The analysis of UK railway accidents and incidents: a comparison of their causal patterns

By

Linda B. Wright

Submitted in fulfillment of the requirements for the Degree
of Doctor of Philosophy at the University of Strathclyde

Centre for Applied Social Psychology
Department of Psychology
University of Strathclyde

2002

Acknowledgements

This thesis could not have been completed without the support and co-operation of a number of people. They are:

John Davies, supervisor and friend, for all his support during my seven years at Strathclyde University.

Eleanor Courtney, for her friendship, discussions, encouragement and willingness to undertake inter-rater reliability trials.

Alastair Ross and Brendan Wallace for their discussions and help with inter-rater reliability trials.

All the CIRAS staff (too numerous to name individually) past and present for their hard work and belief in the principles of CIRAS.

ScotRail management for having the courage and foresight to pilot the CIRAS system in 1996.

Robert Plant, Operations Manager, ScotRail Railways Ltd for providing information and for helping me to understand the complexities of the railway industry.

The many Drivers who gave of their time to report near misses and be interviewed. Without them and their frank information this thesis could not have been completed.

To the CIRAS Trust and Steering Committee who granted permission to use the CIRAS data.

Last but not least, thanks to my husband, Tjerk for all his amazing support and my family (Mum, Dad and Hazel) for their belief that it would eventually be finished.

Abstract

An essential assumption for the usefulness of near miss reporting is the common cause hypothesis: the causal pathways of near misses are supposed to be similar to those of actual accidents (such as injuries and damages). The common cause hypothesis was originally proposed by Heinrich (1931) in his seminal book "Industrial Accident Prevention". Since then, the hypothesis has been alternately supported and rejected based on a confounded view of the interdependence of severity, frequency and causation. The evidence from various studies is examined and it is concluded that the hypothesis has not been properly tested. Thus this thesis tests the validity of the common cause hypothesis.

In order to develop the methodology to test the common cause hypothesis analytical work in the area of incident analysis and reporting was required. Thus this thesis also outlines the approaches to accident and incident analysis and makes several recommendations regarding the use of taxonomies and reporting systems. A reporting and analysis system (CIRAS) for the collection and analysis of near misses and unsafe acts and practices was developed and implemented for use in the UK railway industry. This reporting and analysis system formed the basis for the test of the common cause hypothesis.

Data used to empirically test the common cause hypothesis come from one company of the UK railway industry. Three types of data were used: incidents resulting in 'fatality & injury', 'damage' or 'near miss'. A total of 240 incidents were collected via management reports and a voluntary reporting system. All incidents were coded for causal factors according to the CIRAS (Confidential Incident Reporting and Analysis System) taxonomy. A total of 750 causal factors were assigned to the 240 incidents. Analysis was performed on a comparison of the proportion of codes occurring at all three consequence levels using Chi-square analysis. Results: The CIRAS taxonomy consists of 21 individual causal factors. Only three of these factors (knowledge based errors, training and procedures) were significantly different across the three severity levels. It is therefore concluded that this research provides qualified support for the common cause hypothesis.

Table of contents

List of Figures

List of Tables

List of Abbreviations

ASLEF: Association of Signallers, Locomotive Engineers and Firemen

ASRS: Aviation Safety Reporting System

ASSET: Assessment of Safety Significant Event Teams

AWS: Automatic Warning System

BASIS: British Airways Safety Information System

BR: British Rail

BTP: British Transport Police

CFIT: Controlled Flight into Terrain

CHIRP: Confidential Human Factors Reporting Programme

CIRAS: Confidential Incident Reporting and Analysis System

DRA: Driver's Reminder Appliance

DTM: Driver Team Manager

GEMS: Generic Error Modelling System

HMRI: Her Majesty's Railway Inspectorate

HPES: Human Performance Enhancement System

LOCA: Loss of Containment Accident

LTI: Lost Time Injury

MARS: Marine Accident Reporting System

MERS-TM: Medical Event Reporting System for Transfusion Medicine

MORT: Management Oversight and Risk Tree

NCCAN: National Center on Child Abuse and Neglect

PRISMA: Prevention and Recovery Information System for Monitoring and Analysis

P-Way: Permanent Way Staff

RMT: Rail Maritime Transport

RSC: Railway Safety Case

RSW: Restricted Work Case

SMIS: Safety Management Information System

SPAD: Signal Passed at Danger

S & SD: Safety and Standards Directorate

STEP: Sequentially Timed Events Plotting

TOC: Train Operating Company

TTC: Time to Collision

UoS: University of Strathclyde

WSF: Wrong Side Failure

WSP: Wheel Slip Protection

Chapter 1 The common cause hypothesis described and evaluated

1 Introduction

This chapter describes the common cause hypothesis with detailed reference to the work of Heinrich and others. The arguments that have been presented in the literature both for and against the common cause hypothesis are examined. It is proposed that the hypothesis is fundamentally misunderstood and that this has ramifications for studies, which have so far claimed to have empirically tested the hypothesis. Such studies are examined in light of the data collected. This chapter aims to evaluate the arguments and therefore demonstrate that research to date has failed to adequately test the common cause hypothesis and therefore the idea of a similar common causal pathway for near misses and more serious events can be neither accepted nor rejected at the present time.

1.1 Frequencies and causes of accidents and incidents

In 1931 in his seminal work "Industrial Accident Prevention" H. W. Heinrich claimed that major injury accidents, minor injury accidents and no injury accidents all had the same causes. Although he presented only a small amount of supporting evidence and despite many refutations, the idea of a common causal pathway for events resulting in injury and for those resulting in no adverse consequences appears to have become embedded in the accident prevention community.

It is the aim of this thesis to empirically test the common cause hypothesis (by determining whether the causal patterns of major accidents, serious injuries, property damage incidents and near misses are significantly different or not).

The data for this thesis come solely from the railway domain as this domain provides a rich source of both serious accidents and near misses. The implementation of the CIRAS system (Confidential Incident Reporting and Analysis System) enabled railway near misses to be collected systematically for the first time and hence provides an excellent comparison with actual incidents in this domain.

Testing the common cause hypothesis (i.e. empirically testing whether the various severity levels or the consequences [near misses, minor injuries, major injuries and fatalities] have similar underlying causal mechanisms) is not simply an academic exercise. This theory has been propounded to support the institution of near miss reporting systems, of justifying the time spent analysing minor events in the hopes that major events can be prevented and providing employees with forms and the time to fill them in. In fact, a lot of research has grown up around this theory, and that research may not be in a position to fulfill the promises it makes to decrease the number and severity of accidents by acting on near misses. Conversely, if the common cause hypothesis is validated, then high risk industries and medicine should make a greater effort to stimulate the reporting of near miss events, to spend time collecting and analysing near misses and to follow up with prevention measures that are aimed at decreasing the identified causes.

Chapter 1 provides an historical perspective of the ideas of Heinrich and outlines the logical arguments for and against the common cause hypothesis. This first chapter also describes and discusses the empirical research to date which claims to have tested the hypothesis. Chapter 2 outlines accident investigation processes and presents some detail on accident and incident causation theories. The main thrust of the chapter discusses the relationship between these theories and their implication for causal analysis and taxonomies. Such taxonomies form the basis for testing the common cause hypothesis (as pattern matching is required over a large number of incidents), and hence this chapter provides a clear methodology for testing the hypothesis. A list of the minimal requirements of any taxonomy for determining causation is presented. Chapter 3 then discusses the way in which incident event reports can be collected. A near miss reporting system was implemented on the UK railway industry using the findings and guidelines from Chapter 3. Chapter 4 presents and describes CIRAS (Confidential Incident Reporting and Analysis System) which is currently being used by the UK railway industry for the collection and analysis of near misses and other safety related incidents. The CIRAS taxonomy is fully described and compared to the minimal requirements of a taxonomy presented in the previous chapter. The CIRAS taxonomy is used to code all of the accidents and incidents collected as part of the hypothesis testing. Chapter 5 describes in detail the various types of data collected to test the common cause hypothesis – injury, damage and near miss incidents. All the data come from one railway company. Chapter 6 presents the results of the data analysis, while Chapter 7 presents a discussion of all the results from the thesis. Chapter 8 presents a summary of the results and the way forward for future research.

1.2 An Historical perspective

This section describes the background and thinking that led to the development of the common cause hypothesis.

1.2.2 Background

Heinrich published his seminal work "Industrial Accident Prevention" in 1931, which claimed that accident prevention at that time was aimed at the wrong events (major injuries) - with some revision of text this book survives today, having last been published in 1980 with co-authors Dan Petersen and Nestor Roos. Many of the ideas proposed in 1931 have survived the test of time and proved effective in reducing the number of industrial accidents; others, such as 'the common cause hypothesis' have gained the status of an urban myth (see Hale 2000) despite being omitted from the subsequent versions. In the final edition of 'Industrial Accident Prevention' (1980) the idea of a common causal pathway is rejected as a fundamental confusion between the causes of *severity* and the causes of *frequency*. This section will trace the development of Heinrich's original hypothesis and the subsequent critiques of this line of inquiry.

1.2.3 Action based on all accidents

Heinrich states that the ideas presented in 'Industrial Accident Prevention' stemmed from 17 years of industrial experience (mainly gained in the insurance industry). The aim of the book was to encourage industry to tackle not only those events which resulted in a major injury but also to pay attention and act upon the accidents which resulted in minor injuries, property damage or even no injury at all. The reasoning was that reducing the number of events *per se*, would also lead to a reduction in costs and an increase in productivity. Heinrich re-defined the term 'accident' to include even those events that did not result in an injury. He describes an accident as "an unforeseen, improper or non-planned occurrence....an accident need not necessarily produce an injury". (Heinrich, 1931 pp 93) This was very progressive thinking that required to be based on a scientific approach. In order to convince his audience of the value of this approach, Heinrich tackled the problem from a number of angles. These are described in the sections below.

1.2.4 The hidden costs of accidents

Firstly, he provided evidence that, due to hidden cost of accidents, accident prevention made sound business sense. He contended that the incidental cost of all industrial accidents was four times as great as that of compensation and medical bills. The text highlights a number of cases to demonstrate the calculation of hidden costs. Examples are given in Tables 1 and 2 below.

An example of Heinrich's calculations of the cost of accidents (Actual and Incidental):

Actual costs

Number of accidents	Description of accidents	Compensation and medical bills
3	Fractures & Contusions	$106
18	Rivet burns, cuts, bruises	$76
21	Falling materials	$15
30	Slips and falls	$12
		Total $209

Table 1: Actual costs of accidents during building erection job (Heinrich 1931, p 19)

Incidental costs

Time lost by injured employees (paid directly by employer)	$116
Time lost by other employees	$310
Time lost by superintendent & foremen	$78
Property damage	$158
Payment of forfeits for failure to complete job on time	$200
Portion of overhead cost loss during delay	$75
	Total $937

Table 2: Incidental costs of accidents during a building erection job (Heinrich 1931, p 19)

From the above tables it is clear that the incidental costs of accidents are often greater than the actual immediate cost of the accident (in terms of medical and compensation bills).

Hidden costs apply in the same way to major and minor injuries and Heinrich asserted that the ratio of 4:1 exists for major injuries, minor injuries and accidents with no injury.

The purpose of calculating the effects of hidden costs was not merely to show that all types of accidents cost more than was originally thought by the industrial community as Hale (2000) asserts, but also to demonstrate the relevance of tackling all accident types with the goal of prevention. Heinrich himself expresses it thus:

> *"This four to one ratio should prove to be a powerful stimulus to preventive action……to employers who are apathetic toward accident prevention, or who are mildly interested only for humanitarian reason. To employers who feel that they cannot afford to establish safety on an organized basis."* *(Heinrich, 1931 pp17-18)*

In 1966 Bird performed a similar study into the hidden cost of accidents and calculated the ratio to be of 6:1. This difference in ratio does nothing to detract from

Heinrich's work and in fact emphasises the fact that accident prevention (if effective) could easily save employers much money.

1.2.5 The interconnectedness of safety and productivity

In order to convince those who still failed to see the need for accident prevention, Heinrich proposed that safety and productivity were interconnected, and that the methods which were used to increase productivity, quality and control were the same methods that contributed to safety. Petersen (1989) believes that this expressed Heinrich's best thinking, but it was the one principle of Heinrich's that was most overlooked by safety professionals.

In line with his views on safety and productivity it was also Heinrich's contention that a safe factory was also more productive than an unsafe factory. A study of safety and production, performed by the Committee on Safety and production of the American Engineering Council and published in 1928 (Heinrich 1931, p 33) reinforced Heinrich's view of the interdependence of safety and productivity by concluding that a safe factory was eleven times more likely to be productive than an unsafe factory.

Thus Heinrich concluded that tackling all accidents as well as major injuries would lead to a reduction in costs and increased productivity.

1.2.6 Frequency of occurrence

In addition to reduced costs and increased productivity, Heinrich further established the need to aim prevention at all accident outcomes by providing evidence on the frequency of occurrence of the various accident types (severity levels) within an industrial context. Heinrich (1931) examined insurance claims for disabling injuries and suggested that for every mishap resulting in an injury there were many other similar incidents which resulted in no injury. He estimated that for every 330 potential injury accidents, 1 will result in a disabling injury, 29 in minor injuries and 300 will lead to no injury. Although Heinrich found some variation between different types of industry, he states that for all disabling injuries, the figures average to those shown in the triangle model below in Figure 1.

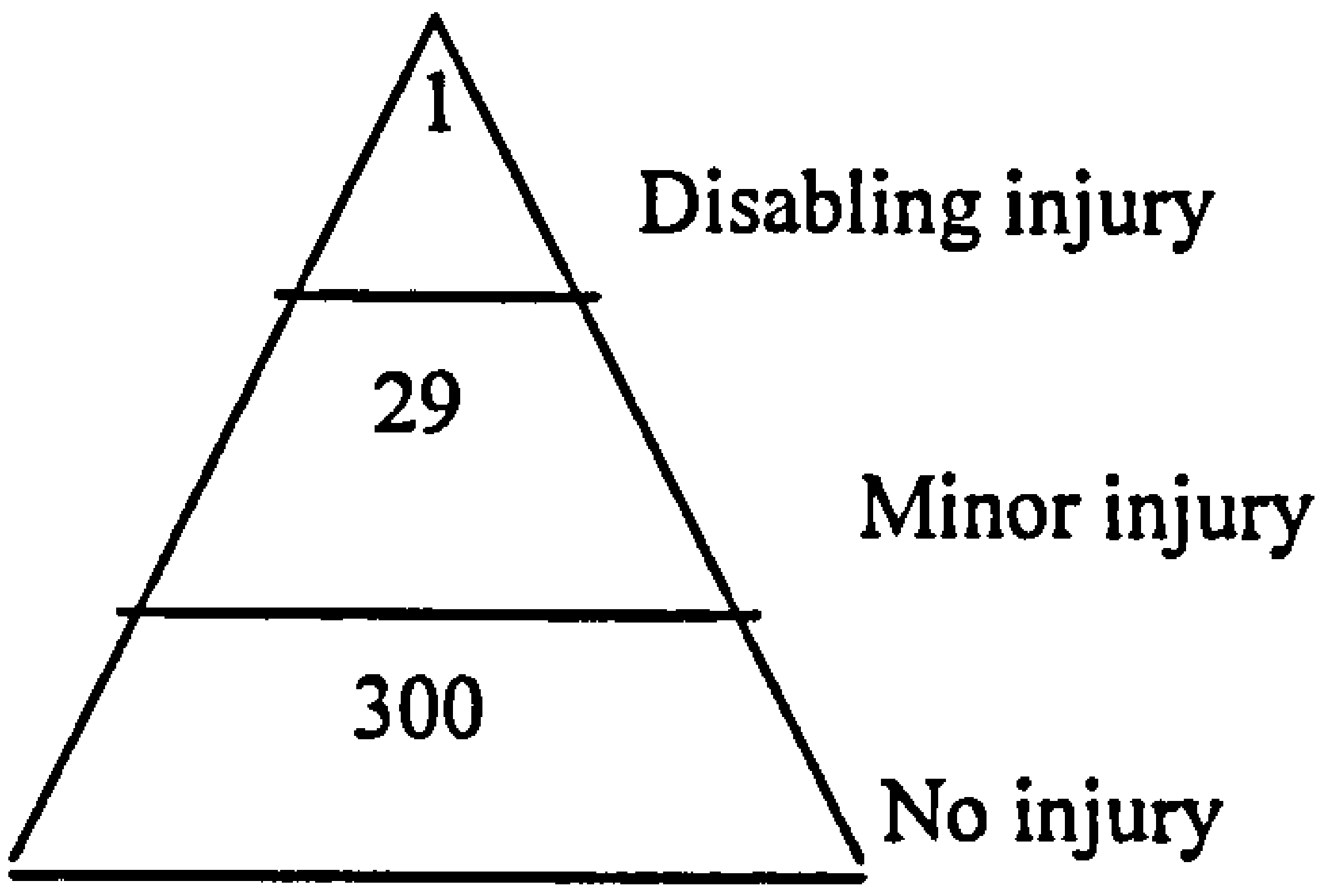


Figure 1: Heinrich's ratio triangle

Other similar studies examining the relationship between major and minor events have also been carried out (e.g. Bird 1966 & 1976; Skiba 1985). Bird (1976) analysed 1, 753, 498 accident reports from 297 companies while working for the Insurance Company of North America. Unlike Heinrich's study, which was based on "accidents of the same kind and involving the same person", Bird widened the scope to include different types of industry and accident. The results of Bird's analysis are shown in the triangle model below:

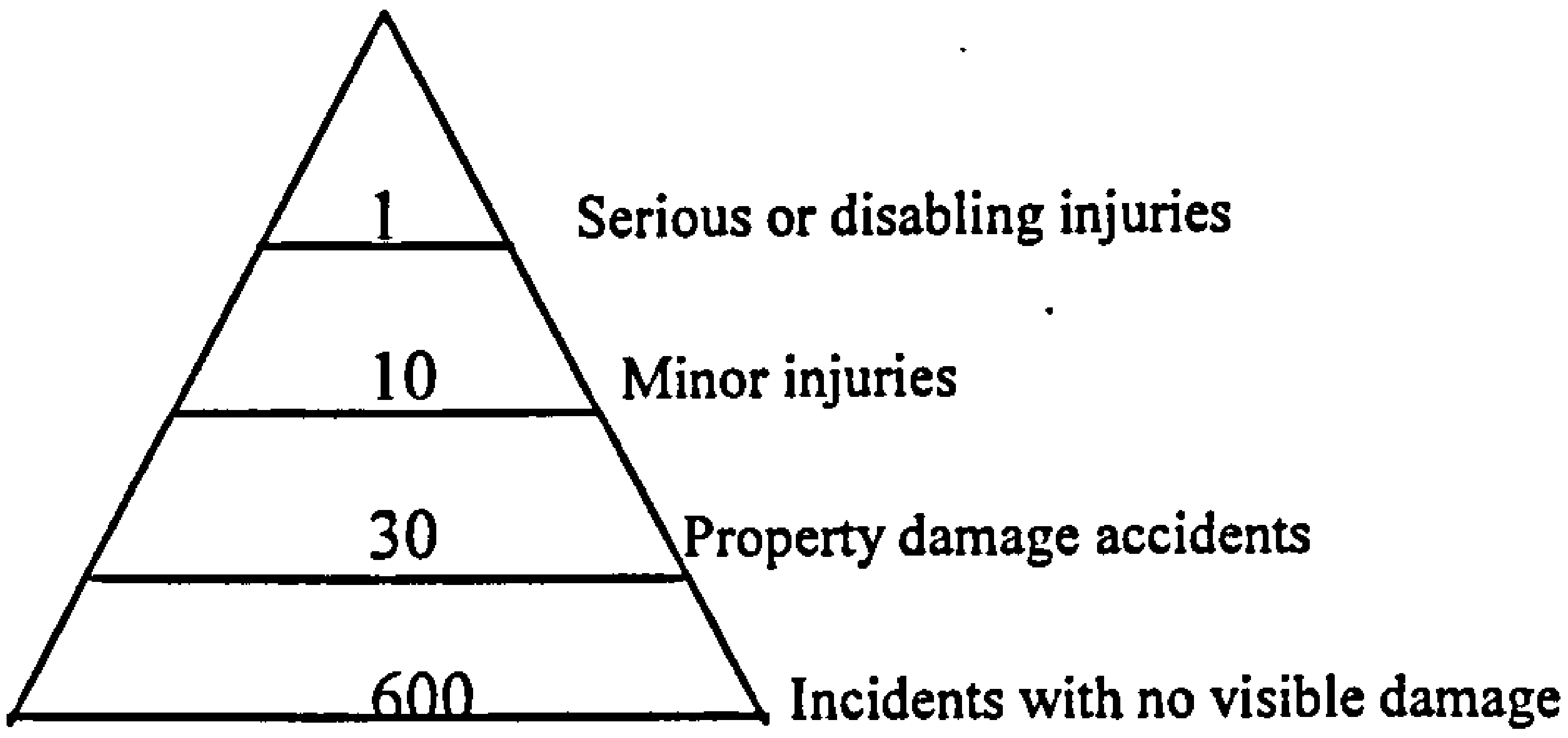


Figure 2: The Bird accident ratio study

Bird's results as shown above in Figure 2, represent an average for all industries. Swain (1974) found the same ratio as Bird. Skiba (1985) has a similar pyramid of work accidents taken from West-German accident records. This provides a set of ratios which suggest that for every fatal accident there are an estimated 70,000 near accidents. It should be noted that the cited accident figures from different data sets provide incompatible ratios. This makes sense, as each industry and work activity will have its own set of ratios of injuries, property damage incidents and near misses.

This can be clearly illustrated by the fact that railway maintenance workers have the unfortunate dichotomy to be at either end of the triangle i.e. due to the nature of their work they are more likely to be involved in either a near miss or a fatality. However, it is unclear from any of the ratio studies, if these ratios are stable over time and can be replicated year on year for the industries or if it is simply an amalgamation of all data collected. While Hyos and Zimolong (1988) state that an empirical ratio does not imply any causal relationship between the different levels of the triangle, they do conclude that reduction or elimination of near misses will improve system safety and reduce the number of injuries. Such a conclusion appears to be based on the fact that the ratios are taken to be stable rather than descriptive.

However, it cannot be determined from the above ratio studies (or from Heinrich's study) whether the established ratios are stable and therefore predictive (i.e. as the number of minor injuries increases so too does the number of property damage accidents and serious or disabling injuries etc., and vice versa a decrease in minor injuries should lead to a decrease in serious or disabling types of accident), or if it is simply a description of the number of incidents which occurred in a given time period. This distinction is fundamental to all future arguments: the claims suggesting that a decrease in near misses will automatically result in a proportional decrease in serious injuries would require a *stable* ratio.

Like Heinrich, Bird concluded that his study presented a strong case for aiming prevention at the events at lower levels in the triangle in an attempt to reduce the number of serious accidents and disabling injuries. He suggested that accident prevention initiatives were failing to target resources and actions appropriately. He suggested that if the more frequently occurring incidents were tackled (i.e. the near misses or incidents with no visible damage) then this would reduce the number of serious accidents (major and minor injuries) which occur. This of course only holds true if the ratio relationship is stable.

Bird himself states:

> *"The 1-10-30-600 relationships in the ratio would seem to indicate quite clearly how foolish it is to direct our total effort at the relatively few events terminating in serious or disabling injury when there are 630 property damage or no-loss incidents occurring that provide a much larger basis for more effective control of total accident losses."* *(Bird, 1976, pp167)*

The idea of tackling less serious incidents (i.e. those at lower levels of the triangle) in order to reduce the more serious accidents, is therefore based on the assumption that a stable ratio between minor and major incidents exists. Where the ratios are stable and not merely descriptive, a causal connection is implied, so that reducing incidents at one level of the iceberg must produce a similar effect at other levels of the iceberg.

This and Heinrich's assertions in 1931 has led to the assumption that the causal mechanisms of minor and major incidents are the same or similar (van der Schaaf 1992; Swain 1974; Ferry 1988).

However, Groeneweg (1998, p 51) provides evidence culled from a number of sources (traffic conflict studies and industry) that the pyramid ratios not only differ within and between companies and type of work but that the ratios themselves are not stable. He cites statistics collected and analysed by the Oil Industry International Exploration and Production Forum which show only a weak relationship between the number of fatalities and the number of Lost Time Injuries (LTIs) and Restricted Work Cases (RWCs)[1]. The correlation between LTIs and fatalities was .656 and between RWCs and fatalities .467. Between RCWs and LTIs the correlation was .269. This suggests that the number of LTIs does to some extent predict the number of fatalities, but the RWCs are not a reliable predictor of either fatalities or LTIs. Groeneweg provides a number of explanations for this finding, including the following:

1. Data on RWCs were not collected by all the companies returning data to the co-ordinator of this statistical study;
2. The various companies had different criteria relating to a LTI and RCW;

---

[1] Restricted work case is defined as injury at work that does not lead to absence the day after the accident because of alternative job assignment.

3. RWCs are not necessarily the result of an accident or incident but are mainly operational decisions.

Groeneweg uses these findings and the differences in data collection protocols to illustrate the point that accident statistics are not a good measure of the safety of a company. However, the question for this thesis is not how well the number of near misses will predict the number of fatalities or minor accidents, but how well the causal patterns of near misses predict the causal pattern of minor injuries, serious accidents and fatalities.

