific level of the individual micro codes.

8. Inter-rater reliability

Inter-rater reliability is a vital (and often neglected) part of any analysis system. Data analysed via the CIRAS system are subject to periodic inter-rater reliability trials. Index of concordance was above 80% for each trial. To ensure the data used in this study were also reliably coded two independent raters (experienced in using the coding scheme) coded a total of 14 incidents from various classes of event used in this study. This resulted in an index of concordance of 78.4%.

9. Results and discussion

Fig. 1 shows how the four macro causal codes are distributed over the three levels of severity. A Chi-square test for proportions showed non-significant differences.

At the level of macro codes (i.e. the superordinate categories of technical, proximal, intermediate and distal) no significant differences were found in the proportion of causal codes between the three severity outcomes (injury, damage and near miss). However, despite the fact that these results

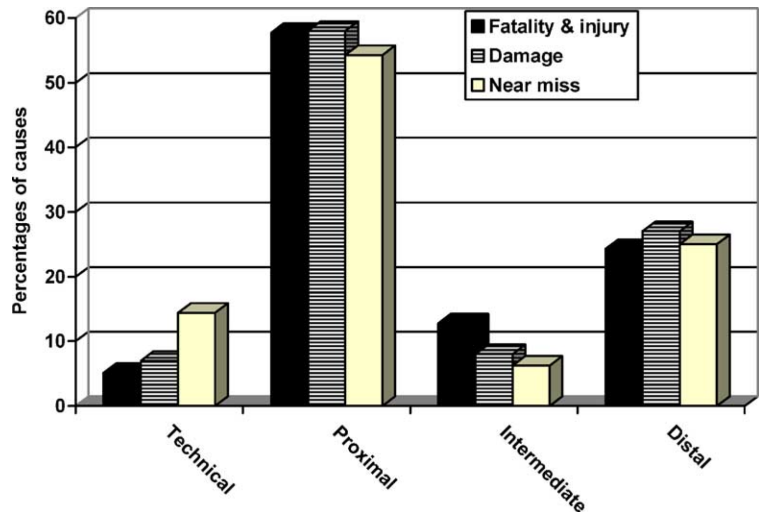

Fig. 1. Macro causal codes by level of severity.

are supportive of the common cause hypothesis, this macro level analysis tests the common cause hypothesis at a very general level.

At the more specific level of individual (micro) codes only three of 21 causal factors are significantly different, namely knowledge based errors, training and procedures. In all three cases, 'fatality & injury' have a greater proportion of these codes assigned than 'near miss' incidents. In terms of the differences found in knowledge based errors, Embrey and Lucas [15] suggest that knowledge based errors are more likely to be detected by someone other than the individual who made the error. It may therefore be the case that individuals reporting via the CIRAS system are unaware of knowledge based errors that they have committed and despite being interviewed these have not come to light.

There are also a number of possible explanations for the higher proportions of the causal codes training and procedures. Firstly, it may be the case that training and procedure causal factors are more prevalent in incidents with a more serious outcome. However, as training and procedures were high on the company agenda, it is more likely that such issues have been more frequently recognised by managers during the formal inquiries than have been revealed by staff during interview after submitting CIRAS reports. Issues such as training and procedures are traditionally management-driven and factors that management expect to have an impact on adverse events and therefore these are more likely to be identified as causal factors in incidents that are investigated by managers.

Therefore, overall, these findings provide qualified support for the common cause hypothesis within the railway domain.

As this study was limited to one domain, using one type of causal taxonomy, it is recommended that further empirical tests of the common cause hypothesis be performed for a number of different domains, and with other types of taxonomies. This would provide more robust evidence of the applicability of the theory.

References

[1] H.W. Heinrich, Industrial Accident Prevention, McGraw-Hill, New York, 1931.
[2] F.E. Bird, Damage Control, Insurance Company of North America, Philadelphia, 1966.
[3] S. Salminen, J. Saari, K.L. Saarela, T. Rasanen, Fatal and non-fatal occupational accidents: identical versus differential causation, Safety Sci. 15 (2) (1992) 109–118.
[4] D.C. Petersen, Techniques of Safety Management, McGraw-Hill, New York, 1971.
[5] D.C. Petersen, Techniques of Safety Management, second ed., McGraw-Hill, New York, 1978.
[6] O. Svenson, The accident evolution and barrier function (AEB) model applied to incident analysis in the processing industries, Risk Anal. 11 (3) (1991) 449–507.
[7] T.W. van der Schaaf, L. Kanse, Errors and error recovery, in: P.F. Elzer, R.H. Kluwe, B. Boussoffara, (Eds.), Human Error and System Design and Management, Springer-Verlag, London, 2000 (lecture notes in control and information sciences, vol. 253, pp. 27–38).
[8] A. Salomiemi, H. Oksanen, Accidents and fatal accidents—some paradoxes, Safety Sci. 29 (1998) 59–66.
[9] S.R. Lozada-Larson, K.R. Laughery, Do identical circumstances precede major and minor injuries? in: Rising to New Heights, Proceedings of the 31st Annual Meeting of the Human Factors Society, vol. 1, New York, 1987, pp. 200–204.
[10] G. Tinline, M.S. Wright, Further development of an audit technique for the evaluation and management of risk. Tasks 7 & 8. Final report C2278. A Study for the Health and Safety Executive, VROM & Norsk Hydro, London, 1993. Four Elements.
[11] A.M. Williamson, A. Feyer, D.R. Cairns, Industry differences in accident causation, Safety Sci. 24 (1) (1996) 1–12.
[12] H.S. Shannon, D.P. Manning, Differences between lost-time and non-lost-time industrial accidents, J. Occup. Accid. 2 (1980) 265–272.
[13] H.S. Kaplan, J.B. Battles, Q.S. Mercer, The hazard analysis action decision table, a means to limit unnecessary change in a medical event reporting system, in: the Proceedings of Enhancing Patient

Safety and Reducing Errors in Healthcare, Rancho Mirage, CA, November 8–10 1998, p. 1008.

[14] L. Wright, Towards an empirical test of the iceberg model, in: the Proceedings of the Seventh European Conference on Cognitive Science Approaches to Process Control, 2000.

[15] D.E. Embrey, D.A. Lucas, The nature of recovery from error, in: L.H.J. Goossens, (Ed.), Human Recovery, in: Proceedings of the COAST A1 Working Group on Risk Analysis and Human Error, 13 October 1987, Delft University of Technology, 1988.

[16] J.C. Flanagan, The critical incident technique, Psychol. Bull. 51 (4) (1954) 327–358.

[17] J.B. Davies, L. Wright, E. Courtney, H. Reid, Confidential incident reporting on the UK railways: The CIRAS system, Cognition, Technology and Work 2 (3) (2000) 117–125.