1.2.7 The value of causal analysis

In order to take effective remedial or preventive action one must know and understand why and how an event has occurred. Heinrich reviewed more than 50,000 accident reports and proposed that the causal analysis of accident generation within most industries was deficient – the causes attributed to the accidents were not the causes of the accidents themselves but the causes of the resulting injuries. These included "causes" such as 'slips and falls', 'falling objects', 'struck by or against', etc. Heinrich provides the following to elucidate his point:

> *"Falls of persons for example, is (sic) obviously a cause of an injury, but it is not the cause of an accident. A person is injured because of a fall. The fall itself is*

the accident, and it is the cause of the accident that must be known in order to effect a remedy." (Heinrich, 1931, pp 39)

It was apparent to Heinrich that the type of injury sustained was being confused with the type of accident which resulted in such injuries: this is still often the case in industry and research today. While Heinrich attested that causal analysis was carried out poorly and in a confused manner, he did not suggest that collecting data on the types of accident (such as slips, trips and falls) should be discontinued for they provided useful information. Indeed such classification systems remain in use today. However, the type of accident should not be confused with or used as a replacement for classification and identification of the causes.

In order to clarify the 'cause' of an accident, Heinrich took his definition of accident causation from the U.S. Bureau of Labor Statistics:

> *"That the accident should be charged to that condition or circumstance which would have prevented the accident; but if there be more than one such condition or circumstance, then to the one most easily preventable." (Heinrich, 1931, pp 42)*

There is much to contend with in this definition: not least that the idea of multiple causation is acknowledged and then bypassed in favour of finding the most easily preventable cause.

Heinrich proposed a classification system, which he claimed, would be adequate for all accidents and all industries. He divided accident causes into supervisory and physical. Beneath each of these superordinate categories, he provided a number of classifications such as faulty instruction, unsafe practice and poor housekeeping. Each of these classifications then had a number of detailed sub-categories from which to select. It can be seen that Heinrich's taxonomy fails to clearly identify 'supervisory' causes including such as 'faulty instruction' which is a supervisory problem along with 'inability of employee', 'mentally unfit' and 'lack of concentration' all of which are not necessarily under the control or management of the supervisor. Furthermore 'poor housekeeping' is included as a 'physical' cause despite the fact that this is clearly under the control of supervisors. However, despite these weaknesses in Heinrich's taxonomy it is nevertheless advanced for its time. Heinrich provided a set of explanations and examples for the terminology he chose. Accidents were caused either by Supervisory or Physical failures. A representation of this early taxonomy is shown in Table 3 below.

ACCIDENT CAUSES	
Supervisory	**Physical**
Faulty Instruction: a. None b. Not enforced c. Incomplete d. Erroneous	Physical hazards a. Ineffectively guarded b. Unguarded
Inability of employee a. Inexperience b. Unskilled c. Ignorant d. Poor Judgement	Poor housekeeping a. Improperly piled or stored equipment b. Congestion
Poor discipline a. Disobedience of rules b. Interference by others c. Fooling	Defective equipment a. Miscellaneous materials and equipment b. Tools c. Machines
Lack of concentration a. Attention distracted b. Inattention	Unsafe building condition a. Fire protection b. exits c. Floors d. Openings e. Misc.
Unsafe practice a. Chance taking b. Short cuts c. Haste	Improper working conditions a. Ventilation b. Sanitation c. Light
Mentally unfit a. Sluggish or fatigued b. Violent temper c. Excitability	Improper planning a. Layout of operations b. Layout of machinery c. Unsafe processes
Physically unfit a. Defective b. Fatigued c. Weak	Improper dress or apparel a. No goggles, gloves, masks, etc b. Unsuitable – long sleeves, high heels, defective, etc

Table 3: Heinrich's 1931 taxonomy of accident causes (p 46)

1.2.8 The common cause hypothesis

Having used a variety of methods to convince his intended audience of the benefits of analysing all types of accident, Heinrich then turned to actual data. Using the taxonomy described in section 1.2.7 above, Heinrich applied his causal analysis to 50,000 accidents. He does not provide data on how this was accomplished, nor any great detail on the use of the taxonomy. There are no reliability studies in the use of the classification scheme and it appears from the text that Heinrich in fact assigned all the causes himself. Further, Heinrich himself states that only one immediate cause was assigned for each accident – given our understanding of accident causation today, the data are somewhat limited. These difficulties notwithstanding, Heinrich's results appeared to point convincingly to the fact that similar causes exist for no injury accidents and for minor and major injury accidents. The data used are from the superordinate categories only for the "supervisory" causes, while only mechanical hazards have been included from the "physical" causes. This is not discussed or explained in the text. A graphical representation of Heinrich's results is provided in Figure 3 below.

Despite the unanswered questions regarding the data shown below, Heinrich's data provide compelling evidence for the similarity in causal patterns between major injuries, minor injuries and no injury accidents. In the 1931 edition, he himself claims the following:

> *"The truth of this deduction is strikingly demonstrated through the discovery – as recorded on the chart – that the frequency of major injuries varies directly with the frequency of no-injury or potential-injury accidents. This result would be expected to follow, as the predominant causes of no-injury accidents are, in average cases, identical with the predominant causes of major injuries –* ***and incidentally of minor injuries as well.****" (Heinrich, 1931 pp 90) [bold author's own]*

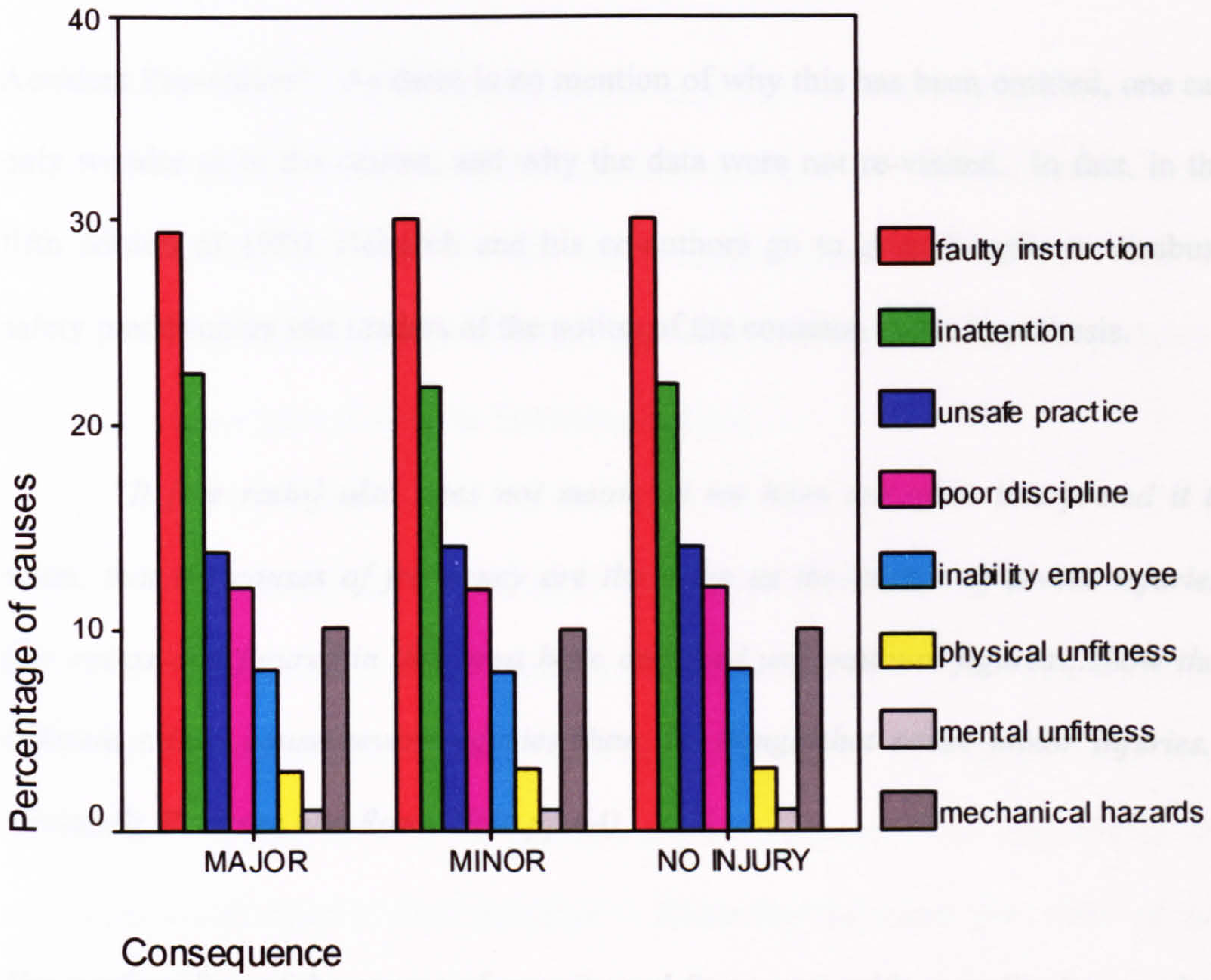


Figure 3: Heinrich's data on causes of major, minor and no injury accidents

The findings displayed on the graph above would suggest that the immediate cause of major, minor and no injury accidents are the same – of course this method of causal analysis takes a simple, linear view of accident causation and fails to deal with multiple causes. The message for industry was clear – collect, analyse and effectively tackle all minor events and the number of serious accidents should decrease. It is surprising therefore, that all references to this data analysis and the similar cause hypothesis are absent from all subsequent versions of "Industrial

Accident Prevention". As there is no mention of why this has been omitted, one can only wonder as to the causes, and why the data were not re-visited. In fact, in the fifth edition of 1980, Heinrich and his co-authors go to great lengths to disabuse safety practitioners and readers of the notion of the common cause hypothesis.

> *"It [the ratio] also does not mean, as we have too often interpreted it to mean, that the causes of frequency are the same as the causes of severe injuries. Our ratios and figures in this area have confused us...national figures...show that different things cause severe injuries than the things that cause minor injuries." (Heinrich, Petersen and Roos, 1980 pp.64)*

The confounding of the causes of severity and frequency, with causality has resulted in a common causal mechanism being rejected on the basis that reducing frequency has not reduced severity by a similar percentage. This does not logically follow. A similar cause hypothesis does not logically follow from the original data on ratios, but from the data on the causal analysis performed by Heinrich. The ratio type models that are used to describe the frequency of occurrence of various severity levels may be merely descriptive and say nothing about cause or the type of relationship which may exist between frequency of accidents and the levels of severity of accidents. Such ratio models have not even been shown to be stable. From the many data sets studied, it can be seen that the ratios for different work types and countries vary enormously. Thus any work in this area is immediately bound by whatever ratio is chosen and the confines of that particular ratio. Heinrich, Petersen and Roos (1980) cite various figures that do not fit with Heinrich's original ratio model to prove that there is no common causal pathway, and support the argument with the following figures:

> *"We can readily see that the types of accidents that result in temporary total disablilities are different from the types of accident that result in permanent partial disabilities or fatalities.....handling materials accounts for 25% of all temporary total disabilities and 21% of all permanent partial injuries, but only 6% of all permanent total injuries and fatalities. Electricity accounts for 13% of all permanent totals and fatalities but accounts for a negligible percentage of temporary totals and permanent partials. These percentages would not differ if the causes of frequency and severity were the same. They are not the same. There are different sets of circumstances surrounding severity. Thus, if we want to control serious injuries we should try to predict where they will happen. Today we can often do just that. " (Heinrich, Petersen and Roos, 1980 pp 64 - 65 & Petersen, 1979 pp 19 - 20)*

So, Heinrich and his co-authors repudiate a similar cause hypothesis on the basis that the figures for different industries do not fit with the original triangle model. However, earlier in the same chapter it is argued quite successfully that the original figures were averages of different people and different accident types, and that different industries are likely to have different ratios.

> *"It does not mean, for instance, that there would be the same ratio for an office worker and a steel erector. It might mean they could be averaged to this (or a similar ratio); but certainly neither of these extremes would fit the ratio." (Heinrich, Petersen and Roos, 1980 pp 64)*

It would appear that Heinrich is prepared to undermine the original model in order to suggest that there is no proof for a similar cause hypothesis. It is unfortunate that he has failed to realise that the existence of a ratio of severity and frequency neither adds nor detracts from a similar cause hypothesis. A rather tautologous kind of logic is at work here: Heinrich is confusing severity and frequency ratios with causal mechanisms.

Heinrich further attacks the similar cause hypothesis by suggesting that a reduction in frequency does not correspond to exactly the same reduction in severity. Again, he is confusing the causes of frequency and severity, with the causal mechanisms of incidents.

"Statistics show that we have only been partially successful in reducing severity by attacking frequency. In the last 40 years National Safety Council figures show an 80 percent reduction in the frequency rate. During that period the same source shows only a 72 percent reduction in the severity rate, a 67 percent reduction in the fatal and permanent total rate, and a 63 percent reduction in the permanent partial disability rate." (Heinrich, Petersen and Roos, 1980 pp 65 & Petersen 1979 pp 23)

Given the ratios provided by Heinrich in 1931, and given that an 80% reduction in frequency corresponds to a 72% reduction in severity for national figures (i.e. all types of work and activity), this does not look like limited success. Rather than suggesting that the different severity levels are not related, these figures appear to provide support for the hypothesis that severity and frequency are related in some way: with a reduction in one leading to a reduction in the other (although not of exactly the same proportions). Heinrich suggested that the above figures did not provide enough proof that the various severity levels are related causally. However, the interdependency of severity and frequency are something different from underlying causal mechanisms. Rather than suggesting that underlying causal mechanisms are dis-similar, Heinrich's figures suggest that severity and frequency are not related in the rather straightforward way first conceived of in the triangle models.

The confusion between severity, frequency and causality has in part arisen as the terminology used by Heinrich and subsequent authors is unclear and open to interpretation. In Heinrich's accident triangle, 'frequency' is how often a given incident occurs, while 'severity' relates to the consequences of the incident. The relationship between the frequency of a given incident and the severity level of incidents has given rise to a confusion between the causes of severity and frequency (i.e. the reason why a given incident results in a severity level a number of times) and the underlying causes of incidents. The causes of frequency and severity relate to the probability of a particular task resulting in an incident and the probability of that incident causing a certain consequence. For example, in the railway industry, the frequency of an accident at an AOCL (automatic open level crossing locally controlled) depends on how often vehicles and trains meet at the crossing, and on how often something goes wrong when this meeting occurs (such as the vehicle driver trying to beat the train, the warning lights not working, or the driver falling asleep etc.). The severity of the accident resulting from the failure of the train and the vehicle to maintain separation depends upon a number of factors which are independent of the cause of the accident, (such as the type of vehicle, the speed of the vehicle and the train, the weight of the train, the number of passengers etc.). This example is provided to demonstrate that the interdependence of frequency and severity has no bearing on the underlying causal mechanisms of accidents and incidents.

1.3 Development of Heinrich's principles

Heinrich provided safety practitioners and researchers of today with plenty of material for study. He provided strong arguments for collecting and analysing minor deviations and near misses. Various authors have expanded upon his triangle models and theorised about the relationship which exists in such models. This section now critically examines the perspectives following on from Heinrich. A summary of his points is presented in order to compare his ideas with those of the present day.

Heinrich's Fundamental Principles for Accident Prevention:

1. Principles of safety and production are alike
2. Executive interest is necessary for effective accident prevention [This is the first time that such a principle appears in literature – it is widely acknowledged in industry today.]
3. Cause analysis is necessary for all accident types (major, minor and no-injury accidents). [In the first edition of "Industrial Accident Prevention" only, does he make claims regarding a common causal pathway for all accident types.]
4. For every major injury accident there are 29 minor injury accidents and 300 no injury accidents.
5. Selection and application of remedy should be based on the causes
6. Corrective practice should be enforced (sic)

Heinrich (1951) pp20

As this thesis is concerned only with item 3 of the above list, this section on present day theory will concentrate mainly on the similar cause hypothesis and triangle or iceberg models.

1.3.1 Current views of Heinrich's theory

Echoes of Heinrich's reasoning can be found in the work of van der Schaaf (1991) who states that the collection and analysis of near miss data have the following purposes:

1. to gain qualitative insight into how (small) failures or errors develop into near misses
2. to arrive at a statistically reliable quantitative insight into the occurrence of factors or combination of factors giving rise to incidents
3. To maintain a certain level of alertness to danger, especially when the rates of actual injuries and other accidents are already low within an organisation.

Reason (1997) fully supports these points while describing problems in the data used to validate the various 'iceberg' models. Van der Schaaf points out that near misses occur more frequently than actual accidents; the ratio studies have clearly demonstrated this point. Whether the order of magnitude is 300:1 as Petersen (1989) suggests or as defined by the ratios described in 1.2, the fact remains that more opportunities exist to register near misses and perform analysis as a means to identifying problem areas. Van der Schaaf (1991) defines the relationships between accidents, near misses and behavioural acts as follows: near misses are sandwiched between actual, but rare, accidents above and a large number of behavioural acts (errors and recoveries) below. Incidents at the top level of the pyramid are more visible than those below. It is assumed that the various levels within the pyramid are directly related. The pyramid represents a continuum of events from the bottom to the top of the triangle i.e. that near misses and accidents are directly related in terms of cause, and therefore, an overlapping set of root causes should be identifiable between the various outcome levels of the model. It is also assumed that incident analysis will attempt to progress as far to the bottom of the triangle as possible, and therefore that accidents will have a greater number of identified root causes than either near misses or behavioural acts. The model is outlined in Figure 4 below.

Sutherland, Makin and Cox (2000) are also proponents of tackling incidents at lower levels of the triangle in order to reduce the more serious accidents. They recommend a proactive behaviour-based approach and the encouragement of reporting near accidents. They cite figures from Tye and Pearson (1974/5) cited in HSE (1992) as providing evidence for tackling near misses rather than fatalities and lost time accidents. They acknowledge that near accidents are excellent predictors of future accidents (Tarrants, 1980).

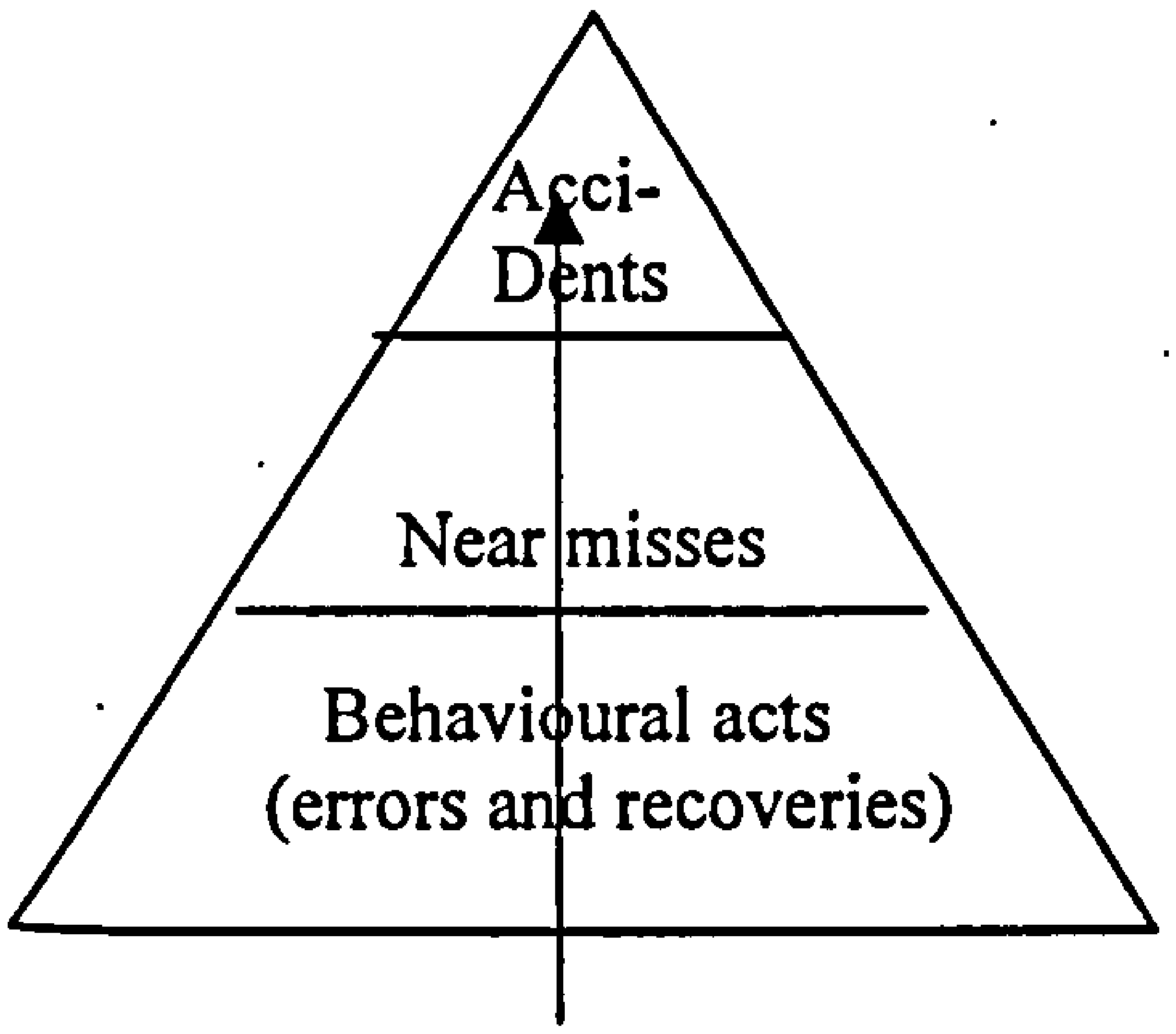


Figure 4: Qualitative iceberg model of the relationship between accidents, near misses and behavioural acts (after van der Schaaf, 1991)

Despite the criticisms by Groeneweg in relation to the stability of the ratio discussed earlier in this chapter, he asserts that the idea behind the pyramid models is valid thus:

> *"fatalities are built on the same foundations as Lost Time Injuries and unsafe acts and should not be treated separately". (Groeneweg, 1998, pp 44)*

Thus it is apparent that while the idea of a stable ratio is in doubt, the propagation of incidents as a continuum from unsafe acts to accidents appears to have become accepted. The acceptance of such a continuum does not depend upon a stable ratio, but would suggest that similar causal patterns would be found in accidents, incidents and near misses. However, other researchers, notably Petersen and Hale have criticised this approach to the propagation of accidents.

Petersen's (1979 & 1980) objections to the common cause hypothesis are based on the fact that a decrease in frequency of accidents has not led to a similar decrease in the severity of accidents. He states that if the causes of severity and frequency were the same then the decrease in the percentages of different severity levels would not differ. Arguments have already been provided, discussing the confusion between a ratio relationship and a causal relationship.

This thesis aims to test the proper interpretation of the common cause hypothesis in one domain that involves a variety of tasks within the railway industry. The limitations of studying one domain are acknowledged and will be expanded upon in Chapter 4, which discusses the data collection. The thesis is not concerned with triangle models per se, nor in frequency of exposure but in determining if there are similar causal mechanisms or pathways underlying incidents which occur at different levels of the triangle.

Hale (2000) does not fully support the same cause hypothesis, and is surprised by the vigour with which it has been adopted. He states that intuitively one can grasp that any given accident is the culmination of a process that can be stopped at a number of points. In his opinion the Heinrich and Bird studies suggest some overlap in causes, but also a considerable degree of difference. In order to substantiate the differences Hale (op cit) cites a number of studies and also includes reference to Heinrich et al's

5th edition of Industrial Accident Prevention which concluded that decreasing frequency had only a partial effect on decreasing severity. Hale also discusses the results of a number of other researchers which apparently refute the common cause hypothesis. These shall be discussed in section 1.4. At this point, it is necessary only to state that the majority of these studies are not actually testing the common cause hypothesis using causal analysis, but using inappropriate data resulting in confusing results and continued debate. Hale and Hale (1972) in an extensive literature review make reference to the common cause hypothesis and at that time highlighted the lack of empirical testing of the hypothesis. Despite their assertions that the hypothesis required rigorous testing before being fully accepted into the literature and safety culture, the hypothesis remains untested (at least in any meaningful and reliable way). Hale accepts a number of these studies as having tested the common cause hypothesis despite their obvious lack of using causal data and cites further accident figures regarding aggregated ratio data as further proof of a different cause scenario. Unfortunately, at this point Hale also has confused the ratio debate with the causal debate. Similarity of causation does not rely on, nor does it bear any relation to the frequency of occurrence.

Hale does not completely dispute the possibility of common causal mechanisms. However, he does present a well-balanced argument for limiting the applicability of the hypothesis. Tinline and Wright (1993) analysed loss of containment accidents and lost time injuries in chemical plants and showed little or no correlation between the occurrence of the two within or between companies. Lost time injuries therefore could not be used as an indication of major hazard safety. Similarly, in unpublished work Hale found that the lost time injuries at a Dutch chemical plant almost all occurred during non-process-related activities, such as walking around the plant and using stairways. He suggests then that it is crucial to know whether practices and conditions could result in more serious outcomes. This approach is based on Heinrich's original work which took the occurrence of a major injury and then looked at the possible ways in which near miss incidents and similar sequences of event resulted in a consequence less severe. Intuitively it can be understood that any given accident is a culmination of a process that could or can be stopped at any point. The process may be halted by way of barriers and/ or recovery mechanisms. A number of authors (Kontogiannis 1999, van der Schaaf and Kanse 2000) are currently investigating the phenomenon of recovery and how it impacts positively in the prevention of more serious accidents. Hale therefore suggests that such factors will also have an effect on causal pathways. The important issue then, is to determine which near misses or low-level events have the potential to result in more serious consequences. Hale is open to the idea that certain low level near miss events have the same causal pathways as more serious events. In terms of accident prevention he is interested in whether it can be determined from early accident sequences which events have the potential to become more serious, which events will be recovered before damage occurs and which events will only ever lead to minor consequences. He suggests that only the potentially serious near misses, and not an amalgamation of all near misses are likely to have similar causal pathways as the more serious consequence events.

He therefore suggests a different type of pyramid, one that takes this process of ability to stop the event at given points into account. Thus if the pyramid is defined in terms of the deviation sequence, it will be a pyramid for a given type of event. Hale's new pyramid to link major and minor accidents is shown below in Figure 5. For Hale, major injuries, minor injuries and property damage accidents are actually the top of a different set of pyramids, each of which has its own set of deviations, losses of control and a range of damage resulting from its own damage process. However, he posits that such a pyramid would only hold true for similar sequences and therefore similar work types. Hale does not expound upon what would constitute a similar sequence. Before this hypothesis could be tested it would be necessary to identify what constitutes a similar sequence or type of activity. e.g. similar duties such as train driving, signalling, maintenance work or similarities in terms of types of event, such as Signal Passed at Danger, Controlled Flight Into Terrain (CFIT) or Loss of Containment Accident (LOCA).

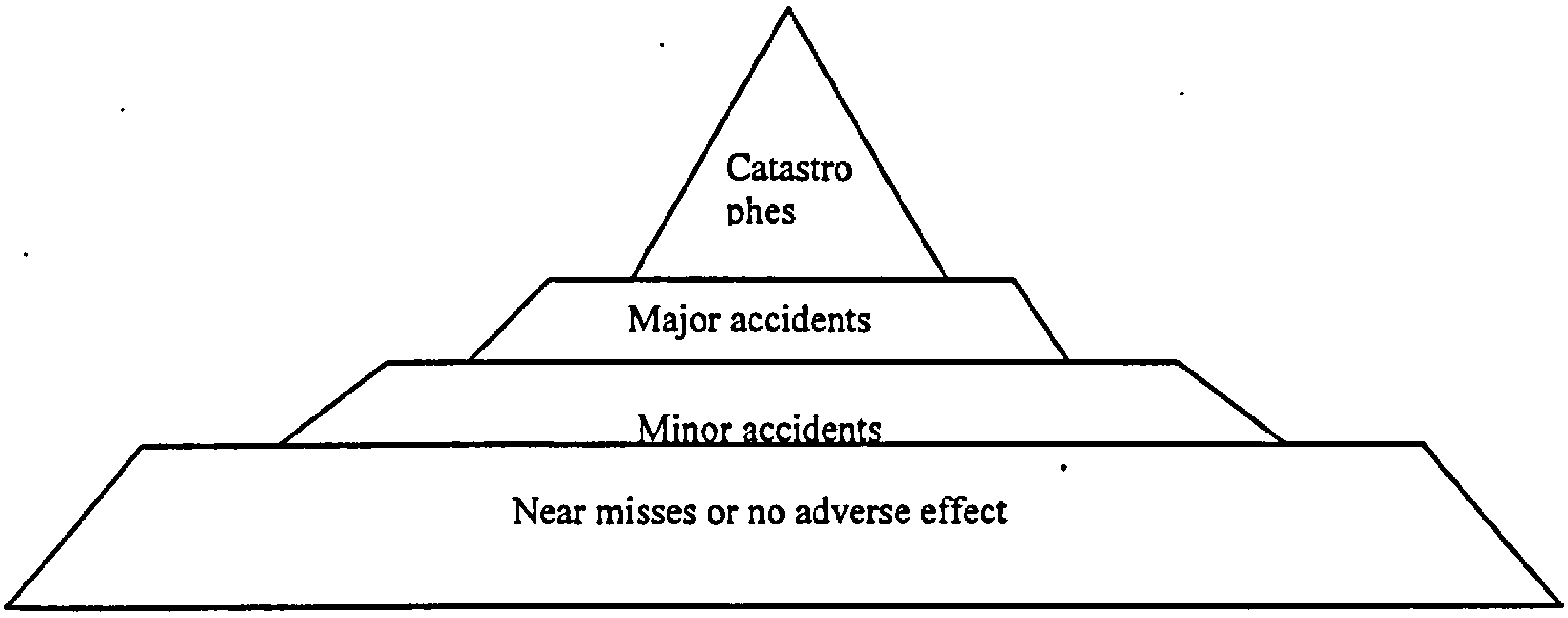


Figure 5: A new pyramid to link major and minor accidents (after Hale 2000)

As can be seen from Figure 5, only a small number of events will progress onto the most severe consequence level, while others have the potential to become major and minor incidents. The majority of events remain at the level of near miss via timely detection, recovery or barriers. While Hale's pyramid model refines the original work done by Heinrich and Bird, it has not as yet been tested. The similar cause hypothesis for the iceberg model has also not been fully or properly tested.

However, despite the objections to the common cause hypothesis Peterson and Hale still believe that the collection and analysis of small scale events will lead to more efficient and safe operations especially where the minor events have the potential for being serious. Where a minor event has the potential for being serious it would suggest that a continuum of the event would lead to a serious injury and that therefore at least some underlying and contributing causes were the same. It also suggests that some form of recovery has prevented a minor event from progressing to the stage where it becomes a fully-fledged accident. This of course has implications not only for the common cause model (which would need to be adapted to take into account that only some near misses have the potential for becoming more serious), but also for incident reporting schemes.

A problem for reporting schemes therefore lies in how to decide which reports have the potential to become more serious: should the reporter be depended upon, or an internal/ external expert. In the CIRAS reporting system (described in chapter 3) it was decided to have the reporter rate the potential consequences of the incident being reported. This however resulted in one third of reported incidents having the potential for a fatality. There are two possible reasons for this. Firstly, the reporters may want their reports to be taken seriously and so exaggerate the potential consequences. Secondly, reporters might discuss the potential consequences in terms of a worst case scenario (e.g. all possible failures occurring together). A severity and frequency rating was developed for use by the administrators of the system, but this proved to be unreliable and subject to individual bias and understanding.

If the potential for more serious consequences is to become a factor in determining similar causal patterns and events upon which action should be taken, clear guidance needs to designed and presented. For example, in the railway industry the potential for a more serious event is based ultimately upon the location of the incident rather the causal factors. A SPAD (Signal Passed at Danger) can occur at many different locations – where there are no conflicting movements possible due to track layout or catch points the incident is recorded as having no potential for damage or fatality. However, the causes of such a SPAD may be the same as the causes of another SPAD at a different location which is not fitted with catch points. This is of course a complicated issue and requires clarity regarding what constitutes a potentially more serious event.

1.4 Common cause hypothesis testing – a critical evaluation of current research

A review of the literature reveals only a handful of studies that claim to have attempted to test the common cause hypothesis. Furthermore, the majority of these have taken as a starting point a confounded view of the iceberg model; that is, they have confused the two dimensions of severity and frequency with causality, leading to conclusions which are not supported by the type of data used in the research. A further problem of comparing different studies has been highlighted by Lortie et al (1999). They examined the variables retained and descriptors used by accident studies published during the years 1986 – 1995. They found 72 papers from which 37 of the articles concentrated on the analysis of accident circumstances. However, despite such a large number of studies, they conclude that accident data are not systematically collected, have no similar frame of reference and appear to have evolved in an idiosyncratic manner. Such differences between accident data result in difficulties in comparing between studies. Further they conclude that the analysis of data focussed on the accident circumstances and rarely centred on data relating to activity, work site, causes or incidents. Such a focus is far from that favoured in accident causation models. Obviously, this prevents a valid testing of, or conclusions to be drawn about the common cause hypothesis. In section 1.4.2 a number of studies are discussed with reference to the type of data used and the relationship of the conclusions to the data.

1.4.1 A confounded view

From the discussion already presented in this first chapter, it is apparent that frequency, severity and causal mechanism have become inextricably linked (Heinrch op cit; Petersen op cit). It appears that researchers have confused the causes of severity and frequency with 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 use 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 types. Causality has no bearing on the ratio relationship propounded by the iceberg model and vice versa.

Despite the many studies that have ratified the ratio model (e.g. Bird 1966, Skiba 1985), there has been no attempt to link the ratios to causal mechanisms, in terms of the underlying causal pathways related to incidents and accidents. The triangles discussed previously suggest that exposure i.e. frequency and severity are related, but do not make any claims about causality in a broader sense. Subsequently, using ratio models to either support or reject a similar cause mechanism or pathway is based on a confounded view of the iceberg / triangle models and is not valid.

1.4.2 Empirical evidence evaluated

This section describes a number of studies that have made conclusions about the common cause hypothesis. Despite an extensive literature search only a handful of studies could be found which claimed to have empirically tested the hypothesis. As will be demonstrated in the following discussion, very few of the studies reported have actually tested the common cause hypothesis, although they have been used subsequently by various authors (e.g. Hale 2000) to refute the hypothesis. Some have concentrated on the ratio relationship between near misses, injuries and more serious consequences, an assumption which rests upon a confounded view of the common cause hypothesis. Others, have understood the distinction between frequency and severity versus causality, but have failed to use data that is appropriate to causes. Traffic conflict studies especially have provided mainly anecdotal evidence that can be used to support or refute the hypothesis – these will be discussed in section 1.4.2.3. Finally, those studies that have used data that is appropriate to test the hypothesis have failed to describe the data adequately and provide transparency as to how statistics were applied. As a result, no firm conclusions can be reached regarding the validity and applicability of the common cause hypothesis. The studies are presented below according to the type of data used i.e. inappropriate and appropriate data. It is important that this distinction is made and demonstrated by recourse to the data as many objections to the common cause hypothesis are based upon inappropriately conducted studies.

1.4.2.1 Ratio data studies

Two studies are described here which used frequency data and the invalidation of a ratio relationship to conclude that the common cause hypothesis was not valid.

In a 1998 study by Saloniemi & Oksanen the relationship between fatal accidents in the workplace and macrostructures of production life in Finland were examined. The central question of the research was to determine how business cycles are related to fatal accidents. Accordingly, manufacturing and construction industries were chosen. However, a subsidiary question was also asked of the data: "are the causal chains behind fatal and non-fatal accidents identical or different?". Re-stated, the authors suggest this was an empirical test of the common-cause theory.

The data used in this study consisted of the following information for both manufacturing and construction: fatality rate and number of fatal accidents; accident frequency; number of wage earners; working hours and unemployment rate. For the construction industry only, the number of cubic metres under construction was also included. Using these data, the authors conclude that there is no support for a pro-cyclic relationship between accidents and economic activity. However, one surprise finding was that declining production (in cubic metres under construction) was followed by an increase in the fatality rate. The authors used this finding to suggest that their study does not corroborate the common cause hypothesis. Specifically they state "the results are consistent with earlier findings which emphasise the specific nature of fatal accidents, their own distinctive logic and their own causes". What this actually shows is that the iceberg model (in terms of ratios of minor to major accidents) is not validated for the construction industry under specific conditions (i.e. decline in metres under construction). This says nothing about similar or dis-similar causes of either fatalities or accidents. It would appear that the authors have confused causal pathways with the ratio produced by the iceberg model.

As the construction industry does not fit the view of the triangle shaped ratio model, the authors conclude that this is evidence of differential causal mechanisms. Such a confounded view which inter-relates causal mechanisms and ratio data leads to conclusions regarding underlying causes that cannot possibly come from data on number of accidents and variables relating to economics (i.e. number of wage earners and number of hours worked). Saloniemi and Oksanen state that the risk of accidents is twice as high among non-skilled workers as among carpenters. However, carpenters have distinctly more accidents requiring one month or more sick leave (they go on to suggest that this may be due to labour market resources i.e. carpenters are more likely to be able to afford longer sick periods). So having deduced that the ratios for these different types of workers do not fit the original pattern displayed in the iceberg models, they conclude that different causes must underlie non-serious, serious and fatal accidents. Interestingly, they also conclude their paper with the following sentence, "the statistical material used in this study is limited and could hardly have provided any conclusive solution either to the problem of the causation of different kinds of accidents or...". Indeed, the statistical material used for this study cannot provide any insight into the causal patterns involved in major and minor accidents. The data can only point to discrepancies or problems associated with the original ratio idea proposed by the iceberg style studies, and not to underlying causal mechanisms.

Tinline & Wright (1993) analysed loss of containment accidents and lost time incidents in chemical plants and showed there was no correlation between the occurrence of the two within or between companies. Again, such research is based on the ratio assumption of the iceberg model and not on the actual causes of the incidents. The two are not related: from the frequency data presented there can be no claims made about causal mechanisms.

1.4.2.2. Causal data studies

This section describes research which has attempted to test the hypothesis using data based on causes of accidents.

A study by Williamson, Feyer & Cairns (1996) compared the causes of fatalities (i.e. one level of the iceberg model) across all industrial groups in Australia. In order to test the various causal mechanisms of fatalities, up to three precursor events and contributing factors were assigned to each fatality, using a classification system. Analysis of the data showed remarkable similarities across all industry types with two thirds of all fatalities being described by only four combinations of precursor events and contributing factors. Although this study researched only the top level of the Heinrich/ Bird triangle models (fatalities), it does provide evidence for similar causal mechanisms at that level of the pyramid. Unfortunately the study was not expanded to include other levels of the pyramid, and so no conclusions can be made about similar causal pathways from the bottom to the top of the iceberg i.e. between near misses, serious accidents and fatalities.

Shannon and Manning (1980) concluded that lost time accidents and non-lost time accidents resulted from very different causes. This conclusion was based on the comparison of 2, 428 accidents reported by employees of the Ford Motor Company at Halewood, England. The accident causes were assigned by the employees and then a sample was randomly validated by Safety Engineers. In 42 of 56 cases, the Safety Engineers agreed with the employee's assessment of cause. The actual causes are not described in the paper nor are they compared. The authors have compared the number of events which preceded the accident with the distribution of accidents in lost time and non-lost time categories. In fact they state that the number of accidents and differences in absolute numbers of preceding events leading to lost time and non-lost time accidents are sufficiently large for normal statistical tests to be unnecessary. It is clear that the authors have not compared causes and so the conclusions regarding the common cause hypothesis are not valid.

Lozada-Larsen & Laughery (1987) lend their support to the common causation hypothesis, citing their study of 7, 131 employee accidents as definitive proof that identical patterns of causation occur across the various severity levels. The paper clearly differentiates between major and minor injury accidents, although near misses are not included in the analyses. The paper clearly states that the identical causation hypothesis is tested and then goes on to describe the data used. These were 6, 435 minor injuries which were compared with 408 major injuries: all were extracted from 5 years of accident data extracted from the records held by Shell Oils, Deer Park, Texas. The data appear to have been coded by the company and there are no inter-rater reliability details.

The first test of the hypothesis examined the frequencies of various activities that were taking place prior to the accident occurring. Such activities included: assembling/ disassembling equipment, handling, moving materials, and manually operating valve, among others. The results are presented in Table 4 below:

Activity prior to accident	**Major injury accidents**	**Minor injury accidents**
Assembling, disassembling equipment	29%	28%
Handling – moving materials	26%	21%
Manually operating valve	10%	11%
In transit within work area	10%	9%
Changing elevations	6%	6%
Housekeeping	5%	5%

Table 4: Most frequent prior activities for major and minor accidents after Lozada-Larsen & Laughery (1987)

The authors state that the above results "indicate qualified support for the identical causation hypothesis". However, while such data provides an interesting insight into the type of activities taking place prior to both major and minor injury accidents it is irrelevant as it does not shed light on causal mechanisms. For example "handling materials" is not the cause of an injury, but the way in which the materials were handled, or the procedures relating to the way in which the materials should be handled and other circumstances surrounding the handling of materials is the cause. The authors have failed to understand the difference between the cause of an event and what activity was taking place when the event occurred.

This first 'test' of the hypothesis was followed by comparing "accident events" and "incident events" for both major and minor injury accidents. Again the authors conclude the results (which show a similar distribution of percentages across the two groups provides support for the identical causation hypothesis). However, once again these data are not valid to test the hypothesis. These data included variables such as "impact with object", "material noncontainment", "caught between", "cut by" and others. Heinrich is clear that the event resulting in an injury is not the cause of the accident. "Impact with object" is the cause of the injury sustained, but is not the cause of the accident or event that caused the impact. Such data are not causal explanations but descriptions of the injury event. While this study may provide useful information for the industry concerned on the type of accidents occurring during which normal, routine activities it is not a valid test of the common cause hypothesis.

Salminen et al (1992) improve upon previous studies which claim to have tested the similar cause hypothesis. They have clearly identified the problems inherent in previous studies thus "the hypotheses have been examined by comparing the types of accidents which do not necessarily describe the causation of accidents". While this study shows promise and progresses beyond identifying and comparing types of accidents, it provides a rather loose classification of "accident factors" and still failed to compare the combinations of causes present in serious and fatal accidents. There was also no comparison of fatalities and serious accidents with near misses. However, if the hypothesis is valid the same set of causes should be present between the three different levels of the pyramid and so this should not have any effect on the results.

Salminen et al (1992) found that their study did not support a similar cause hypothesis for serious accidents and fatalities. The study focuses on the differences (or similarities) in 99 serious accidents, of which 20 were fatalities; hence 20 fatalities were compared with 79 non-fatal accidents. Accident factors (assumed to be the causal mechanisms) were assigned to each accident, from a possible 14 factors; with an average of 12 assigned to each accident. The consensus between two independent raters was 79%. A Kolmogorov-Smirnov test was performed between the two groups which revealed statistically significant differences ($p<0.05$) in the distribution of the 14 factors between the two groups. Thus the conclusion was that the common cause hypothesis was disproved. However, this only demonstrates that more or less causes were involved in the different outcomes of fatal and non-fatal accidents: the authors failed to perform any analysis on the relative proportions of factors assigned to the accidents. The table below provides a summary of the categories and the percentages of each factor associated with the different level of event. It can be seen that the non-fatal and fatal accidents differ on only a few of the accident factors.

Accident factor	**Fatal accidents (n=273) %**	**Non-fatal accidents (n=916) %**
Perceiving and recognising danger	7	6
Work habit	11	17
Other factors related to victim	10	11
Co-workers	7	4
Foremen or customers	1	3
Organisation factors	11	7
Insufficient planning	5	6
Technical shortcomings	8	9
Workspace	8	8
Tools	14	15
Handling of materials	11	8
Clothing	3	1
Weather	3	1
Other	1	3

Table 5: Distribution of accident factors in fatal and non-fatal accidents (after Salminen et al (1992)

There are a number of criticisms that are relevant to this research. Firstly, the 14 accident factors used for the study appear to be an amalgamation of contributing causal factors (e.g. work habits, insufficient planning), situations outwith the control of employees (e.g. weather, workspace), and subjective measures (e.g. perceiving and recognising danger). As such it is unclear whether the authors are clearly identifying and testing causal patterns. Secondly, it is disappointing that no analysis was performed at the level of individual causes to determine where the differences between serious accidents and fatalities may lie. Thirdly, the criteria for inclusion in the category 'serious accident' perhaps requires to be more specific – the authors state that "all accidents...were classified as serious including the situations in which the possibility of a serious accident was obvious". Thus non-serious consequences were included in the 'serious' category if they had the potential to become serious: there is no information on how this distinction was made. This approach would appear to be amalgamating the bottom layers of the triangle and comparing them against only the fatalities. As these accidents were obtained from all serious occupational accidents in Southern Finland it is unclear whether certain industries are over-represented in either of the samples, and indeed the authors do not discuss the issue of the types of industry represented in the fatalities or serious accident categories.

These objections notwithstanding, this research is obviously important as the authors have clearly differentiated between the similar cause hypothesis and the ratio relationship posited by the iceberg studies.

Kaplan, Battles & Mercer (1998) present data extracted from MERS-TM (the Medical Event Reporting System for Tranfusion Medicine). This is a confidential reporting system for failures that occur during the process of obtaining and transfusing blood and blood products. Incidents are categorised according to a taxonomy that includes Technical, Human and Organisational failures. Within each of these major categories there are a number of more detailed codes. The data presented however, only relate to the higher-order categories. Reported events are divided into actual events and near misses. Severity levels are applied to both types of event: for actual events the severity level is a reflection of the actual consequence, and for near misses it is a reflection of the potential severity had the incident not been recovered, halted or detected. A number of permutations of causes by event category and severity level were compared using chi-square analysis. A rather conservative significance level of 0.001 was selected. A total of 371 reported events were analysed as part of the study. Results are summarised in Table 6 below.

Severity level (actual and potential)	**Near misses compared with actual events**
Level 1 (fatality or serious outcome)	N.S
Level 2 (Minor and transient injury)	N.S
Level 3 (No ill effects)	P<0.001

Table 6: Results of MERS-TM analysis (adapted from Kaplan et al, 1998)

These results appear to be based on all categories of cause being combined, as no deeper analysis was performed to determine where the differences in terms of Technical, Human and Organisational category occurred. Kaplan et al (op cit) appear to have performed a comparison of near misses and actual events without taking severity level (actual and potential) into consideration and this was reported as being significant although no statistics or details were provided. There was no attempt to determine if differences occurred at the various levels of cause comprising the Technical, Human and Organisational categories. The authors conclude that the data support the conclusion that actual events and near misses share the same causal factors when analysing the more serious (actual or potential) events. This seems to lend some support for Hale's theory (section 1.3.1) which states that the more potentially serious near misses are the most important in terms of prevention. This study is a major step closer to testing the common cause hypothesis. It must however remain inconclusive as the data are not fully described, there is no deeper analysis performed and the results are based on a rather conservative estimate of significance (0.001).

Wright (2000) compared the causes of near misses and unsafe acts as reported to the railway confidential reporting system called CIRAS (Confidential Incident Reporting and Analysis System). Causes were assigned according to the CIRAS taxonomy which is fully described in chapter 3. The results showed some differences in terms of Technical and Organisational causes between the two groups. It was concluded that the two groups studied and the results were insufficient to provide any conclusive evidence to either support or reject the common cause hypothesis.

1.4.2.3 Traffic conflict studies

Traffic accident rates and conflict studies have been used by a number of authors (Hyden 1987, Hale 2000, Brown 1991) to both support and call the common cause hypothesis into doubt. These studies however, have not focussed on testing the hypothesis per se and the data reported does not support the conclusions which have been drawn from them. In terms of road traffic accidents and near misses, Hyden (1987) suggested that near misses and actual crashes could be viewed as a continuum, as the processes involved in near misses were often similar to those resulting in a crash. However, there is little more than anecdotal evidence to support this in traffic conflict studies. van der Horst (1991) has established that a TTC (time to collision) of 1.5s distinguishes well between normal and critical behaviour of road users. Hale (2000) takes this as evidence of different causes between accidents and near misses. However, van der Horst makes no claims about causality, he merely states that studying near misses (in terms of traffic conflicts) may help shed light on the differences between near misses and actual accidents. The results of the traffic conflict observations were not compared to actual accident data, and hence no conclusions can be drawn as regards the common cause hypothesis. Brown (1991) did compare observations of traffic with actual accident data for the same location. However, the research was not specifically focussed on the common cause hypothesis and so presents only a limited discussion of some differences found between near miss occurences and actual crashes. For example a total of 665 conflicts observed over a three day period were compared to 378 actual crashes in the previous five year period. The frequency of occurrence of different types of conflict (e.g. conflicts at ramp, right hand turn at ramp) were compared to the frequency of occurrence of actual accidents that occurred during similar types of movement. Accident records showed 142 left hand turn crashes while the researchers observed very few conflicts (actual number not provided). The author concludes that there are differences between near misses and actual crashes based on this and other similar findings. However, once again, this is the use of frequency data and not causal data. Traffic conflict studies could present a good opportunity to test the validity of the common cause hypothesis, but to date the data have not been used for this purpose, and so any conclusions based on such studies must remain tentative at best.

Traffic conflict studies may have implications for the railway industry. Similar studies could be performed using data from train data recorders, which could provide the industry with information on the critical time for when a SPAD will occur and cannot be avoided. Such information could also be used to determine successful versus non-successful recoveries from missing signals or failing to react to signals.

1.4.3 A summary of the empirical common cause studies

It is clear from section 1.4 above that the common cause hypothesis is yet to be comprehensively tested. For the sake of clarity the following table provides an overview of the various papers discussed above, and presents the type of data used, whether the authors confused causality and the existence of an observable ratio between minor and major incidents and the conclusions reached.

Reference	**Type of data used**	**Claims re causality (yes or no)**	**Confounded view of the iceberg model**	**Conclusions**
Saloniemi & Oksanen (1998)	Frequency data for major and minor accidents in manufacturing and construction industries.	Yes	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.
Lozada-Larsen & Laughery (1997).	Frequency data Comparison of the type of activity taking place prior to accident occurring e.g. during manufacture.	Yes	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.
Tineline & Wright, (1993)	Frequency data comparing the occurrence of lost time	Yes	Yes. Confuses ratio data for causal data.	As ratios not in agreement with original iceberg theory, concluded that different causal

	accidents and loss of containment accidents			mechanisms present between major and minor
Williamson, Feyer & Cairns (1996)	Data classified according to taxonomy of causes. Only for fatalities, all industrial types.	Yes, only for fatalities.	No.	Similar causes for all fatalities.
Shannon & Manning (1980)	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	Yes.	As differences observed between the number of accident events assigned to the consequence (lost time or non-lost time) concluded different causes.
Salminen, Saari, Saarela & Rasanen (1992)	Finnish accident research model of 14 factors applied to 20 fatalities and 79 serious accidents.	Yes	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.
Kaplan, Battles & Mercer (1998)	Causal factors according to the classification of MERS-TM: Technical, Human and Organisational factors for near misses and actual events.	Yes	No	Authors' state this data supports the common cause hypothesis – but only under certain severity conditions. Conservative significance level chosen.
Wright (2000)	Causal factors according to CIRAS: Technical, Proximal,	Yes	No	Results based on preliminary analysis of data and comparison graphically. Differences noted between Technical

	Intermediate and Distal for near misses and unsafe acts.			and Organisational causes between the near misses and unsafe acts.

Table 7: A review of the literature on causality claims

From the table above it can be seen that an adequate test of the common cause hypothesis has not yet been performed. Further, there are obvious limitations in a number of the studies (in terms of data used and a clear understanding of the hypothesis). An adequate test of the common cause hypothesis involves comparing the relative contributions of the various causes to the incident consequences.

1.5 Summary and conclusions

This chapter has provided an overview of the genesis of the common cause hypothesis and discussed the controversy that still exists today. Despite the lack of definitive research many authors have strong opinions on the relevance of a common causal pathway for near misses, injuries and more serious accidents. A discussion of the supporting evidence was provided. The chapter concluded with a review of the current literature available on testing the common cause hypothesis.

It was demonstrated that the studies so far have failed to grasp the difference between causal relationships and ratio relationships and were basing results on a confounded view of the iceberg model. Furthermore, those studies which have recognised this problem have failed to test the hypothesis in a satisfactory way: i.e. comparing causal mechanisms of the various levels of the iceberg.

The problems of the comparison of accident and incident data were described. It is apparent that industry, researchers and safety practitioners have failed to collect data of a standard or similar type. This has resulted in a variety of data types (accident events, precursor events, preceding event, injury type, injury cause and event causes) that exclude the possibility of comparing data sets across industries. Common data standards between industries or at least across industry types (such as the railway industry, the petro-chemical industry) would lead to meaningful comparisons of data sets across or between industries. This lack of standardisation of terminology and data collection also leads to confusion in the interpretation of results in general, and in particular has led to a lack of understanding regarding what constitutes a valid and reliable test of the common cause hypothesis. Researchers and those in industry should clearly define the terms they are using, the data they have collected and fully report the basis for their analysis.

It is concluded that the common cause hypothesis has not as yet been adequately tested and that such a test to be described in this thesis is both relevant and important. A valid test of the common cause hypothesis involves an analytical approach to the problem (as described in this chapter), followed by developing a method by which to test the hypothesis, the development of a system to collect the data, appropriate analysis of the data and finally a comparison of the relative contributions of causal factors (the proportions of the various causes) that result in different levels of consequence. The method is described in the following chapters.

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 completely different patterns of causes, then industry has been concentrating on levels of severity (near misses, small failures) which will 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 and to ensure that causal analysis techniques become more widespread.

Chapter 2 Accident analysis and the use of taxonomies

2 Event analysis - purposes and methods

This chapter begins with an examination of the purposes of accident and incident analysis, and discusses the differences between two main approaches: the traditional, judicial approach and the organisational learning approach to event analysis. No matter the purpose of the event investigation and analysis, analysts and investigators will hold a view of the nature of causality and the way in which an accident or incident is propagated. Therefore this chapter also briefly introduces various accident models before moving onto the methods which are used to analyse events. The role of taxonomic approaches is discussed and the minimal requirements of such a classification system for accidents and incidents is proposed. Finally, the problems associated with accident and incident investigation and taxonomic classification systems are discussed.

2.1 The goals of accident/ incident analysis

For many, the goals of accident analysis are clear: the accident is investigated in order to determine the causes, and then, prevention measures are derived which should prevent future accidents from occurring (Heinrich 1931, 1980; Bird 1976; Ferry 1988). However, prevention and reduction of risk is not always the aim of accident analysis. There are many reasons for companies and investigators to perform accident analysis: these may be to satisfy regulators; to determine responsibility, blame and liability; to satisfy insurers that all that could be done to prevent the incident had been done; to understand why the event happened or to make improvements based on the causes of incidents i.e. to learn from past failures. Benner (2000) identified 44 different reasons for performing accident analysis based on discussions with investigators, such as to support civil litigation, establish liability, settle compensation claims, find and prosecute violations, determine causes and prevent accidents. These differing and often conflicting goals result in different types and levels of analysis being performed. The two main goals of accident analysis, namely apportioning blame and determining causes in order to learn from them are discussed in the following sections.

2.1.2 Traditional approaches

In a traditional judicial approach (Hale, 1997) the objectives of event analysis are to point the finger of blame at one individual or a group of individuals who failed to perform their duties correctly, or who were negligent. Such an approach is epitomised in the public inquiries which feature in the British response to disasters and are becoming more prevalent when disasters occur on the UK railway infrastructure (e.g. Fennell Inquiry, 1988; Hidden Inquiry,1989). The purpose of accident investigation and analysis in this model is to uncover the chain of events leading to the incident and to determine where the duty of care was breached. It is assumed that all parties involved in the accident had a duty of care (as individuals or companies) to protect customers or users of the system against injury and damage. Therefore when injury or damage does occur, this duty has not been fulfilled by one or all of the parties involved, and so the judicial approach will determine where blame and responsibility (and therefore compensation) lies. Until recent times, the focus of inquiries tended to be aimed at finding responsible individuals, although for the last few decades emphasis has also been placed on the organisational failings (Reason 1997) as well as the role played by individuals in the accident. The aims and shortcomings of the judicial approach are highlighted by Svenson et al (1999) by reference to a Swedish court case (Lundberg, 1992) in which three patients undergoing kidney dialysis died. The result of the court case was that one individual, the senior nurse, was found responsible for the deaths and subsequently given a suspended prison sentence. Svenson et al (op cit) contend that within the legal framework, the simplest form of accident analysis was performed (accident resulted from single cause). The resulting action to prevent a similar incident from occurring in the above scenario was to remove the nurse. However, in a complex socio-technical system such as the renal unit where there were design failings (emergency stop and alarms on the dialysis machines were the same), they suggest that the finding of one cause does not adequately explain how the accident happened. In order to test this hypothesis, Svenson and his colleagues later compared the legal anaylsis with analyses of psychology experts and engineering experts. The results showed that the legal analysis was the least comprehensive. The analyses performed by both psychologists and engineers found that 74% and 63% respectively attributed responsibility to agents other than the nurse. Further the psychology experts found a mean of 9 errors while the engineering experts found a mean of 9.95 errors (in both sets technical and human factors errors were found). This far exceeds the number of errors found by the original judicial inquiry; which found one individual responsible and only her errors. Svenson et al (op cit) use these findings to argue that the legal framework can be inefficient or even counter-productive in promoting improvements to the safety of complex integrated systems. Baram (1997) also suggests that regulatory and legal approaches result in defensive management attitudes and little organisational learning due to decreased motivation for employees to make reports (due to shame, blame and liability) and the company being concerned only with penalties and liabilities rather than learning from incidents.

Another aspect of the traditional approach is that it tackles each accident as a separate entity and does not require a structured approach to accident analysis in terms of a methodology. Each accident is analysed individually and there may be no commonalities between different inquiries in terms of method of analysis and investigation. The aim is not to aid the organisation in better safety performance, rather the objectives are to determine the specifics of the particular case, to apportion blame and sometimes also to aid prevention of a future recurrence of the same accident.

Such an approach may well identify some causes of a particular accident, but whether it aids organisational learning remains a matter of conjecture. Rosenthal (1997) suggests that there is no easy answer to the question "Is the goal of using accident event analysis to gather information aimed at improving process safety regulatory and safety management systems compatible with the legal system's goal of establishing possible corporate responsibility as a basis for corporate punishment and victim compensation?" (Hale, Wilpert and Frietag (eds) 1997 pp179).

While this thesis is not concerned with the traditional judicial approach per se, data from formal company inquiries have been used, and it is therefore important to understand the different goals between an inquiry approach and that of an organisational learning approach. That is not to say that organisational learning cannot take place following a formal inquiry, however formal inquiries are based on individual cases and fulfill a multitude of purposes: apportioning blame; satisfying the regulators and preventing the same accident from happening again.

2.1.3 Organisational learning and safety improvement

The alternative approach to event analysis is that of learning from failures and making improvements. The focus of event analysis is not on finding individuals to blame but identifying why humans, equipment or systems failed and taking measures to prevent a future recurrence of events caused by the same factors. Wagenaar and van der Schrier (1997) are clear that the only meaningful objective of accident analysis is to prevent future accidents. Given that the aim is prevention, then it is clear that the analysis method should systematically uncover the systemic causes of the accident and result ultimately in a list of aspects of the operation that can be changed. Wagenaar and van der Schrier (op cit) clearly identify the range of causes that should be included in the analysis thus: "...causes that existed not only at the specific time of the accident but had existed for a longer period, that remain present after the accident, and that have the power to contribute in a significant manner to future accidents" i.e. both active and latent failures (Reason 1990).

Van der Schaaf (1991) outlines three purposes behind the organisational learning approach:

1. Modelling and learning about new failure modes

Within this objective, van der Schaaf suggests that the specific goal is to "identify likely factors or system elements" qualitatively in order to identify precursors of future accidents. From this qualitative insight there are two ways to reduce the likelihood of future incidents and accidents. Firstly by eliminating error-inducing factors, and secondly by introducing or strengthening recovery-promoting factors. Recovery factors are those factors which contribute to a full or partial recovery of a situation after an error or failure has occurred, and so prevent or reduce the negative consequences of the error or failure. In order for successful recovery actions to be strengthened, they must be identifiable and part of the modelling process (see Van der Schaaf 1988; van der Schaaf and Kanse 2000 for a detailed explanation). Further, he suggests that another type of qualitative insight may be the identification of new or unusual combinations of factors which resulted in an incident or accident. As new types of events are always possible (often what we fail to predict does happen), this means that the set of possible accident scenarios is increased. However, a completely pre-coded system is unlikely to be able to include such events. This is not unrealistic – no taxonomic system can possibly list all possible failure modes. This type of analysis then supplements the pre-coded system.

2. Monitoring the frequency of known failures

This objective requires pattern recognition across a number of minor accidents or of more major accidents into pre-coded categories. Using such a system it is possible to determine which factors or combinations of factors occur most frequently. This would then form the basis for a rational decision making process on where to concentrate resources in an attempt to improve safety in the most efficient way. If such a system were also linked to actions undertaken, then the effectiveness of interventions could also be monitored i.e. has the intervention led to a decrease in the particular factors they were aimed at?

3. Maintain alertness and motivate staff

Descriptions of past events can be used for training and discussion purposes, and for safety promotion campaigns.

These three objectives can be realised using a variety of methods for collecting and analysing data. Van der Schaaf (op cit) suggests seven steps or modules which provide the building blocks for designing a near miss management system (see Chapter 3). However, the same steps can be used to describe the modules necessary for a successful event reporting and analysis system that aims to process actual accidents and incidents and not only near misses.

The problems and choices to be made concerning the detection of relevant events and the type of data to select are discussed elsewhere (see for example Hale, Wilpert & Freitag eds. 1997). While these problems are acknowledged, they are not central to a discussion on the method selected to analyse events. However, these issues will be re-visited in Chapter 3 (Near miss reporting systems) and Chapter 4 (The CIRAS system).

2.2 Accident causation models

The analysis, classification and coding of event data is usually based upon a model of accident causation and / or a theory of behaviour. This section will discuss a number of theories of accident causation and relate them to different analysis methods.

Accident causation models have evolved from the most basic one-cause model, to the complex interaction of the organisational accident. Both accident causation theories and accident investigation methodologies have implications for how deeply an event will be analysed, for how far back in time the analysis proceeds, when the analysis stops (a stopping rule) and for the number and type of corrective actions that will be suggested. Many accident causation models exist. Benner (1985) examined and rated 14 accident models and 17 accident investigation methodologies used by U.S. government agencies. The models identified ranged from the single cause model to the events process model and the energy flow model. Benner (op cit) was of the opinion that the basic underlying philosophy of accident causation would affect the type of analysis and investigation technique adopted. Although he rated the different types of accident models he did not fully describe the models. Further the rating of the models was subjective. Benner states that this could be improved upon and should be performed in order to recommend the most appropriate models and ensure standardisation in accident analysis.

Hoyos and Zimolong (1988) provide an overview of analytical frameworks for accident causation and the corresponding investigation methods. They identify four main approaches: systems theory; sequential models; integrative models; and the epidemiological approach. Subsumed within each of these four approaches lie numerous accident causation models and classification techniques: a brief description of the salient points of the various accident models will be highlighted here and a selection of the methodologies will be introduced.

According to systems theory an accident results from a defect in parts of a system or the interaction between them. Within the systems theory approach there are a variety of sub-theories e.g. systems ergonomics emphasises the man-machine-environment taxonomy (e.g. Fault Tree Analysis, Meister, 1971), while the industrial engineering systems view focuses on the control of the production process from a management view (e.g. MORT [Management Oversight and Risk Tree] , Johnson 1975, Adams 1976). In Fault Tree Analysis the accident or failure itself is the beginning of the analysis. A logic tree is then constructed to show the relationship that may occur between faults in the system.

Sequential models view accidents as a sequence or chain of events. Heinrich's 1959 Domino Theory is a good example of such a model of accident causation. An accident will be prevented if one of the dominoes can be removed and thus prevent the rest from falling and therefore completing the sequence. Sequential analysis is performed in order to determine which of the five dominoes – ancestry, social environment, fault of person, unsafe act or unsafe condition – need to be removed. A description of the accident sequence is the starting point for identifying the situation that can explain why the incident happened.

The integrative models are a combination of the human and machine constituents. These are combined to form an action taking into account information processing. The action performed may result in adaptation resulting in system homeostasis or in maladaptation resulting in a disruption to the system. Such system deviations are the beginning of the accident. Kjellen (1984) called this the deviation concept. Accidents are seen as a combination of the factors of the man-machine system i.e. technical, organisation, social, economic and individual.

A study by Andersson and Menckel (1995) identified eleven conceptually different models. The models ranged from the simplistic to the complicated and all had in common three main dimensions of time, level and direction. These theories were related to the epidemiological approach and the interaction between host, agent and environment, and mainly focussed on the issue of primary, secondary and tertiary prevention methodologies.

Whatever philosophical approach is used to define accident causation, it is clear that within all of these models, causation can range from the simplistic to the complex. Benner (2000) proposes that accident perceptions with their associated theories include the following: the single event perception and one cause theory; chain events perception and domino theory; the determinant variable perception and factorial theory; the multi-linear events sequences perception and process theory. Svenson (1999) similarly suggests the following four main ways of looking at accident causation:

One cause model

At the simplest level the view of accident causation is one cause leading to an accident (see Figure 6). In such a model therefore, analysis ends when one cause (usually the one immediately preceding the accident) is identified. Removal of this cause will prevent a recurrence of the event, assuming sole and sufficient cause. From a legal perspective this model is attractive because it focuses on only one cause and this makes it easier to find someone to blame.

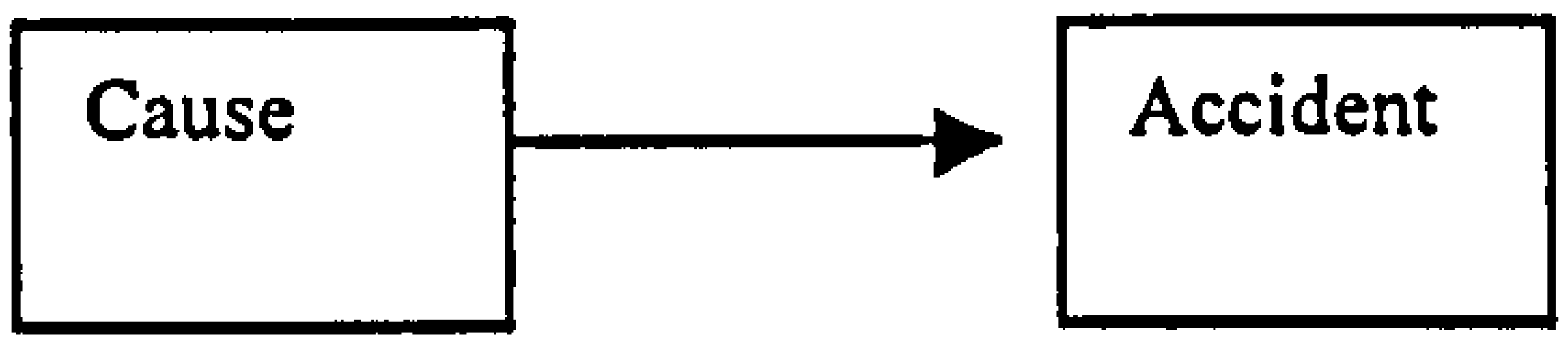


Figure 6: The single cause model

Chain or sequence of events (also called multiple steps causal model)

In this model one cause or event may lead to a second event and so on (see Figure 7). This model is only slightly more sophisticated than the one cause model, and is in fact the one cause model extended backwards in time to the cause behind the cause of the accident.

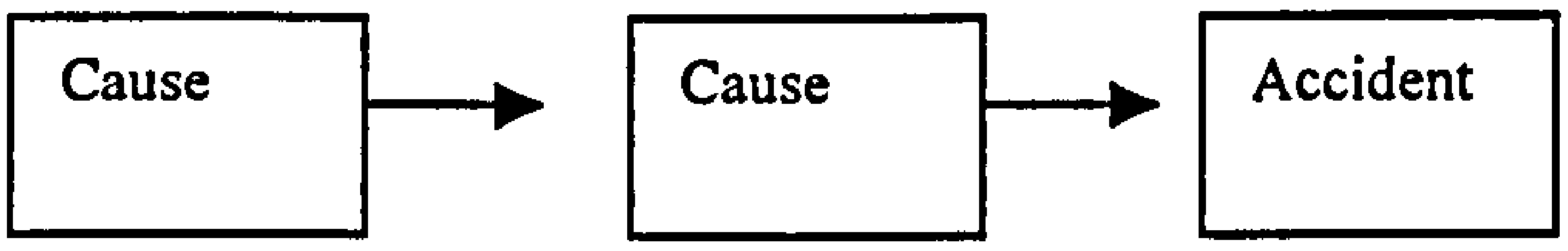


Figure 7:Multiple steps model

Multiple-cause model

In this model, several factors coincide and are linked to the occurrence of an abnormal event (see Figure 8). Individual causes may be necessary or sufficient to result in an accident. In this model accidents are viewed as the result of different single cause factors simultaneously contributing to the incident.

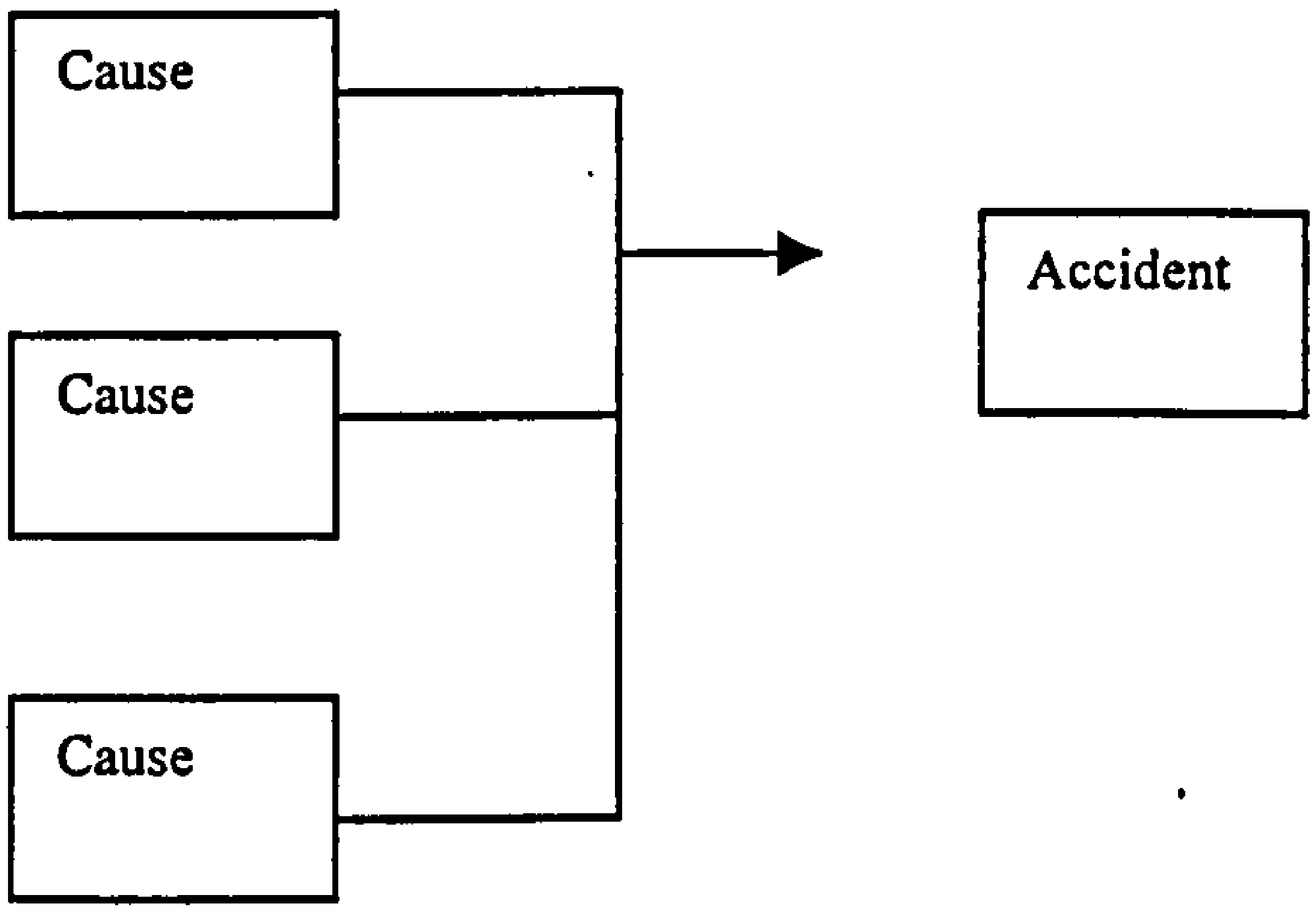


Figure 8: Multiple cause model

Multiple steps joint effects model

This model assumes multiple causes and traces back in time to discover the causes of the causes. This is illustrated in Figure 9 below.

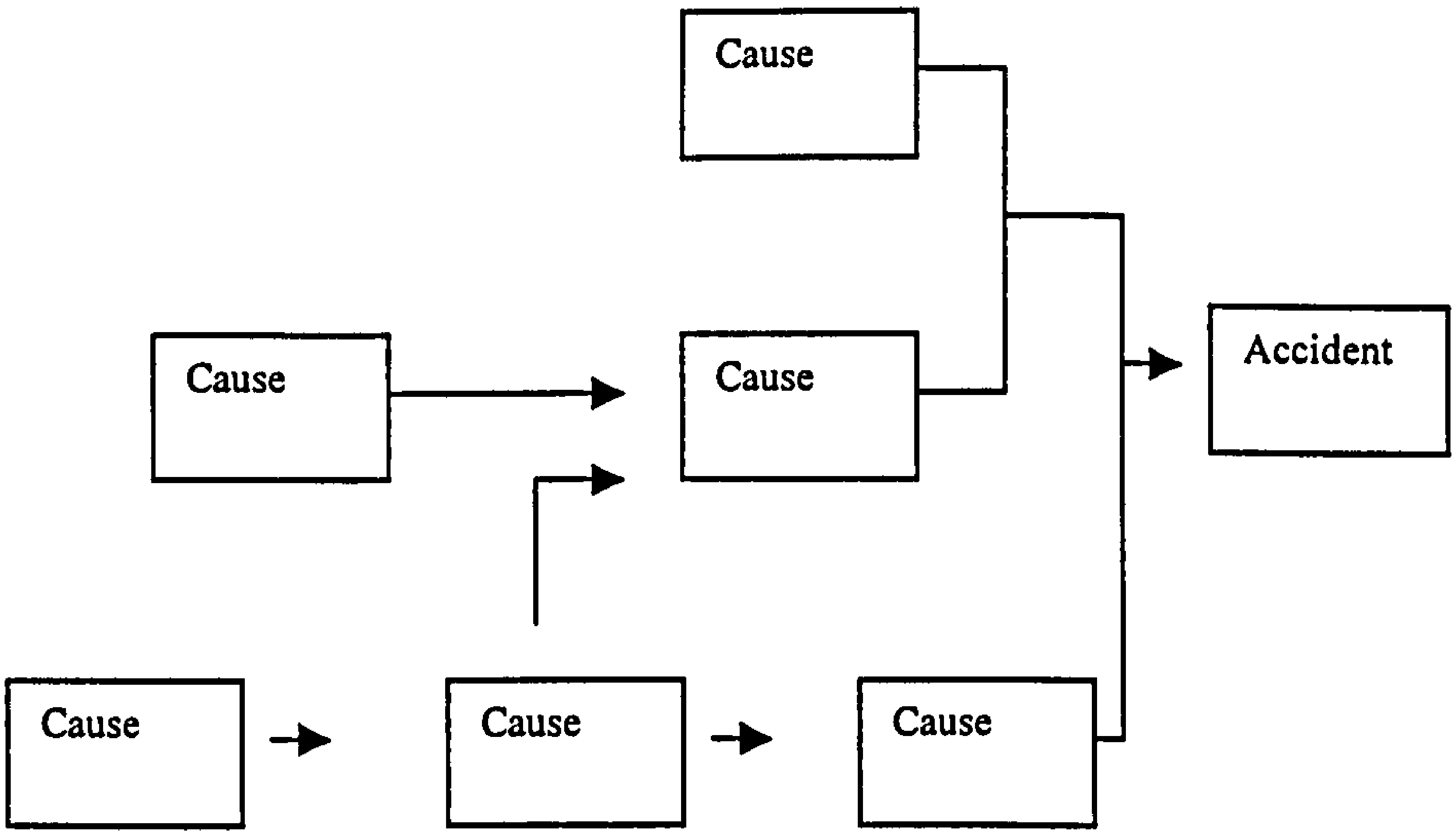


Figure 9: Multiple steps joint effects model

None of the above models will individually satisfy fully the way in which all types of accident or incident evolve. Some events may indeed be the result of a single cause, while others are the result of a more complex interaction of a number of factors. The trick is for analysts to avoid becoming blinkered in their approach: analysts should be open to the different possibilities and possible combinations that can lead to an unwanted event when performing analysis, rather than rigidly using one model to the exclusion of other possibilities. The next section traces the move from individual human error to the organisational accident.

2.2.1 From human error to organisational accident

While the view of accidents has moved from simple to an increasingly complex interaction of events, there has likewise been a move from simply blaming the active failures of individuals (unsafe acts and rule violations) by staff at the sharp end to increasingly including the failures of organisations (latent failures). The need to shift the focus from individual front line operators to the organisational failures was realised in the conclusions of the Zeebrugge and King's Cross inquiries (Sheen, 1987; Fennell 1988). These inquiries stated that rather than being the cause of the accidents the human operators were the inheritors of a defective system, riddled with poor design, conflicting goals and bad management decisions. Reason (1990; 1997) formalised the contribution of latent failures following examination of major disasters in a wide range of complex technologies using the 'resident pathogen' metaphor. Latent pathogens are likened to the 'resident pathogens' within the human body, which combine with external factors to bring about disease. Like cancer and heart disease, accidents in complex systems do not arise from a single cause. It is assumed that latent failures originate in the errors of high-level decision makers and only come to light when they combine or result in a failure at the sharp end which leads to an accident. It is further assumed that latent failures are present within the system for some time before they are made visible via an accident sequence. The focus on latent failures does not mean that human errors should be excluded from the analysis. As Reason himself says (op cit) the work of improving the immediate human-machine interface is still an important factor in accident reduction. Rather a combination of human errors, technical failures and latent failures will provide a fuller, more robust picture of accident causation and highlight the areas, or combination of areas where improvements to the whole safety system may be made.

2.3 Classification possibilities

Accidents and incidents may be classified (i.e. the contributing causes can be classified) according to models of error, behaviour and task. The following section will concentrate on step 4, Classification, of van der Schaaf's model, i.e. the approaches taken to event analysis and classification of causes. It has already been stated that the purposes or objectives of event analysis can be various and that this has an impact upon the way the events are analysed. However, even within a similarity of purpose, there exist a variety of ways in which to classify, code and analyse the data.

2.3.1 Taxonomies: Goals and requirements

Taxonomies provide a system for classifying accident and incident causes. They contain pre-defined categories (often including errors, latent failures and technical faults) which enable accident causes to be identified according to a generic set of possible causes. The sections below outline the goals and requirements of taxonomies.

2.3.2 Goals

The goal of any taxonomy used to code event data is to build a database of causes to enable the comparision of causes across multiple events. This requires pattern recognition across a number of minor accidents or of more major accidents into pre-coded categories. If there are no pre-coded categories then there can be no meaningful comparison. Using such a system it is possible to determine which factors or combinations of factors occur most frequently. This would then allow preventive measures to be based on the factors that occur most frequently, rather than on one major incident. The examination and analysis of single events does not allow for such a comparison and learning, prevention and improvements can therefore only take place from individual events. Such a reactive policy may fail to reduce the most frequently occurring causes and tackle only those causes which have resulted in high profile accidents.

2.3.3 The minimal requirements of a taxonomy

From the discussion on accident causation theories it is clear that current models assume that accidents are not always simply a one to one correspondence (e.g. one individual makes a single mistake and the result is a major accident). Rather they are more likely to be a combination of active and latent failures combining in a complex way to ultimately result in the catastrophic event or accident. Taxonomies have attempted to reconcile this by addressing human errors and organisational failures as well as technical faults. Consequently, a proliferation of taxonomies and methodologies have been designed to encompass various types of human and cognitive errors, organisational and technical failures. Many have been designed for specific industries (e.g. nuclear power, aviation) and are not directly transferable to other domains. Consequently, it is difficult to compare taxonomies and evaluate their usefulness. It is necessary therefore, that a common set of requirements that are built upon a common foundation, are developed that enable the evaluation of taxonomies, in order that users can be confident that the most appropriate and fit for purpose method is selected and used. Such an approach is not new (e.g. Benner 1985 subjectively rated 14 analysis methods). However, little research has actually been performed in the area of defining criteria and even less in evaluating taxonomies according to specified criteria. This section outlines a number of requirements that have been proposed for any taxonomy (and briefly discusses evaluations) to ensure that it provides the most benefit for the user.

A minimal set of requirements for any taxonomy used to identify accident causes has been identified by van der Schaaf (op cit) as involving the following:

1. Should be based on a suitable model of human behaviour
2. Should include technical components, human behaviour, and organisational and managerial causes.
3. Should be comprehensive in a qualitative sense (i.e. should be able to handle accidents, damages and near misses; cover both failures and recoveries)
4. Suggest ways of influencing and rectifying the root causes identified

There are a number of analysis methods that fulfil these requirements (although the majority do not incorporate identification of recovery strategies) such as MORT, STEP, TRIPOD, CIRAS and PRISMA among others. However, there are other criteria that can be used to evaluate the methods used. A number of authors suggest ways of determining the most appropriate method for classifying causes. Fahlbruch and Wilpert (1997) propose criteria that are relevant for choosing an appropriate analysis method. These are:

1. Theoretical adequacy
2. Manageability
3. Economy
4. Comprehensiveness
5. Problem solving characteristics

This set of criteria is similar to that proposed by van der Schaaf and has been applied to three event analysis approaches used in Nuclear Power Plants. Neither of the above criteria includes the issue of reliability (the consistency of coding between different raters). The table below provides a comparison of three methods (ASSET, HPES and MORT) as performed by Fahlbruch and Wilpert (op cit). The authors state that the metrics are fuzzy and not devoid of a certain degree of arbitrariness. It is not explained upon what basis the scores were assigned.

	ASSET	HPES	MORT
Theory	0	-	0/+
Manageability	0	+	0/+
Economy	-	0	+
Comprehensiveness	+	+	+
Problem solving	+	-	-

Table 8: Characteristics of event analysis methods (+ = given; 0 = limited; - = not given)

Wagenaar and van der Schrier (1997) and Groeneweg (1998) suggest any method should meet the following criteria:

1. Be revealing i.e. distinguish between events and underlying causes
2. Be quantitative, enabling results to be computed and accumulated across many accidents
3. Be valid i.e. that causes will also be present in future accident scenarios, while not missing any important causes of future accidents
4. Be reliable i.e. that various independent analysts should reach the same conclusions
5. Be practical i.e. the analysis can be performed without employing rare specialists, or costing too much in terms of time and money
6. Be consequential i.e. should formulate and rank order recommendations for accident prevention.

This set of criteria adds the issue of reliability but does not say anything about the theoretical model underlying the method of analysis. Nevertheless it is a useful set of criteria against which to compare various methods. Wagenaar et al (op cit) used these criteria to compare five different analysis methods, namely Fault Tree Analysis, MORT (Management Oversight and Risk Tree), STEP (Sequentially Timed Events Plotting) SCAT (Systemic Causal Analysis Technique) and TRIPOD.

The comparison of five accident techniques based on the above criteria are shown in Table 9 below: (it is assumed although not explained in the text, that (–) means does not fulfill the criterion, while (+) denotes that the criterion is fulfilled.

	FTA	MORT	STEP	SCAT	TRIPOD
Revealing	-	+	-	+	+
Quantitative	-	-/+	-	-	+
Valid	?	?	-	?	?+
Reliable	-	?	?	?	+
Practical	-/+	-/+	+	+	+
Consequential	-/+	-/+	+	-/+	-/+

Table 9: A comparison of five accident analysis techniques (from Wagenaar and van der Schrier 1997)

The lack of testing for validity found for all techniques is a result of the difficulty of measurement. The authors only consider the aspect of validity which is concerned with the accuracy of future predictions. The prediction of future accident scenarios however is based on the assumption that nothing of importance has changed. However, if an analysis technique aims to change the operating system and succeeds in doing so then it would be expected that future accidents will have different causes from past ones.

It is also worth noting that reliability (i.e. that various independent analysts reach the same conclusions regarding the causes of the accident or incident – also called consistency or inter-rater consensus) has not been tested for three of the five methods. This is an important omission for any taxonomy and is not a complicated process.

Despite the fact that TRIPOD does not generate recommendations, it has been ranked (-/+) for the Consequential category by the authors. This ranking was justified by the following statement "the TRIPOD method is not designed to produce specific recommendations, as we think this is mainly a job for the organisations themselves. The local flavour will have a substantial influence on the choice of recommendations; generating own solutions will enhance the ownership feelings of the recommendations". While this may in fact be justification for not producing recommendations it does not explain the positive ranking associated with TRIPOD for the 'consequential' category in Table 9 above.

While this evaluation of different methods is important research and provides some basis for selecting a method, the authors do not provide details of how the evaluation was performed (e.g. subjective judgement, expert opinion) or the level of acquaintance with the various methods. Further the authors state that the applicability of a method will depend on the reason for using the method. This may in part be true, but any taxonomy in use should satisfy at least a basic set of criteria to ensure it meets a minimum standard.

Shorrock (2002) suggests a comprehensive list of twelve criteria based on those proposed by Kirwan (1992). These are:

1. Comprehensiveness: the ability to discriminate and classify a comprehensive range of errors and influencing factors
2. Consistency: (also called reliability) the degree to which the method leads to consistent analyses between different users and with the same user over time
3. Life cycle applicability: the degree to which the technique can be used throughout the formative and summative phases of system design lifecycle
4. Predictive accuracy: the degree to which the technique is able to accurately predict potential errors
5. Theoretical validity: whether the technique is based on a model of human performance with a theoretically plausible internal structure
6. Context validity: whether the technique adequately captures contextual information
7. Flexibility: whether the technique enables different levels of analysis according to the needs of the project, expertise of the user or known information
8. Usefulness: whether the technique suggests or can generate error reduction or mitigation measures
9. Resource efficiency (training): the time taken to become proficient in the use of the technique

10. Resource efficiency (usage): the amount of time required to collect supporting information and conduct analysis
11. Usability: the ease of use of the technique
12. Auditability: the degree to which the system lends itself to auditable documentation

As with other criteria, the relative importance of the criteria are not rated by the author. However, a small survey of the importance of the criteria was carried out with eight human factors specialists working in the domain of air traffic management. They rated comprehensiveness as the most important with auditability as the least important for their particular domain.

There is obviously some cross over between the various sets of criteria. Shappell and Wiegmann (1997) are clear that an underlying theoretical model is essential for any accident analysis tool, and that the taxonomy used should take into account latent factors and human failures.

From the sets of criteria produced by the above authors, it is now proposed that the following list describes the *minimal* requirements of a taxonomy[*]:

---

[*] The minimal requirements of a taxonomy are currently being tested via a questionnaire study with relevant professionals.

1. Based on a suitable underlying theory of human behaviour
2. Should include technical components, human behaviour, and organisational and managerial causes.
3. Be reliable i.e. that various independent analysts should reach the same conclusions
4. Should be comprehensive in a qualitative sense (i.e. should be able to handle accidents, damages and near misses; cover both failures and recoveries)
5. Be quantitative, enabling results to be computed and accumulated across many accidents
6. Be consequential. Suggest ways of influencing and rectifying the root causes identified
7. Be revealing i.e. distinguish between events and underlying causes

A number of the criteria from Shorrock (2002) are not included here as a minimal requirement, such as auditability, flexibility, resource efficiency. These are aspects which are viewed as beneficial and optional (though not necessary to ensure that the taxonomy will aid in the process of assigning cause and helping to improve the safety of a system) and may be used in determining whether to institute a system (e.g. cost factors, training). The implementation of a method will, of course, depend to some extent upon cost. However, turning this into a minimal requirement may detract from the efficacy and applicability of the analysis process selected, as currently all minimal requirements are viewed as equally important. The validity aspect (called predictive accuracy by Shorrock (op cit)) is as yet an undetermined measure for taxonomies which retrospectively assign causes to incidents and accidents. As such it is not included in this list of minimal requirements. However, as robust methods develop to check the validity of analysis methods then this item could be added. It is viewed here as an attractive addition but not as a requirement. It is suggested that taxonomies should at least meet the minimal requirements (and preferably some of the optional ones) before being used in a 'real life' setting. Optional criteria for taxonomic approaches are therefore outlined below.

1. Validity
2. Auditability
3. Resource efficiency (training)
4. Resource efficiency (usage)
5. Usability
6. Life-cycle applicability
7. Flexibility

There is obviously research needed on the best way of ascertaining a method to determine whether taxonomies meet the requirements, rather than simply relying on the subjective measures that have been so far used

2.4 Criticisms of taxonomic approaches

There are a number of problems associated with the taxonomic approach to the analysis of events and assignment of causal factors to categories. A number of these difficulties are related solely to taxonomic approaches, while others are also a problem for the analysis of events in general and are independent of the approach used. This section will highlight general problems of accident analysis (which also apply to taxonomic approaches) and then focus on the problems that are specific to taxonomic approaches to the classification of causal factors.

2.4.1 Problems for the analysis of incidents

Whatever type of analysis is used to describe the factors that led to an incident there are a number of problems for analysts. Such general problems are listed below:

1. The investigation technique:

Any investigation technique should provide an adequate description of the incident in order to feed the system of analysis. If the quality and quantity of information obtained is insufficient, then the analyst is left with an incomplete description and unable to fully analyse the causes. There are many reasons why an investigation may not uncover a full description of the incident, or why the depth of the investigation is insufficient. These are listed below:

1a. The investigator's view of causation: Svenson (1999) and Benner (2000) have identified a variety of causation perceptions from simple one cause models to more complicated multiple causation models. When an investigator has satisfied him/herself that the causes (according to their preconception) have been identified then the investigation is likely to be halted. For example if the view of accident causation is a one cause model then when a necessary and sufficient cause is found the investigator may decide to terminate the investigation, thus losing the possibility of uncovering further causes via a more thorough investigation.

1b. The significance of the consequences: if the consequences of an event are great (e.g. fatality or major damage), then the event is likely to be investigated in depth. However, if a minor near miss has occurred it is unlikely that the investigation will be to the same level as the more serious incidents. However, this is not always the case: where a near miss has the potential to have become a serious incident then the depth of investigation is likely to be of a greater depth than is the case for non-consequential near misses. The depth of an investigation may also be a function of demands of the regulator or other official bodies. For example in the medical domain, a recent report 'Doing Less harm' (2001) proposes to categorise events according to severity on a colour coded scheme (ranging from red for serious to green for minor). According to the severity level there is guidance provided on the average length of time that investigation and root cause analysis should take to complete. For incidents categorised as red, it is suggested that investigation and root cause analysis take on the average between 2 and 20 days; for incidents categorised as orange this average time is reduced to between 1 and 3 days; yellow incidents should be investigated and analysed in an average of no more than 1 – 5 hours while green events should not be analysed for root causes at all. Such differences in investigation procedures may result in a real or artificial difference in the number of causal factors found for serious incidents versus less serious incidents. In either case, the comparison between different levels of severity is confounded by the differences in data collection.

1c. Time constraints: how much time can be spent in the investigation of events depends on any number of factors, e.g. how many events occur and have to be investigated; other roles undertaken by the investigator; number of analysts; regulatory requirements (deadlines for submitting reports etc); and management decisions.

1d. Eye witness evidence: another reason for an investigation remaining incomplete is unrelated to the characteristics of the investigator or the company, but relates to the account of eye witnesses or those involved. Eye witnesses or staff under suspicion may present defensive accounts for fear of blame and punishment or bringing punishment on colleagues (Adams & Hartwell 1977). Eye witness accounts may be also be contradictory – thus leaving the analyst / investigator to choose the most likely account or the one which matches the theory of the investigator. In some cases key witnesses are no longer able to provide a description of the incident due to death or injury. In such cases the investigator can only rely on the physical evidence and provide probabilities on the causes. Such is often the case in railway accidents when drivers are killed following a Signal Passed at Danger resulting in a head-on collision (e.g. Accident at Newton, 1991). Despite the depth of the investigation and the skill of the investigator involved, why the driver passed the signal will not be uncovered, and a full reconstruction is hence impossible, although the fact that a signal was passed is indisputable.

2. The detection of incidents: in order to analyse incidents they must first be detected. Automatic detection of some types of incident is limited to some high-tech high-consequence industries, and even these systems do not detect near misses or low grade events in all cases. Reporting systems which are not based on the automatic detection of incidents rely on reports from individuals who have witnessed or been involved in an accident or incident to report them. This is the case for both voluntary and mandatory reporting systems, although there may greater incentives, such as the absence of punishment to report to mandatory systems. This in effect, leads to fewer incidents being reported than actually occur. Johnson (2002) provides

evidence that a voluntary reporting scheme for adverse drug events detected less than 10% of the events found by manually combing medical records. Such a problem may lead to the under-estimation of certain types of problem. On the other hand, systems that automatically detect some incidents are limited by technology to detecting only certain types of incident. For example in the railway industry Signals Passed At Danger are automatically detected by the Signaller due to track circuit placement and has an almost 100% detection rate. This has led to a reactive policy to reduce that type of event, despite the fact that the potential for injuries and fatalities from the majority of SPADs is negligible due to physical barriers, while other events are given a lower priority.

3. Analytical biases: Biases and failures of reasoning from cognitive psychology (e.g. Tversky & Kahneman 1982; Johnson-Laird & Wason 1970) can also apply to accident investigation and analysis. Johnson (op cit) highlights a number of cognitive biases and discusses the way in which they affect the analysis of incidents in particular. These can effect any type of analysis.

3a. Confirmation and frequency bias: Confirmation bias results when the analyst attempts to ensure that the causal analysis of an incident fits with preconceived ideas about how the incident occurred. Frequency bias results when analysts become familiar with certain causal factors that have been observed or assigned regularly in past events. Future events are then also likely to be attributed to these frequently observed factors. Previous work performed in the nuclear industry (Wallace et al 2002) demonstrated the frequency bias, with the coding scheme in use producing replicable data patterns ("self checking not used" coded as the most likely cause of all incidents) from data that was demonstrated to have a lack of inter-rater reliability.

3b. Recognition bias: this arises when analysts have a limited repertoire of causal factors and hence try to fit the incident, despite issues of complexity, into one of the factors from their limited field.

3c. Author bias: this results when individuals are not willing to accept the results of causal analysis performed by someone other than themselves. For example, this could easily result in cases where there is conflict of interest between the company the analyst works for and the company who were involved in the accident.

3d. Political, sponsor and professional bias. Johnson (op cit) suggests that these types of bias result when the results of analysis are influenced by the political situation, professional recognition or the outcome for the sponsor. He calls these influences on causal analysis pressure from above, below and beside. However, such biases are not restricted to accident analysis but may also be present in many different types of research and investigation.

4. Counter-factual reasoning: a number of authors have discussed the role that counter-factual reasoning plays in accident analysis (Ferry, 1988; Ladkin, 2000). Counter-factual reasoning is used to disregard those factors without which an incident would still have occurred. However, Johnson (op cit) suggests that this can cause numerous problems for accident investigation and analysis. How can the analyst or the investigator be sure that an accident or incident would still have occurred if a causal factor had not been present.

The above list is not exhaustive, however it does present some of the problems associated with accident/ incident investigation and analysis at a general level. There is no simple solution to the majority of the problems described above, although awareness of where errors can be made is the first step towards avoiding the pit falls. The next section turns to the problems and criticisms that have been levelled solely against the taxonomic approach to event analysis.

2.4.2 Limitations and problems of taxonomic approaches

The following problems relate particularly to taxonomic approaches.

1. Reliability of assigning codes (also called consistency and inter-rater consensus): any classification system, whether for assigning causal factors to pre-defined categories, for assigning marks to students essays or for classifying mental illness (e.g. DSM IV) needs to be used reliably. There are two types of reliability that are important for taxonomies: inter-rater reliability and intra-rater reliability. Inter-rater reliability is the level of agreement between different analysts; while intra-rater reliability is the agreement of the same analyst with an event that he has already coded previously. The issue of inter-rater reliability for taxonomies that purport to define the causal factors involved in an incident, is highly important: the assigned categories are the basis on which improvements are made and prevention strategies based. Any taxonomy in use should be demonstrated to have an acceptable level of inter-rater reliability. Lekberg (1997) has demonstrated that there are fundamental differences in the ways different experts analyse incidents in the Swedish nuclear power industry. Wallace et al (2002) performed an inter-rater reliability trial with experts in a nuclear power plant and found a very low of agreement between experts using an established coding system that contained 196 codes and was in use in a number of plants. The poor reliability was attributed to a lack of logic in the coding scheme which had developed and grown over a number of years.

The consistency of codes assigned by one analyst over time is also of concern for classification systems. Changes in the coding of previously coded events may be related to 3d above (professional, political and sponsor bias), or to changing views of the analyst over time. If the same individual codes similar incidents differently, or the same incident differently after time has lapsed, then this again has implications for the comparison of the data and ultimately for the strategies used to prevent future recurrences.

The reliability (both inter and intra) of the coding of any taxonomy should be tested thoroughly before implementation. Training for new analysts should include reliability trials before competence in using the taxonomy is declared. Further frequent reliability trials should be performed on any taxonomy that is in use. If implemented, such measures would remove the criticisms of a lack of inter-rater reliability in the use of taxonomies.

2. Static coding structure: currently most taxonomies are based on state of the art theories of human behaviour. However, as theories advance it is unlikely that taxonomies will be updated as quickly as theory progresses. This means that the causal factors will be limited to previous ideas. Johnson (2002) suggests that few of the changes from the revised version of Reason's (1990) GEMs taxonomy have been included in taxonomies. Analysts may also retain distinctions in a database that no longer reflect the way in which tasks or activities are organised. Further, if alterations are made to the coding scheme then either the whole database requires a re-coding exercise or some of the data becomes outdated at best, and obsolete or

unuseable at worst. However, despite these problems, which are not insurmountable, the alternative remains to analyse each incident separately and have no common coding structure. The taxonomic approach has the advantage that causes can be accumulated over time and patterns observed – where there is no structure to the application of causal factors, there can also be no patterns identified. Hence prevention is based on individual cases and political factors rather than an accumulation of events with similar causal patterns. If new codes are added then the choice of re-coding or not using obsolete data must be taken into account. Where there are a significant number of reports, re-coding may not be an option due to time constraints or lack of staff.

3. Incomplete coverage: no one taxonomy can include all possible causal factors. This is not disputed. However no system can hope to include all possible causes of accidents or incidents (an attempt to do so would result in an enormous number of codes which would preclude the possibility of comparison, as many codes would contain insufficient numbers for any meaningful analysis). The solution for this problem that has been adopted by the PRISMA system is to have a category 'X' unclassifiable. These events are then regularly analysed to determine the number and the possible types of code that would apply. A new code may be added to the classification over time, if necessary and beneficial (using the codicils from 2. above).

There are both advantages and disadvantages to a pre-coded system. Fahlbruch and Wilpert (1997) state that a structured approach (such as a taxonomy or checklist) assures comprehensiveness by making it hard to overlook important potential factors. Further they state that such an approach has advantages over non-predetermined systems as it limits the problem solving freedom of the analysts. However, they also state that a taxonomic approach may also lead to a cause that is present in the accident scenario, but not in the taxonomy being overlooked. Fischoff et al (1978) have demonstrated that experts tend to overlook missing branches in causal trees. Further Sheahy (1979) has demonstrated that items placed at the beginning of a list are more likely to be chosen than those at the end of a list. However, despite these limitations, the advantages of a taxonomic approach are clear: patterns of causal factors can be identified and prevention strategies can then be based on an amalgamation of cases rather than on a single significant event.

2.5 Summary and conclusions

This chapter has provided an overview of the different purposes and goals of incident analysis; however, the most important and therefore the main goal of incident analysis should always be prevention. The different approaches to accident causation have been highlighted and it was concluded that a more robust picture of accident causation would be provided, by taking account of technical failures, human errors and latent failures, rather than focussing on individuals, technical faults or limiting the scope to a single cause.

The requirements for analysing incidents were discussed and a set of criteria for a taxonomy proposed. Finally, the problems related to accident investigation and analysis in general, and for taxonomic approaches in particular were discussed. It was concluded that despite the problems associated with taxonomic approaches, they provide advantages over single event analysis in terms of enabling prevention strategies to be devised and implemented.

In order to test the common cause hypothesis, it is necessary to analyse data using an appropriate taxonomy. As important as the method of analysis is the method of collecting near misses to be analysed. The next chapter discusses the ways in which to collect safety related event data and highlights the choices to be made in relation to setting up, implementing and maintaining a reporting system.

Chapter 3 Event reporting and analysis systems

3 Introduction

This chapter discusses the issues associated with designing and implementing a reporting system. In particular the chapter concentrates on the methodological issues that require to be addressed prior to any such system being launched. This chapter forms the methodological basis for the implementation of the CIRAS system (described in the following chapter) which was used to collect the near miss data used to test the common cause hypothesis.

3.1 The goal(s) of the reporting system

The goal or goals of a reporting system should be determined from the outset. The goal(s) of the system then drives the way the system is organised in terms of data input (the types of event that should be reported), data processing (the method selected to analyse the data, and how the data are selected for analysis) and output (implemented changes, monitoring and feedback to staff). Van der Schaaf's (1992) table shown below provides an overview of the modules necessary for a near miss framework. This framework outlines the various purposes and goals of a reporting system, and details the various modules necessary to fulfill these goals. For example, if the goal of the reporting system is to monitor the occurrence of known failures or problems, then the goal of the reporting system, and the particular failure that is being monitored should be clearly explained to staff. In this case, the reporting system is only interested in obtaining and analysing reports on this particular issue, and all other types of report are ignored.

Module	Modelling	Monitoring	Alertness
1. Detection recognition and reporting	Everything	Known problems only	Recognising and reporting
2. Selection of events for analysis according to purpose(s)	New reports only	**Not relevant**	Convincing
3. Description of all relevant hardware, human and organisational factors	Detailed	**Not relevant or very superficial**	Detailed examples of new and old hazards
4.Classification according to socio-technical model	Flexible: looking for new root causes	Routine: standard set of root causes	**Not relevant**
5.Computation statistical analysis to uncover certain (patterns of) factors	**Not relevant.** Only single events considered	Periodic analysis of updated large database	**Not relevant**
6. Interpretation and Implementation translation of statistical results into corrective and preventive measures	Finding (new) ways of improving prevention and recovery	**Not relevant/ Already prescribed by module 4.**	Near misses as precursors; focus on recovery mechanisms
7. Evaluation measuring the effectiveness of proposed measures after their implementation	**Not relevant**	Comparing predicted and actual effects of implemented measures	**Not relevant**

Table 10: An overview of different versions of the basic framework according to different purposes (from van der Schaaf 1992 p 37).

The table above represents the inputs to the system (modules 1 – 3), the way these inputs are processed (modules 4 – 6) and the output and evaluation which follows the preventive measures.

In terms of input, the designers of the reporting system need to decide on the type of events that the system is being designed to collect. There are a variety of choices. For example, major accidents, minor deviations and near misses, events that were successfully recovered, particular types of event such as SPADs (Signals Passed at Danger), LOCA (Loss of coolant accident) or unsafe practices and situations, etc. A near miss reporting system concentrates on near misses (including recovery actions), and often also on unsafe situations and practices as well. Giving a wide definition of near misses encourages staff to make reports and makes it easier for staff to decide to make a report. For an overview of the reasons why staff may not make a report see van der Schaaf and Kanse (2002). Once a reporting system receives reports, decisions are required on which reports to analyse in depth. The selection of reports for analysis becomes an increasingly important task with an increase in reports. When a reporting system is first launched it is likely that initially all reports can be analysed (in terms of manpower), however, this situation is likely to change rapidly as reporting rates increase. Hence, the analysts and the managers require to clearly identify the incident reports that will be analysed in depth, those that will be analysed in a minimal way and those which are identified simply as repeat incidents. Freitag and Hale (1997) call this a compromise between 'generating enough events to learn from and avoiding swamping the analysis system with too much work which will cost more that its added value'. The rationale for deciding on the appropriate level of analysis (from full analysis to no analysis) should be clear and easy to follow by all analysts.

The analysis of reported incidents is discussed in Chapter 2, in particular the use of a taxonomy for coding event causes and enabling pattern recognition across a number of events rather than learning from single events. The chosen analysis system should be compatible with the minimal requirements as outlined in the previous chapter. Furthermore, management need to provide adequate training in the use of the analysis system or hire experts to perform the analysis.

In terms of the final module 'evaluation, the actions implemented require to be monitored for success in reducing similar incidents. A system should be put in place to enable the monitoring and success of implemented prevention measures.

Although the above table provides a summary of all the steps required for the design of a successful reporting system, there are other issues that also need to be addressed when implementing a reporting system.

3.2 Management support

In order for any reporting scheme to be successful, generate and act on reports it is necessary that it is given full support at the highest level of the company (Lucas

1991). Ives (1991) provides evidence of management 'killing off' a reporting system through the unnecessary re-organisation of a successfully operating system. It is therefore important that management understand the purposes of the reporting system and foster its use rather create barriers.

3.3 Management attitude to near miss reporting schemes

Lucas (1991) cites three factors that are directly under management control which are vital for the success of near miss reporting (not necessarily only for a third party scheme, but also for in-house near miss reporting initiatives). These are anonymity, forgiveness and feedback.

3.3.1 Anonymity

Whether a near miss reporting scheme should be confidential or anonymous depends on the goals of the system, the depth of information required, the safety culture of the organisation and the way in which information is reported. If the reporting system allows anonymous reports to be submitted, then information gained from the first contact is the sum and total of all information that can be gained – an anonymous system prevents re-contact with the reporter. Confidential systems may result in the reporter becoming anonymous once the report has been fully processed. This is the case for example in both ASRS and CIRAS. For anonymity in such cases, read confidentiality which in effect allows the reporter to remain anonymous as far as the company disciplinary systems are concerned. CIRAS (which will be discussed in Chapter 3) is a *confidential system* to enable the reporter to be re-contacted and further information obtained.

Both confidential and non-confidential reporting schemes share the same analytical objectives, but confidentiality provides the reassurance that people need if they are to report many of the events or circumstances of most interest to risk management. Reporting often takes a great deal of courage, to overcome the fear of ridicule, embarrassment and retribution. This makes maintaining the confidence and trust of reporters essential. Harrison (1991) gives the following advantages of confidential reporting schemes:

1.Improves the completeness of accident and incident reports
2. Helps overcome the barriers associated with near miss reporting
3. Increases the information to build up a human error database
4. Increases suggestions for improvements.

3.3.1.1 The advantages of confidential reporting over anonymous reporting

There has been much discussion in the literature about the issue of confidentiality versus anonymity in reporting systems. Like confidential schemes, anonymous reporting schemes aim to overcome the barriers associated with self-reporting of errors and accidents. However, the disadvantages of anonymous reporting outweigh the benefits gained. This is because anonymous reports cannot be followed-up to obtain further details, nor can the source be verified. Therefore such a system is more likely to attract 'nonsense' reports or personally motivated reports designed to settle scores. An anonymous system is reliant upon the initial report for all details – any missing information remains missing. This can have a deleterious effect on analysis and therefore upon subsequent attempts to address the causes of incidents. Conversely, a system that provides confidentiality can verify the source and perform interviews with reporters to elicit further details. There are however, limitations to the confidential approach which also apply to anonymous reporting: reports cannot be verified by witnesses or by other staff who may have been involved in the incident, nor can technical evidence be requested from the company. Such limitations however, are balanced by the willingness to report and the type of information provided by reporters.

The majority of third party reporting systems e.g. CHIRP, ASRS and MARS (Marine Accident Reporting System) all provide confidentiality rather than anonymity for the reasons stated above.

3.3.2 Forgiveness

In terms of forgiveness, this means that management should not take punitive action against reporters no matter the circumstances of the report. When the reporting system is confidential, this is taken as a given. Much attention has been given to the punitive aspects of reporting near misses and accidents via a company or third party reporting system. Confidential systems are introduced in order to circumvent the disciplinary process in an attempt to gain information about what was happening 'at the coal face' rather than to find out who was breaking the rules or not performing adequately. Such systems are often no-blame, in so far as individual reporters are guaranteed not to be disciplined if they report something that they have been involved in only via the confidential system. However, if they are involved in an incident which is known to management and for which disciplinary action would be the normal outcome and also report it to the confidential system, they are not guaranteed protection from discipline. In other words, whilst confidential reports are treated confidentially and there is no blame or discipline attached to such reports, should the incident come to light via other means, then the individual will be treated as they would normally, whether a confidential report is filed or not. In fact, as management is never provided with the names of individuals making reports, there is no opportunity to cross reference confidential reports with reports or incidents that come to light through normal channels or investigations.

When ASRS was first introduced it provided immunity from prosecution for those submitting reports. Of course this had the effect of generating reports on the same incident by a number of individuals present at the time of the incident – all thereby ensuring that they were immune from the disciplinary process. Immunity from discipline can be counter-productive to the aims of a confidential reporting system as it distorts the number and quality of reports received (duplicate reports are counted separately). This results in statistical evidence that is based not on the number of near misses actually occurring but on the number of individuals who witnessed or were involved in the same incident. Thus immunity from discipline should not be an integral part of any near miss reporting system.

Much discussion has taken place as to whether a reporting scheme should be no-blame, provide immunity or simply exist within an enlightened culture. Berman (1996) suggests that a no-blame culture is both difficult to achieve and potentially self-defeating, and proposes instead a culture of enlightened response, although he fails to clearly identify how this could be achieved. Reason (2001) also discusses the need for an enlightened or 'just' culture specifically in relation to the railway industry. He suggests that a no-blame culture is neither feasible nor desirable. He does however, acknowledge the difficulty of reaching the state of a just culture: which depends upon understanding motivations as well as actions. In such a situation, managers have to decide which acts deserve punishment and which do not, and where to draw the line. This can in itself create what appears to be an unjust system. Furthermore, unless staff are able to discuss their motivations and actions openly, how can managers and those in a position to make judgements on the underlying motivations and required punishments, perform their function without prejudice or bias. Marx (2001) attempts to address these issues and provide guidance on establishing policies for a just culture.

The desirability of a just culture is not disputed here, but the feasibility of moving directly from a blame culture to a just culture (for example in the railway industry) seems overly optimistic. Until it is clearly demonstrated that a just culture exists, and until employees trust their managers with information without the fear or recrimination, near miss reporting will remain an under-utilised resource. A number of authors (Adams and Hartwell 1977, Webb et al 1989) have linked under-reporting of incidents and suppression of information to the apportionment of blame and disciplinary action i.e. to a blame culture. In the UK railway industry a combination of the blame culture and staff perceptions of the utility of making reports has resulted in under-reporting. Clarke (1998) found that the pattern of intended under-reporting indicated a 'risk management' cultural approach (Lucas 1991) in the industry, which serves to emphasise specific types of incident (SPADs and Wrong Side Failures [WSFs]) which should be reported to the detriment of a broader range of potential problems. Clarke (op cit) also found that only 3% of Drivers would report rule breaking by a colleague. Confidential reporting is the first step towards re-gaining the trust of staff and establishing a reporting culture. A 'just' culture is hopefully the next step.

3.3.3 Feedback

The third factor which Lucas (op cit) identifies as vital for the success of near miss reporting is feedback. Reporting schemes may be readily accepted and embraced by staff – reports may be forthcoming in the start up phase. However, if staff are to continue to use the system and keep making reports then they have to see concrete results and receive feedback on what actions their reports have generated. This may be done through individual feedback or via a publication which all potential reporters receive. This ensures that reports are not seen as entering a 'black hole' and disappearing without trace.

3.4 Reporting routes

Decisions also need to be made on the route via which reports can be made. This decision may be constrained by the type of system that is implemented (e.g. users of a fully anonymous system may be unlikely to use a telephone route out of office hours as this might require the reporter to leave their telephone number on an answering machine). There are a number of possible routes that are suitable for making reports. These are as follows: via a form, electronically, via telephone.

When a report is made on paper there are a number of choices in terms of design and information requested. The form may contain a number of questions for the reporter to answer (such as date, time, equipment being used, place of event, etc), or it may contain a fairly open question such as 'describe the event you wish to report and how/ why you think it happened'. If the reporting scheme is not anonymous then there should be space for the name and contact details of the reporter. Thought should be given to the method of eliciting reports – to make it as easy as possible for reporters reply paid envelopes could be provided along with a supply of forms. Forms should be available in places that are easy for potential reporters to access and not on request from senior managers.

Similarly, if the report is to be made electronically, similar decisions on the design of the questions to be answered require to be made. In the case of CIRS, reports are made electronically and are anonymous – however it is likely that an individual could be traced via the internet, and further assurances may be required in such cases. Also, unless reporters have access to computer technology, an electronic reporting system is likely to have limited success.

There are a number of advantages and disadvantages to instituting telephone reporting. The advantages relate to the ease and rapidity of making a report. Further, the person making the report is directly in contact with someone from the reporting system and thus there is a greater sense of involvement and personalisation, and possibly more information provided by the reporter. The disadvantages are that if the reporter calls out of hours or when the analysts are unavailable, then they may not leave a message or a contact number. CIRAS accepts both phone calls and forms and has found that, in general, reporters are willing to leave a contact name and number on the answering service. Telephone reporting is likely to increase the staff required to man the reporting system. However, should telephone reporting be an attractive option, then it is made more attractive to potential reporters if the number is Freephone and involves no cost to the reporter.

3.5 Incentives for reporting

A near miss reporting scheme may wish to increase the number of reports it receives by providing incentives for staff to submit reports. Such promotional campaigns should not be undertaken lightly, and the dangers of doing so should be taken into account. The rewarding of reports may lead to biases in the data which would not otherwise be present; trivial reports may be made in order to claim the rewards, or fallacious reports may be generated in order to receive rewards. Also after stopping the rewards, reporting rates may fall – thus confounding the reasons for the fall in reports. An example from the literature may help to highlight the dangers of rewarding employees for making reports concerning safety. The smallpox case is taken from Makin and Sutherland (1991). "..the international health organisations undertook a concerted campaign to eradicate smallpox...the 'front line troops' for the campaign were health visitors. Each of these had a geographical area for which they were responsible. In order to motivate the health visitors a bonus scheme was introduced. Arguing that the final goal was eradication of smallpox, a scheme was devised whereby each visitor was rewarded according to the absence of smallpox in their area. However, although the visitors consistently earned good bonuses, smallpox remained endemic. When considered from the visitors' perspective the reasons for this apparently paradoxical situation becomes clear. If you are rewarded for lack of cases, the incentive is to turn a blind eye. When in doubt don't report. The system is obviously open to abuse. Management finally realised the potential for abuse, and the reward system was turned on its head. Instead of being rewarded for the absence of cases, visitors were now rewarded for finding cases. The results were dramatic, undiscovered cases now came to the attention of the authorities and could be treated." This case highlights the dangers of providing rewards for making reports, and therefore such reward systems should not be part of an event reporting system.

3.6 Preparation and planning

Before a reporting system (confidential or otherwise) can be launched in a company there must adequate planning and preparation in all of the issues discussed in the sections above. This includes providing and designing for the following:

- a route via which reports can be (decisions include whether it is electronic, paper-based, form-driven, telephone);
- adequate staff and training should be provided to deal with the reports (this includes interviewing and investigation techniques, familiarity with the analysis process);
- appropriate analysis method;
- publicity and briefing to staff to ensure they are aware of the scheme, can recognise what to report and know how make reports;
- appropriate feedback channels.

As an example of how important preparation and planning are the CIRAS system (which is described in detail in the next chapter) took over one year at the planning stage before it was implemented.

3.7 Evaluation

Once the reporting system has been designed and implemented review and evaluation of the system is necessary to ensure that the system fulfils the goal(s) that were the impetus for initiating the system. Such evaluations may include the effectiveness of specific measures or a comparison of near miss analysis with other safety performance indicators (van der Schaaf 1991). However, evaluation is difficult to achieve in practice for the following reasons:

- Multiple interventions usually take place in parallel (ideally to test the effectiveness of a particular corrective measure it would be implemented in isolation and the effect could be measured. As corrective measures also take place at the same time as many other company initiatives and rule changes it is very difficult to assess how successful any one corrective measure is).
- The time lag for changes to occur and take effect is unknown – hence multiple interventions may have taken place and it is difficult to be sure which intervention is responsible for the changes measured.
- Organisations and their culture and rules change over time.

3.8 Summary and conclusions

This chapter has highlighted the various issues that need to be taken into consideration when designing and implementing a reporting system for safety events such as near misses.

The goal(s) of the reporting system should be decided upon in the first instance, and this then has implications for the design of the elements that comprise the reporting system. Support from senior management is vital for the success of any reporting system, and management can foster trust in a reporting system by considering the importance of anonymity (confidentiality), forgiveness (no-blame) and feedback. A confidential system provides advantages over anonymous reporting systems, as it enables reporters to be re-contacted and have trust in the system. Reporting routes were discussed, including paper methods and electronic methods. No one method for reporting is best in all circumstances and the method or methods chosen should reflect the goals and aims of the system, and the abilities of the reporters. Potential reporters should be made aware of the system via briefing and publicity material. Incentives used to increase reporting and encourage reporters should be avoided as they can distort the type and quality of data received by the reporting system. The analysis system should be carefully selected and it should fulfil the criteria outlined in chapter 2. Analysts and users of the system should be trained in the system and fully understand both the analysis and the investigation process. Finally, before launching a system adequate preparation and planning should take place.

The next chapter introduces an incident reporting and analysis system implemented in the UK railway system: CIRAS (confidential incident reporting and analysis system). The system is fully described in terms of the reasons for introduction, the aims and goals, operation and the taxonomy used. The taxonomy is compared to the criteria defined in Chapter 2 in order to evaluate the appropriateness of using the taxonomy as the analysis methodology for testing the common cause hypothesis. The data and coding scheme used by the CIRAS system then forms the basis for an empirical test of the common cause hypothesis.

Chapter 4 CIRAS and the CIRAS taxonomy

4 The Confidential Incident Reporting and Analysis System (CIRAS)

This chapter introduces and describes CIRAS, the confidential reporting programme operating in the UK railway industry. The rationale for the introduction of the system is discussed. The CIRAS system was initially introduced in September of 1996 and expanded to cover the entire UK railway network in June 2000, at which time all companies holding a Railway Safety Case were mandated to join the system. As this thesis is only using data collected prior to the system becoming national, the description of the system is mainly focussed on the years 1996 – 2000, although a small description of the way the system was introduced and organised nationally is included. The CIRAS taxonomy used prior to the expansion of the CIRAS system is fully described. The validity of using the taxonomy to test the common cause hypothesis is discussed with reference to the requirements of a taxonomy as outlined in Chapter 2.

4.1 Introduction to CIRAS

CIRAS is the Confidential Incident Reporting and Analysis System currently being used in order to identify and deal with human factors problems on the railways in the UK. CIRAS was initially a response to the contribution of human factors (including human error and latent failures) to incidents, situations and near misses on the railways. An earlier background report by Heybroek (1995) of Vosper Thornycroft commissioned by ScotRail Railways Ltd, pointed out the role of human factors in the rail industry, and the importance of these has also been highlighted in other industries (e.g. the off-shore oil industry; nuclear industry). Furthermore, existing official reporting procedures are often associated with disciplinary action, and this distorts both the nature and number of reports received. This is particularly true in the rail industry where, historically, relationships between workforce and management have sometimes been characterised by mutual mistrust and animosity, rather than co-operation. This results in a tendency for reports to become focused on technical failures and chance happenings (the reports tend to be strategic, defensive and external) with the human element being virtually absent. In some instances, it may even be the case that a near miss or incident with no obvious consequences will be deliberately concealed (i.e. the person concerned feels lucky to "have got away with it this time") due to the perceived disciplinary implications, rather than being seen as something from which others could usefully learn (Frese and van Dyck 1996, Edmondson 1996).

The aim of the system is to collect reports from individuals (Drivers, Signallers and other safety critical employees) of near misses, incidents and error promoting conditions, which would not normally be reported through official channels, and to use this information to enhance existing safety management systems. CIRAS is not intended to replace existing reporting procedures, and is a complementary system which operates in parallel with existing reporting channels. CIRAS is confidential and "blame free", and therefore staff can report not only technical failures, but also operator or human errors without fear of recrimination and discipline.

The CIRAS system by ensuring confidentiality seeks to rectify this imbalance, promote a reporting culture and hence producing causal human factors data that otherwise go undetected and unrecorded. CIRAS is also timely since it complements the privatisation of the rail industry in the UK, opening the door to new, more open management systems and changes in safety culture. The CIRAS system was pioneered and introduced in September of 1996 by ScotRail Railways Ltd (a major Train Operating Company), just as the privatisation process was drawing to a close, and was extended to include all companies that are part of the Railway Group in June 2000, when privatisation had been completed and operating for some time.

4.2 The background to the CIRAS system

Following a report by Heybroek (op cit) CIRAS was introduced as a pilot study in September of 1996 to determine the usefulness of confidential reporting for ScotRail Railways Ltd. This report suggested that rail incidents (e.g. SPADs – Signals Passed at Danger) had similarities to the aviation experience in which technical design and engineering improvements appeared powerless to reduce the incidents caused by human error. Further the report reviewed a number of SPAD incidents and concluded that the investigations failed to identify causes for a number of reasons.

Firstly, the blame culture within the railway industry camouflages the causes (Drivers are unwilling to openly and honestly discuss the incident due to the possibility of disciplinary repercussions – and admitted to Heybroek that if SPAD reporting were a no-blame procedure their accounts of incidents would probably change). This pattern of 'defensive communication' has been well-documented – the goal of which is for workers to protect themselves from the threat of blame or punishment (DeSalvo & Zurcher 1984, Gibb 1961).

Secondly, the framework for the investigation of SPAD incidents is based on an industry standard form, which classifies SPADs according to categories (such as Disregard, Misjudge) and not to causes. Thus investigators fulfill only the requirements of the SPAD reporting form and stop short of identifying causal factors.

Following these findings the report concluded that a confidential reporting programme should be instituted to collect reports of near misses and other safety incidents. Confidential reporting should reduce the fear of making reports, and allow Drivers to be more open and honest about their role in incidents. In turn this should aid industry learning, enable resources to be targetted most effectively and hence reduce the number of incidents occurring.

The CIRAS system was officially introduced in September of 1996, following a period of intensive training and briefing with staff. Following a successful two year pilot period, further companies approached the CIRAS team with requests to join the system. These included a number of train operators, maintenance contractor firms and a rail infrastructure controller. In the June of 2000, the system was mandated by Railtrack Safety and Standards Directorate (now Railway Safety Limited) and introduced to all companies which were a member of the Railway Group i.e. all companies in possession of a Railway Safety Case. The introduction of the system across the UK will be discussed in section 3.5.

4.3 The development of the CIRAS system

In order to move smoothly from the conception of CIRAS to a fully operational system a number of steps were taken. Firstly a Steering Committee was established. In the first instance, this committee existed to ensure the system was developed in a manner which would enable the appropriate and secure collection, analysis and feedback of data regarding near misses. The committee oversaw the development of the system and the rules by which it operated. Once the system was established the committee continued to oversee the system and ensure it produced tangible results. The Steering Committee comprised members from the original team from Vosper Thornycroft who recommended a confidential system be established to complement the existing reporting procedures within the railway company they had investigated.

Members from ScotRail Railways Ltd also sat on this committee in order to ensure that the system would meet their needs and requirements. Members from the University of Strathclyde who were to develop and administer the system were also part of the committee, as was a representative of the Health and Safety Executive. In addition an independent human factors expert was established as the Chair of the Steering Committee. Prior to the establishment of CIRAS as a national system the committee met on a three monthly basis. Prior to each meeting the committee members received a comprehensive report detailing the reports received, an overview of the human factors contained in the reports and a review of the actions taken as a result of reports.

While the Steering Committee existed to develop policy, procedure and generally ensure the system operated effectively and within the remit, a Liaison Committee was established to direct the project at a local level and to respond directly to the reports received from Drivers. The Liaison Committee included representatives from ScotRail Railways Ltd and the corresponding Railtrack Zone, HMRI (Her Majesty's Railway Inspectorate), ASLEF and RMT (trades unions), and the University of Strathclyde. This committee was responsible for compiling feedback to Drivers in the form of a three monthly journal publication (posted direct to the homes of the potential users of the system) and for recommending actions on issues of safety.

The design and operation of the system was based on best practice from other reporting systems and industries. At the time of the development of CIRAS, CHIRPs (Confidential Human Factors Reporting Programme) had already been in operation for the aviation industry for more than a decade. Accordingly, the team visited Peter Tait and his colleagues at CHIRP to discuss the implementation and data analysis performed for the civil aviation industry. Best practice in terms of feedback to staff via a newsletter was adopted following this visit. CHIRPs do not accept anonymous reports (for reasons of clarification, obtaining further information and to prevent spurious reporting) and this was also deemed to be the best route for the CIRAS system to take. As the aviation industry were more advanced than the railway industry in terms of reporting systems, the research team from the University of Strathclyde also visited BASIS (British Airways Safety Information System). Unlike CHIRPs or CIRAS this is an in-house reporting system, which has three tiers of reporting – one of them being a confidential system. For a full description of the BASIS system see O'Leary (2001). Again valuable lessons were learned from the experience of BASIS and the administrator, Mike O'Leary. Finally, an administrator of EUFORCE – a European aviation reporting system - (Captain Paul Wilson) kindly discussed his experiences and provided valuable information on the pitfalls to avoid and the necessary elements for any confidential system. During these learning sessions it became clear that the principles for a confidential reporting system in the railway industry were the same as for the aviation industry but the current systems of analysis used by the aviation industry were not suitable for the railway industry. The databases did not use human factors codes in a way directly transferable. BASIS leaned heavily on team work and crew resource management – the driving function on trains is an individual task and does not rely on team co-operation or the equivalent of a co-pilot. It was therefore necessary to develop a causal coding system that was relevant to the railway industry.

An extensive review of the current literature was undertaken in order to find a suitable model for use in the railway industry. The model that was finally developed by the University of Strathclyde drew heavily from the work of Rasmussen (1986) and Reason (1990) incorporating the Skill- Rule- Knowledge distinctions and the organisational and human causal factors. Further as perception of signals is fundamental to the role of the Driver, perception was also included in the model. Shappell and Wiegmann (2001) also recognise the need for a perceptual code and have included it in their taxonomy called HFACS (Human Factors Analysis and Classification System). A supervisory layer was added to the human and organisational causes as the railway industry is extremely hierarchical. Finally a specific set of technical codes for the railway industry was developed in order to locate the area where technical problems occurred (on sets, at signals etc) rather than to determine the nature of the technical fault. Drawing on the work of van der Schaaf (1991) recovery actions were also incorporated in the model. The original

CIRAS model is fully described in section 3.9. Examples of reports made via CIRAS and preliminary data analysis can be found in Wright et al (2000) and Davies et al (2000).

In parallel with carrying out the development of the CIRAS system, the team were also involved in trend and pattern analysis in the nuclear industry. Lessons were learned from this work. The data from a nuclear plant being researched resulted from an elaborate and very comprehensive coding system comprising some one hundred and seventy codes. The first step was to establish whether the codes had been assigned reliably by the different engineers performing the causal analysis. Full results of this study are related by Wallace et al (2002). However, a summary will provide sufficient detail here. Twenty eight incidents were coded by three experienced coders using the normal system and assigned in the usual way. The average index of concordance was 42%. While there is no generally agreed criteria for acceptance of data from the index of concordance (Caro et al 1979), the National Center on Child Abuse and Neglect (NCCAN 1998) suggest 75% as an acceptable level of agreement. The problem with the nuclear industry's system was that it was highly detailed without any logical hierarchical organisation and therefore raters failed to agree. A further problem with such a comprehensive and in-depth taxonomy is that it takes many years to establish enough data to perform any meaningful statistics. Therefore, it was decided to make the CIRAS taxonomy more general in order to increase reliability and obtain sufficient data for meaningful analysis to take place over a shorter time frame than a more detailed taxonomy would allow. In order to compensate for the more general nature of the taxonomy, a detailed analysis of the transcripts of the interviews (which followed all reports where the reporter was agreeable) was incorporated into the system. This was performed using Sage Publications 'NU:DIST' programme and allowed all text relating to the individual codes to be retrieved easily: thus each causal code was associated with a field in the qualitative data analysis system and could be recalled to provide further details and explanations if and when required. It was hoped that this would provide a solution to the problems of over-generalisation that are often levelled against a static taxonomic approach (see Chapter 2 for a discussion of these problems).

4.4 CIRAS operation

In the simplest terms, safety critical staff voluntarily report safety concerns, unsafe actions and practices direct to CIRAS personnel. Reports are made via a standard reporting form (which asks for a name, phone number and address, and information about the incident in the respondent's own words) or by telephoning directly to a CIRAS employee. Reports are followed up by a telephone or face to face interview where more information is sought including demographic data on time, place, date, length of shift etc. A full list of variables requested at interview is provided in the follow-up form in Appendix 1. Interviews are not performed on either employer premises or time. Interviews are tape-recorded with permission and fully transcribed to facilitate the analysis and feedback of information to the individual companies. These reports are then disidentified, raised with the relevant management (at quarterly Liaison group meetings) and formal feedback is provided on the issues and incidents raised at three-monthly intervals. Clarke (1998) conducted a questionnaire study with 128 train drivers from BR prior to privatisation, and found that lack of action and trust in managers were the major reasons for failure to report incidents. Accordingly the feedback, demonstrating actions that have been taken are essential to maintain enthusiasm and use of the system. The feedback takes two forms. Firstly, there is the CIRAS Journal which is published quarterly and circulated to all employees who are potential users of the system. This provides information on the types of things reported, gives a management response, invites comment and includes a new CIRAS form. It has been demonstrated that such feedback is necessary to ensure the success of any reporting system (see Lucas in van der Schaaf, Lucas & Hale 1991). Secondly, data are classified according to the CIRAS taxonomy, which provides the basis for a more detailed report which is sent to management. This gives a more precise human factors description, includes data analysis, and makes recommendations where appropriate. As with the CIRAS Journal, all information is completely disidentified. Should an individual want personal feedback on the issue or incident they have reported, they are able to contact CIRAS and after providing suitable details on the incident they are provided a personal update. Figure 10 below outlines the basics of the CIRAS process.

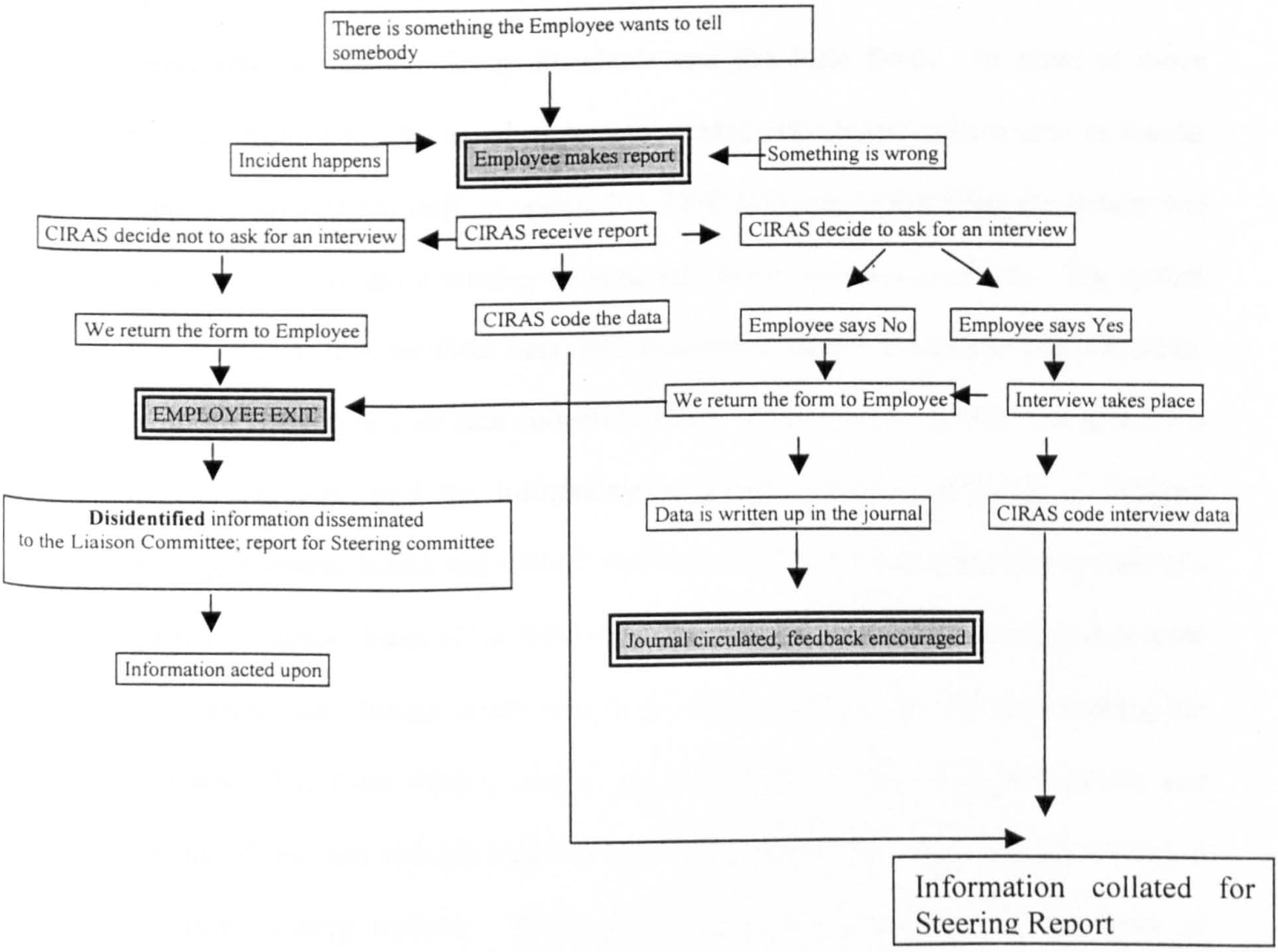


Figure 10: The CIRAS Process

4.5 The national CIRAS system

Following the success of the CIRAS system for the companies enrolled from September 1996, and the interest generated within the industry for such a system, it was mandated in June 2000. This in effect, made it compulsory for all companies who were members of the Railway Group to join the system. The mandate was applied by the then, Railtrack Safety and Standards Directorate who were responsible for safety, Group Standards and the Rule Book. In order to move forward from a locally based system to a nationally based system able to handle reports from +70,000 staff, as opposed to 5,000 involved at that time, the system was logically divided into a number of parts and these were tendered out. The system was divided as follows: three data collection bases (Regions) and one analysis centre (Core). The Regions or data collection bases would receive reports and perform a cursory analysis, feed the information to liaison groups consisting of industry managers, trades unions and CIRAS staff and publish the responses and actions in a quarterly journal. Each of the three receiving bases would publish a separate journal to retain a local flavour which was found to be preferred by the staff reading the journals. The Core facility was to be responsible for the in-depth analysis and coding of the data collected according to the CIRAS taxonomy and for writing 6 monthly industry reports. The Regional bases were awarded to University of Strathclyde (Glasgow); WS Atkins (Warrington) and DERA (Farnborough). The various companies were divided amongst the three Regional Facilities thus providing full coverage of the UK mainland. The Core Facility was awarded to the University of Strathclyde. The Core was responsible for writing Standards to ensure that the

CIRAS system provided a united front and a seamless entity, despite being run by various organisations. The figure below represents the national process.

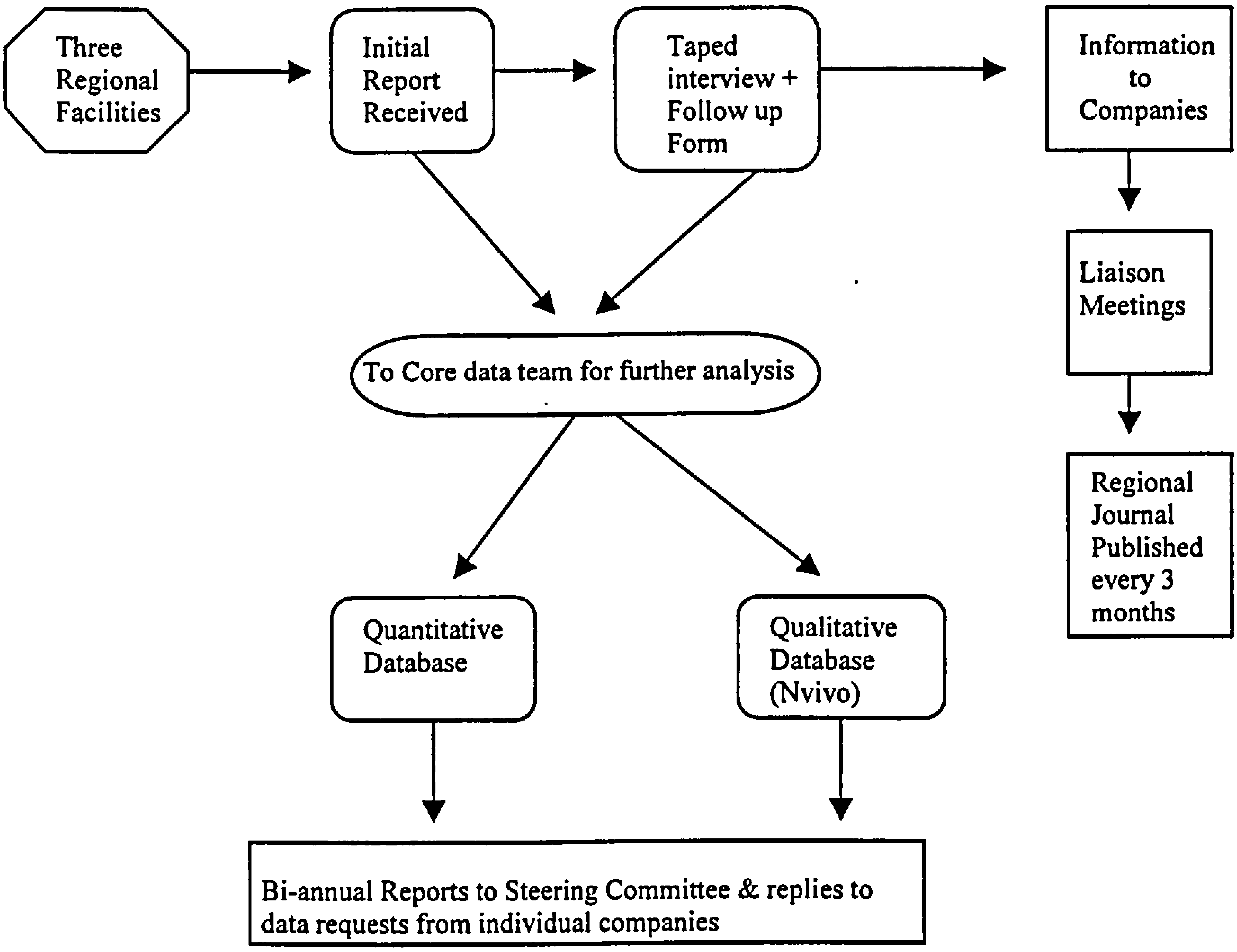


Figure 11: The National CIRAS process

Each Regional Facility works to the CIRAS process outlined in Figure 10 above and all adhere to the CIRAS standards and procedures to ensure confidentiality.

Following the nationalisation of the system, the taxonomy underwent a series of small changes and additions. For the purposes of this thesis, only data collected and analysed during the first four years, when the system was local will be used. Therefore the discussion of the taxonomy relates to the original taxonomy developed for the system from September 1996 to June 2000. While the amendments to the taxonomy are minor, they present difficulties in comparing early and late data (without a re-coding exercise and subsequent reliability trials). Furthermore, the author was involved in the building of the original model and the collection and coding of the data during the period 1996 – 2000.

4.6 CIRAS confidentiality

With a system like CIRAS, confidentiality is clearly of the greatest importance. The system has to be totally trustworthy, and has to be seen to be so. A single breach of confidence or security would cause the system to lose all credibility. The process employed by the CIRAS system to maintain the confidentiality of reporters is outlined in this section.

In order to verify reports and obtain more information, it is essential that staff provide a name and contact telephone number and/ or address. Reports which are submitted without personal details are not included in the CIRAS data base. The confidentiality process begins at the stage of receiving a valid report form or telephone call. At no point will CIRAS provide management with details that would identify an individual.

The data collection process was also designed to maintain the confidentiality of the reporters. In order to ensure this, interviews are not conducted on work premises or time. This means that other staff do not observe reporters speaking with CIRAS personnel, and reporters are not in the position of requesting time off from duty to attend a CIRAS interview. The majority of interviews that are conducted face to face take place either at the reporter's home, or in a quiet location where colleagues are unlikely to be. Telephone interviews are likewise conducted during the reporters off duty hours and usually at their home telephone number. Under no circumstances are telephone interviews performed at the workplace.

In terms of protection and security of the data, the following measures were designed to provide assurances of the confidentiality. The CIRAS database is kept in a burglar-proof and protected location; data are stored on a removable hard drive which is kept in a safe overnight; and the computer itself is 'stand-alone' and non-networked so it cannot be hacked. The original report form is returned to the reporter after the interview is completed, no copies are made or retained and all identifiers are removed when data are entered onto the computer. Thereafter it is impossible for any record to be traced to any individual through the database. Finally, all data are coded in terms of event-number, and no names or locations are recorded. When the system was nationalised, Standards and Procedures were drawn up which maintained the pre-nationalisation security aspects, making the national CIRAS system seamless, despite being administered at different geographical locations, by different companies.

In order to ensure confidence in the system, and relevant knowledge of it, railway staff were briefed on the system by their line managers who were in turn briefed by CIRAS personnel at a two day workshop. Due to logistics and the number of staff involved it was impossible for CIRAS to personally brief all railway staff. However, in order to ensure understanding and answer any questions which staff may have, a selection of safety briefing days was attended by CIRAS personnel on completion of the process by line managers. In general driving staff seemed enthusiastic about the system and understood the principles. As with the introduction of any new system, a minority of Drivers viewed CIRAS with suspicion and distrust. By the time CIRAS was mandated as a national system, these same Drivers were unhappy that **their** reporting system was to be changed to a national one – their sense of ownership gained during three years of operation was undermined by the move to a national scheme.

4.7 Immunity and no-blame

Much attention has been given to the punitive aspects of reporting near misses and accidents via a company or third party reporting system. CIRAS was introduced in order to circumvent the disciplinary process in an attempt to gain information about what was happening 'at the coal face' rather than to find out who was breaking the rules or not performing adequately. The system is no-blame, in so far as individual reporters are guaranteed not to be disciplined if they report something that they have been involved in only via the CIRAS system. However, if they are involved in an incident which is known to management and for which disciplinary action would be the normal outcome and also report it to CIRAS, they are not guaranteed protection from discipline. In other words, whilst CIRAS reports are treated confidentially and there is no blame or discipline attached to a CIRAS report, should the incident come to light via other means, then the individual will be treated as they would normally, whether a CIRAS report is filed or not. In fact, as management is never provided with the names of individuals making reports, there is no opportunity to cross reference CIRAS reports with reports or incidents that come to light through normal channels or investigations.

4.8 The CIRAS data

The data that the CIRAS system is designed to deal with are of two types. Firstly there are textual data in the form of transcribed semi- structured interviews with reporters, and transcriptions of initial reports from the reporting form or telephone report. These form the core of the database. Secondly there are demographic and other data on recorded on hard copies of what is known as a follow up form (see Appendix 1). The follow up data includes age and length of experience of the reporter; the day and time that the incident occurred; the type of train involved; the type of signalling involved and other personal and railway related data. The former data are processed via the CIRAS taxonomy and entered onto the database when coding is completed, while the latter are directly entered onto the database. As both types of data are held on the same database, the information from the follow-up form can easily be cross-referenced with human factors data.

4.9 The CIRAS Taxonomy

Human factors problems extend all the way from the operator to latent failures (Reason 1990). Thus, within the CIRAS system, human factors are broken down into three major categories. These are a) proximal factors, or mistakes that occur 'at the sharp end'; for example, in the driver's cab, or in the signal box; b) intermediate factors, which include maintenance, supervision, rules and communications; and c) distal factors which include procedures, design, management decisions and general company ethos / workforce culture. Within each of these three categories, the CIRAS system breaks event reports down into a finer grained and more specific set of codes. A diagrammatic representation of the human factors model is provided in Figure 12 below.

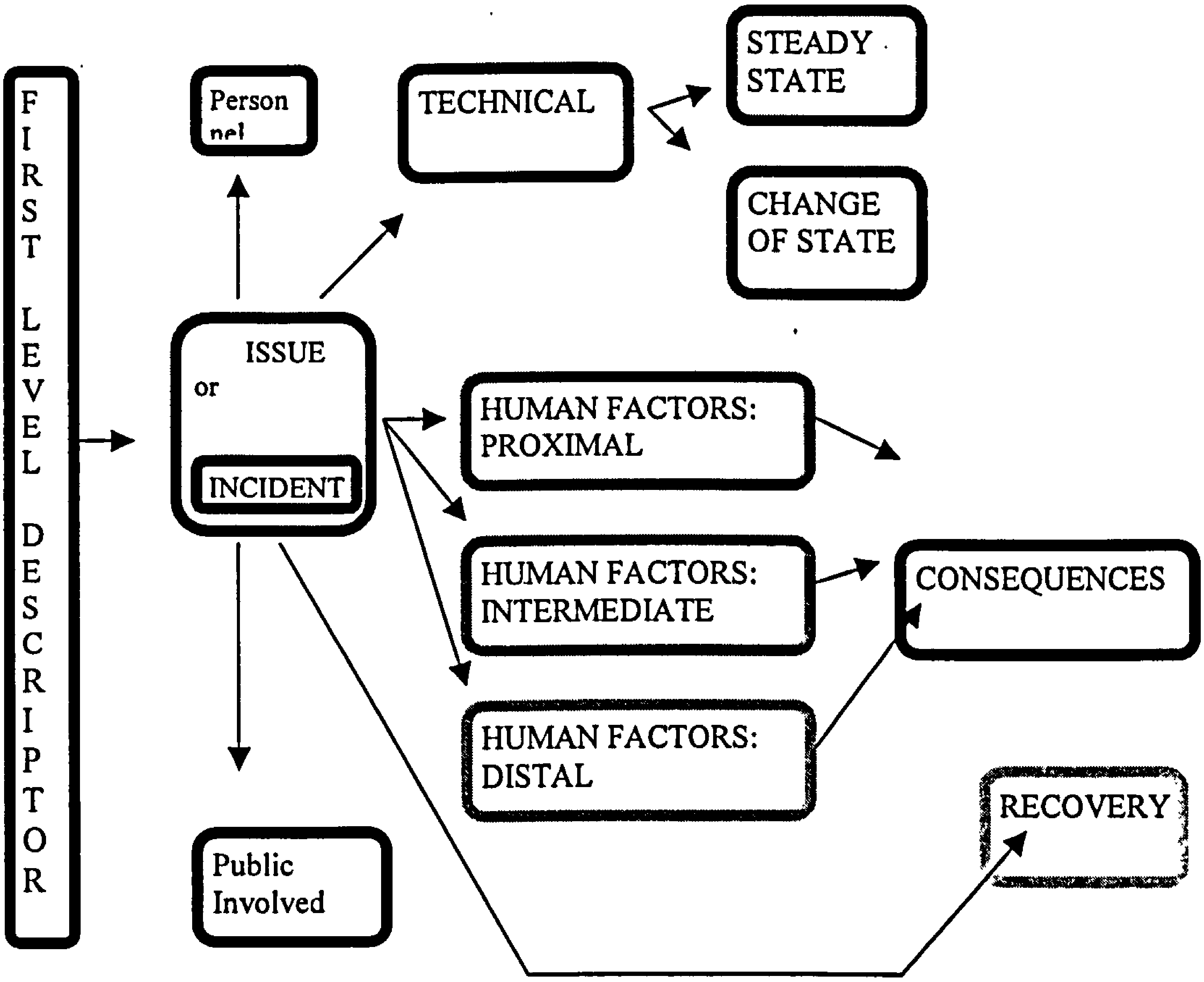


Figure 12: The CIRAS human factors model

The model evolved in collaboration with work being carried out by the Centre for Applied Social Psychology in both the UK rail and nuclear industries. The first step in the CIRAS coding process requires reports to be separated into *incidents* and *issues*. This enables all concerns to be analysed and for feedback to be provided. At interview staff often raise more safety issues than they have detailed on the initial report form or via the initial telephone call. An issue is defined as a concern about personnel or equipment failing to operate as planned but with no specific example given or concern about the behaviour of the general public which has safety implications but with no specific example given. An example of an issue from the CIRAS data is: radio communication should be used to control shunting movements within a depot where the physical layout prevents the driver from seeing the shunter. The radio batteries failed and were not replaced for some months, leading to unauthorised practices amongst drivers and shunters. An incident is defined as a specific example of a failure of personnel or equipment to operate as planned or a specific example of general public behaviour which has safety implications. As a rule an incident will be related in the first person (i.e. the reporter was present when the incident occurred) and will include dates, times and places. An example of an incident from the CIRAS data is: a driver was working nightshift followed by driving passenger trains early in the morning and briefly fell asleep, while travelling at 60 miles per hour, and it was the bell from the AWS magnet that woke him up.

Following the identification of the issues and incidents that comprise a report, the separate issues and incidents are assigned to a first level descriptor. The first level descriptor provides a brief explanation of the type of concern raised. This category is exclusive i.e. only one category may be assigned to each type of report made. A list of the first level descriptions with examples is provided in Table 11 below. This list emerged from discussions with managers in the railway industry and from the data collected and includes some categories (such as near miss, public behaviour) which are used by the railway industry. Issues and incidents are then categorised according to the CIRAS taxonomy.

Initial descriptor	**Example**
Potential/ actual failure to stop	Slides, station overshoots, SPADs
Near miss (on or near the line)	Potential collision with rolling stock, vehicle drivers at level crossings, P-Way staff, members of the public
Public behaviour	Vandals on or near the line, disruptive individuals on train or at stations
Environment	External/ internal conditions. E.g. rail contamination, lineside foliage, in cab features
Briefing/ training	Inadequate/ lack of information on site or arising from pre-arranged briefings
Supervision	Tasks or procedures poorly supervised
Equipment problems	Items/units (sets, radios, monitors, lighting) not working or not available, lack of maintenance, problems with PPE
Shift patterns/ rotas	Driving task at end of long shift, number of consecutive days worked, fatigue resulting from shift pattern
Adapted behaviour	Unauthorised action believed to be the best/ quickest way of completing task. Short cuts and bending of rules. Procedural drift.
Alcohol/ drugs impairment	Staff under the influence of drugs/ alcohol during the course of their work
BT Police response	Poor response or inadequate cover at key locations by British Transport Police
Slips/ lapses	Inadvertently omitting a step in procedure, forgetting equipment, slips, trips and falls.
Checking/ testing	Failure to check essential equipment or infrastructure, problems resulting from this process
Planning	Failure to adequately identify required features in any planning process, or discrepancies resulting from this activity
Other	Events which cannot be placed in any other category

Table 11: CIRAS first level descriptors

Reports are then coded in terms of two main channels, one dealing with human factors, the other dealing with technical failures or problems. Particular events can be coded through either or both of these channels depending on the causal factors. The data-base allows cross referencing of any combination of variables; and specific types of human error can be located for time, place, type of work etc. by cross-referencing technical and human-factors codes with additional data collected via the follow-up form.

4.9.1 Proximal Factors

Proximal causes of events are defined as causes associated with front-line operators e.g. drivers and signallers. These causes are closest in time and often place to the occurrence of an event. At this level the CIRAS model includes skill based, rule based and knowledge based errors (Rasmussen 1986 and Reason 1990). These and additional proximal or 'sharp end' factors are shown in the table below, with examples of the type of error made by the operator (as reported to CIRAS).

Proximal code	Definition	Example
Attention	Lack of concentration leading to slips (physical) and lapses (mental/ cognitive)	Failed to stop at station as singing in cab
Perception	Inability to see or hear specific features	Unable to see signal due to foliage
Knowledge	Lack of knowledge/ inadequate or incorrect knowledge for task	Trainee unaware shunting procedure not authorised
Rule violation	Deliberate breach of rules or procedures	Not using electrical protection when necessary
Rule based error	*Mistaken* use of wrong rule/ procedure in a given situation (note the action is mistaken but not *unintended* i.e. the person *meant* to do the act at the time but didn't *mean* to get the rules wrong).	Opened doors on the wrong side of the train, but in correct manner
Skill based error	A 'slip' or a 'lapse'/ an unintended action.	Driver slipped and fell when entering train
Communication between staff	Failure of communication between front line staff	Driver failed to pass on relevant information to signaller
Fatigue	Tiredness/ fatigue influencing behaviour	Train delayed as driver got out of train to walk along platform

Table 12: Proximal human factors

4.9.2 Intermediate factors

Intermediate causes of events are defined as occurring between proximal and distal causes. Staff involved are somewhat removed in time and place from the occurrence of an event e.g. supervisors and maintenance workers. Intermediate causes include the following: maintenance (inadequate or not carried out), communication failures from managers to staff and training. These and others are shown, with examples in Table 13 below.

Intermediate code	Definition	Example
Communication from staff to management	Failure of frontline staff to communicate with managers/ supervisors	Driver failed to report incident to supervisor
Communication from management to staff	Failure of supervisors to communicate with front line staff	Manager did not provide feedback on reported incident
Maintenance	Inadequate or absent repairs/ maintenance	Monitors not maintained
Training	Insufficient for task, not provided	Training not provided on rarely performed task (isolating doors)

Table 13: Intermediate human factors

4.9.3 Distal factors

Distal factors are defined as organisational or systems related failures. Staff involved are likely to be removed in time and place from the event e.g. managers and designers. The distal causes of events are shown, with examples in the table below.

Distal code	Definition	Example
Top down communication	Failure of senior managers to communicate with staff	Failure to fully explain implications of restructuring
Communication from staff to management	Failure of frontline staff to communicate with top managers/ supervisors	Failure to report information to top management
Procedures	Ambiguous, difficult to follow or absent	Procedure surrounding inoperative AWS open to interpretation.
Design of equipment	Equipment design not fit for purpose	Lifting equipment for removal of train doors inadequately designed
Rostering	Shift scheduling, staffing	Short staffing requiring overtime, rest day working
Objectives	Management priorities	Performance before safety
Culture	Attitudes of work force, macho-style, general company ethos	Get the job done culture

Table 14: Distal human factors

It is important that any coding system should be understandable to the managers who receive information from it. Therefore, CIRAS also has a classification system for technical/ equipment failures which is specific to the railway industry, which includes in-cab/ on-train equipment and infrastructure hardware e.g. signalling systems. This is described in the section below.

4.9.4 Technical codes

The technical codes used by the CIRAS system were originally developed when the system was implemented for Drivers only and therefore were most pertinent to train-related equipment. With the addition of signalling staff and maintenance contractors the technical codes were broadened to take account of the equipment used by such staff grades. The major re-organising of the technical codes took place following the nationalisation of the CIRAS system. While the technical contribution of equipment to incidents should not be overlooked, this aspect is indeed the one which official reporting channels is most familiar with. The first step in identifying technical problems is to identify the location of the equipment failure or design fault. The technical codes in the model are specifically related to the railway industry and include the following: 'sets' i.e. the train (passenger, ballast); 'track' i.e. on the running line; 'stations' at the area leading into or departing from or within the station itself; 'depot' i.e. where trains are maintained or stabled; 'other' anything which is not covered by the previous four locations. Then specific equipment types or groups are identified as either operating normally 'steady' or malfunctioning 'change'. Normally operating equipment that causes a problem is taken to have been inadequately or poorly designed or unfit for purpose. Technical codes should be used if the reporter outlines a technical item as being conditional for an incident or as being a hazard. Technical faults should only be coded if they impact upon the incident or issue in question, i.e. if they are causally related to the reported safety concern. For example, if a Driver is involved in a station overshoot due to misreading the timetable and the radio reception is faulty, there would be no technical code assigned. If however, the Driver experienced poor radio reception and was attempting to achieve a better reception by changing channels and therefore became distracted and lost situation awareness and overshot a station, then the faulty radio would be a contributory factor and thus coded as such. In this way technical factors in the database will only be technical problems which have contributed in some way to an unsafe situation or near miss.

The technical aspects of the CIRAS system are provided in table 15 below. Each of the items may be coded as either 'steady' or 'change'. This list of technical codes (location and equipment type) is not causal but descriptive of the location where the failure or design fault resides. Technical change, Sets, AWS (as an example code) informs the reader that failure of the AWS equipment on board the train was a causal factor in an incident, but does not make any statements about the cause of the failure of the AWS. Such a task requires technical knowledge and access to testing of equipment and expertise beyond the remit and competence of the CIRAS system and team. In testing the common cause hypothesis only the codes 'Technical Steady' and 'Technical Change' were used as the deeper level of description denotes only the site of the technical failure and therefore does not add anything to the testing of the hypothesis.

Location of equipment malfunction, failure or design problem	**Equipment type**
Sets (trains, passenger, freight, ballast)	Brakes AWS (automatic warning system) Lights WSP (wheel slip protection) Doors Radio Other
Track – on the line	Rails Overhead line Lights/ lighting Signs Other
Station	Lights
Other	Radios/ phones Doors Signs Signals

Table 15: Technical codes

4.9.5 Personnel involved

As part of the CIRAS coding structure, the designation of staff who have been involved in the reported issue or incident is recorded. Staff members (e.g. Drivers, Signallers) should only be coded in this category if they were involved in the incident being reported, i.e. were instrumental in causing the incident or adversely affected by it consequentially. For example a Driver may report P-Way staff working without high visibility protective clothing in the vicinity of a platform at the buffer end. Although the Driver is making the report, he is neither instrumental in causing the staff to be without their high visibility clothing and nor does it affect the Driver as s/he has not driven into the platform on which the P-Way gang are working. Staff categories include the following: Driver, Signaller, Maintenance, Supervisor, Conductor and Other. These categories are not used to test the common cause hypothesis and are detailed here to fully describe the CIRAS system. Following the expansion of the system in 2000, this category of code was greatly expanded to include all types of staff working for the railway industry.

4.9.6 Public behaviour and involvement

The CIRAS system also includes a category for incidents that have been largely caused by the public. Vandalism and trespass are of great concern to the railway industry and such incidents have become increasingly worrying – train accidents due to vandalism increased by 41% in 1995 – 1996 (Railway Safety 1996). The railway industry requires the mandatory reporting of incidents of vandalism and these are also passed to British Transport Police to enable them to concentrate resources in the most problematic areas. It is not the aim of CIRAS to become an alternative route for reports of vandalism and trespass. These types of report were often made to CIRAS following being reported through normal channels (often as a response to the lack of perceived action taken by supervisors and BTP). The taxonomy includes a set of very broad codes on where the behaviour of the public has caused impact on the driving function. These are defined as the following: On sets; On track; At stations; Other and the response/ actions of BTP. These data are not used to test the common cause hypothesis. All incidents involving the behaviour of the public were excluded from the analysis for all levels of the iceberg model (see Chapter 4 for a description of the data used to test the hypothesis).

4.9.7 Potential Consequences

CIRAS also seeks to look at the possible *consequences* of near misses or failures, and attempts to discover why accidents and incidents are sometimes *averted* or avoided altogether. This is done through a set of 'consequence' and 'recovery' codes. The potential consequences are only collected for incidents, not for issues. At interview staff making reports are asked by the CIRAS researcher what could have happened had the situation not been recovered in some way. Respondents may provide as many potential consequences as they think could have resulted from the incident being reported. Responses are not prompted in any way by the researchers and are the opinions of those involved in the situation at the time; this may result in potential consequences being exaggerated in an attempt to have the report treated with more seriousness than it warrants (see Chapter 1). The list of potential consequences was derived entirely from the staff during the course of interviews and are shown in Table 16 below. The potential consequences are not used in the analysis to test the common cause hypothesis.

Potential consequences related by staff at interview
Fatality
SPAD
Damage to rolling stock/ track/ other equipment
Discipline
Staff injury
Passenger injury
Collision with another train or buffer end
Derailment
Decrease in attention and concentration
Delay to services
Delay in the event of an emergency situation
Speeding (to make up time)
Fault not rectified or repaired
Station overshoot or failure to stop in time
Under reporting of the situation
Other

Table 16: Potential consequences

4.9.8 Recovery

An important aspect of any system such as CIRAS is to discover the reason why situations and near misses did not develop into fully-fledged accidents. CIRAS seeks to do this via a set of broad recovery codes (see van der Schaaf 1988, and van der Schaaf et al 1991 for a discussion on the benefits of recovery). At interview, reportees are asked why they think the situation they were involved in did not lead to a more serious consequence. The reasons given have been categorised into "procedures" i.e. situation retrieved by application of appropriate procedure; "individual action" i.e. situation retrieved by individual actions involving knowledge based or recognition primed decision; "technical" i.e. situation retrieved by automatic systems or technical barriers; "chance" i.e. situation retrieved by luck, guess work or trial and error. As for the potential consequences, recovery strategies are only gathered for incidents and not for issues. The successful recovery strategies can be used to communicate to staff the various ways that potentially serious incidents were avoided (Lucas 1991). While the recovery codes provide useful information for companies, they are not used in the testing of the common cause hypothesis, and are presented here to detail the full scope of the CIRAS model.

4.9 Coding reliability

Inter-rater reliability is a vital (and often neglected) part of any analysis system. Accordingly, to ensure the data coded are reliable, periodic inter-rater reliability trials were conducted both during the period when CIRAS was a local system and when it was operating at a national level. As this thesis is only concerned with data which were collected and analysed during the period 1996 – 2000, the reliability data described here pertain only to that time.

Inter-rater reliability trials were performed under strict conditions. Raters independently read and coded the reported incidents in the normal way, without discussing or comparing results until the trial was completed. A number of reliability trials were performed during the life cycle of the system. Index of concordance was around 80% for each trial (84.6% and 86.7% for two trials) using 10 randomly selected transcripts comprising 30 and 28 events respectively. The results of a third reliability trial are provided in this section. Reliability between two independent raters for the human factors coding (Technical, Proximal, Intermediate and Distal including all sub-codes) was performed for 10 randomly selected individual transcripts. Codes were assigned on the basis of reading the transcripts in the normal way and then applying the causal codes. Index of concordance was calculated using the following:

$$\frac{\text{Number of agreements x 100}}{\text{Number of agreements + number of disagreements}}$$

Reliability trial data: The data from the most recent reliability trial (1999) are shown in table 17 below. This resulted in an index of concordance of 85% at the last trial. For the coding of the initial descriptor of events, reliability checks were carried out by three independent raters, again using the index of concordance. The results were as follows: Rater 1 with Rater2 = 85.3%; Rater 1 with Rater3 = 88%; Rater 2 with Rater 3 = 85.3%.

These results suggest that we can be confident that the data are being coded in the same way by different raters.

Report No.	Coder 1	Coder 2	Agreements	Disagreements
103	Rule violation Communication manager to staff Procedures Objectives Track - change	Rule violation Communication manager to staff Procedures Objectives Track - change	5	0
102	Perception Procedures Objectives	Perception Procedures Objectives	3	0
81	Perception Rule based Comm between staff Procedures Lights – steady	Perception Knowledge Comm between staff Procedures Lights - steady	4	2
86	Perception Signal - steady	Perception Signal – steady	2	0
118	Perception Design Signals - steady	Perception Design Signals - steady	3	0
105	Communication between staff Rule violation Procedures	Communication between staff Rule violation Procedures	3	0
79	Communication between staff Procedures AWS - change	Communication between staff Communication manager to staff Procedures AWS - change	3	1
73	Perception Maintenance Communication manager to staff Tech Other – change	Perception Maintenance Communication manager to staff Procedures Tech Other – change	4	1
116	Communication between staff Procedures	Communication between staff Procedures	2	0
101	Attention Fatigue Rostering Procedures Objectives	Fatigue Rostering Procedures Objectives	4	1
88	Communication manager to staff Objectives Culture	Objectives Culture	2	1

Table 17: Results of last CIRAS reliability trial

4.10 How CIRAS meets the minimal requirements of a taxonomy

Chapter 2 discussed the minimal requirements of any taxonomy used to identify the causes of accidents or incidents in order to introduce prevention strategies and hence prevent recurrence. This chapter has described the way CIRAS collects data on safety concerns, the way the reports are processed and the analysis of the reports via the taxonomy. The CIRAS system is now compared to the minimum requirements of a taxonomy from the proposed list discussed in chapter 2.

Requirement	**CIRAS**
Theory Based on a suitable underlying theory of human behaviour	Based on GEMS and Active and Latent failures
Technical, human and organisational causes	Includes these categories
Reliability Be reliable i.e. that various independent analysts should reach the same conclusions	Frequent reliability trials. Latest trial result: 85%
Comprehensive Should be comprehensive in a qualitative sense (i.e. should be able to handle accidents, damages and near misses; cover both failures and recoveries)	Currently used in practice for near misses and unsafe behaviour, practices or situations. Is an appropriate method for both actual and potential incidents. Used for this thesis to analyse actual accidents and near misses. Suitable for both
Quantitative Be quantitative, enabling results to be computed and accumulated across many accidents	Data can be accumulated in a variety of ways (e.g. by company, by staff type, by month) and across all incidents reported.
Consequential Suggest ways of influencing and rectifying the root causes identified	CIRAS does not automatically generate suggestions for remedial action, although suggestions are provided in most cases.
Revealing Distinguish between events and underlying causes	The CIRAS system distinguishes between events, underlying causes and provides context information via the qualitative description of contextual factors.

Table 18: How CIRAS meets the minimal requirements of a taxonomy

As can be seen from the table above, the CIRAS analysis and taxonomy meets almost all of the minimal requirements. It is therefore justifiable to use this system to test the common cause hypothesis.

4.11 Summary and conclusions

This chapter has provided an overview of the background and development of the CIRAS system as used in the UK railway industry during the period 1996 – 2000. The methodological issues addressed in Chapter 3, were used as the basis for designing and implementing the system. The system developed from the philosophy that it is better to know what is happening at the 'coal face' in order to take remedial action, rather than to know who is responsible and start disciplinary procedures. The CIRAS methodology and data forms the basis for the test of the common cause hypothesis.

The system was mandated nationally for all companies in possession of a Railway Safety Case in June of 2000. Following 18 months of operation at the national level, the system continues to operate to the basic principles, although the analysis component is no longer performed according to the University of Strathclyde (UoS) model. The University of Strathclyde model was described in detail to provide a basic understanding of the type of human, technical and organisational factors, which can be identified using the system. This system is the basis for testing the common cause hypothesis: incidents from CIRAS reports, SPAD investigations and Formal Inquiries are coded using the University of Strathclyde human factors model prior to comparing the causal codes at the different severity levels. The UoS CIRAS taxonomy was compared to the minimal requirements of a taxonomy as proposed in Chapter 2, and fulfilled almost all of the minimal requirements. It is concluded therefore that the use of this taxonomy to test the common cause hypothesis is defensible.

5 Methodology and data collection

5 Overview

This chapter describes the data collection and selection, the assigning of incidents to the various severity levels used to test the common cause hypothesis and the assignment of the causal human and technical factors that were present during the course of the incidents.

5.1 Introduction to railway industry data

This section outlines the various types of data collected and investigated within the railway industry in order to provide understanding of the type of data available. Within different companies there may be different ways of investigating and analysing incidents. However, the majority of incidents are reportable to SMIS (Safety Management Information System), a database held by the railway body responsible for safety and standards - Railway Safety Limited. This allows analysis and comparisons to be made across the whole of the industry.

Within the railway industry, incidents are observed and investigated in a variety of ways. Incidents come to the attention of a company via automatic detection, self-reporting, reports by members of the public or by observation following the end of a shift (when damage to rolling stock or equipment is identified). A variety of incident types are collected and analysed by railway companies. These include the following:

Signals Passed at Danger (SPADs) without authority; trespass and vandalism incidents; suicides; occupational accidents and injuries; collisions; damage to property and assets; technical faults and failures such as Wrong Side Failures; delays and production issues and near misses.

This thesis concentrates on incidents that Drivers have been involved in, during the course of driving duties such as SPADs, collisions and near misses. For driving incidents, there are a variety of ways in which post-incident investigation is performed. These include Public Inquiries, Formal Inquiries and SPAD investigations.

In rare cases a Public Inquiry is held. This usually follows a major accident involving loss of life (e.g.Ladbroke Grove Rail Inquiry, 2001). Public inquiries were not used in this research as they often take place some time after the event and are performed by lawyers and judges. Problems with the judicial approach are discussed in Chapter 2.

Following a major incident or an incident with the potential for a major loss an internal Formal Inquiry may be performed. A Formal Inquiry may also include members from another railway company (if the incident involves staff from more than one company or results in serious damage to infrastructure) and is then called a

Joint Formal Inquiry (often Railtrack is the other partner in a Joint Formal Inquiry). A Formal Inquiry usually comprises a panel of railway managers and consists of evidence and interviews provided by the staff involved and witnesses, as well as technical and scientific evidence. In the case of a Formal Inquiry the company 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.

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. They are usually performed solely by the Driver Team Manager/ Supervisor (a manager of Drivers, with driving experience him/herself) who is responsible for the Driver involved. They consist of a report by the Driver and any relevant witnesses or related staff (such as Conductor, Guard or Signaller), and a conclusion of cause by the Driver Team Manager.

5.2 Data for the study

The data used in this study come from three sources: from Formal Inquiries, SPAD investigations and from CIRAS reports. Each of these sources will be described in detail in the following sections. Permission was granted from the railway company concerned to use the Formal Inquiry and SPAD investigation data, with agreement to keep the details of the individual incidents confidential. Permission to use the

CIRAS reports was granted from the CIRAS Steering Committee. All incidents come from the same company. Within the same company there is the same cultural background, same rules and procedures and the same staff. Thus the data are not contaminated by differences of culture, in operating procedures or rules, or between company differences in data collection, investigation and interpretation techniques.

It was intended to collect data from the various sources from the same time period (1996 – 2000). However, this resulted in too few Formal Inquiries, as Formal Inquiries are not performed as a matter of course. Therefore data from Formal Inquiries was extended to include the period 1990 – 2000. As far as the author is aware there were no important changes to the company rules and regulations during this extended period. All SPAD investigations and CIRAS reports came from the period from 1996 – 2000.

5.2.1 Formal Inquiries

Formal Inquiries are performed by a panel of highly expert railway managers (usually four or five plus a union representative) 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 the interviews are transcribed verbatim and included in the final report, along with 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. Due to the nature and seriousness of the Formal Inquiry process the author was not involved in these investigations. However, a railway company representative was available to discuss any aspects of the Formal Inquiry which were unclear or warranted further explanation.

A total of 66 Formal Inquiries were gathered from the company: this being the total number of Formal Inquiries completed and available during the period under study. Of the 66 Formal Inquiries collected, 20 reports were discarded. Reasons for discarding Formal Inquiries from the data set included the following:

- information was incomplete;
- the causal factors were not identifiable (e.g. if key staff did not survive the incident);
- the incident was caused by individuals other than railway staff (e.g. vandalism, inappropriate behaviour of vehicle drivers at level crossings);
- the cause of the incident was outwith the control of the company concerned (e.g. landslide).

These criteria resulted in 46 Formal Inquiries being suitable for use in the study. Of the twenty Formal Investigations that were discarded, ten involved the public behaving inappropriately at level crossings, three were caused by staff from other companies and did not relate to the driving function, six contained insufficient information to enable the coding of the incident according to the CIRAS taxonomy, and one was the result of a landslide that the company could not have prevented. Table 19 below provides a summary of the reasons for discarding Formal Inquiry data.

Reasons for excluding Formal Inquiry reports	**Number of Formal Inquiries**	**% of total Formal Inquiries collected (n = 66)**
Public behaviour	10	15%
Insufficient information for causal coding	6	9%
Not related to driving function	3	4.5%
Outwith control of company	1	1.5%

Table 19: Reasons for excluding Formal Inquiries

It is of interest and concern that 9% of the Formal Inquiries did not contain sufficient information for causal coding to be performed. This is not simply an indication of lack of depth or failure of investigation. It is also a reflection of the fact that the causes of some incidents remain unknown due to the specific circumstances. In the case of the death of a key witness the immediate causes of the incident will never be fully known although an Inquiry panel may speculate on what the employee was doing at the time. Further, in a number of the investigations collected, key witnesses provided conflicting evidence: in such cases it can be impossible to determine the sequence of events and the causes. In other cases the incident has been disputed and the investigation centres on proving that the incident took place rather than on determining causes.

5.2.2. Signal Passed at Danger (SPAD) investigations

SPAD (Signal Passed at Danger) investigations take place after a signal has been passed at red without authority. 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 the Driver Team 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 Team 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. However, these reports include the necessary data for the Railway Safety Ltd system called SMIS (Safety Management Information System). Often obtaining this mandatory data for SMIS drives the investigation of the incident. Due to the nature of SPAD investigations and resultant disciplinary action (three SPADs and a Driver is removed from driving duties) the author was not able to be involved in any of the SPAD investigations – reports were collected after they had been completed by the investigating manager.

The SPAD investigations used in this study were gathered after the investigation had been finalised. A total of 116 SPAD investigations were collected; this being the total number available from the company concerned. Of the 116 SPAD investigations collected, 88 were suitable for use in the study. Reasons for discarding SPAD investigations included the following:

- there was insufficient information to enable the coding process to be completed;
- the cause of the incident was a result of public behaviour (either vandalism or the behaviour of vehicle drivers at level crossings)

Two SPAD investigations were discarded as they were due to public behaviour, while 26 did not contain enough information to allow causal coding to be performed. None of the SPADs were outwith the control of the company and none were caused by employees of another railway-related company. A total of 28 SPAD investigations were discarded: Table 20 below provides a summary of the reasons for SPAD investigations not being included in the final data set.

Reasons for excluding SPAD investigations	**Number of SPAD investigations**	**% of total SPAD investigations collected (n = 116)**
Public behaviour	2	1.7%
Insufficient information for causal coding	26	22.4%

Table 20: Reasons for excluding SPAD investigations

As can be noted from Table 20 above, a much greater percentage of SPAD investigations than Formal Inquiries contained insufficient information to enable causal coding to take place. This is of some concern. Although all of the discarded SPAD investigations had railway mandated categories assigned to the incidents, there was little support from the reports for the conclusions reached. The reports usually consist of a minimal description of the incident (usually 10 – 12 lines) written by the Driver, and a summary by the Driver Manager. The Driver is interviewed and the information gathered during the interview process is not always detailed in the investigation report. Due to the discipline associated with SPADs - three infractions and the Driver is removed from driving duties - Drivers are often defensive when reporting the incidents and tend to provide a minimum of evidence. There are two main strategies used by Drivers following their involvement in a SPAD. The first is to blame technical failings (e.g. poor rail conditions or poor brakes). This results in a costly exercise of testing rails and brakes. If these alleged causes are not substantiated, the Driver is fully blamed and the consequences can be harsh if this is not the first offence. The second strategy used by Drivers is to take full responsibility for the SPAD, blaming it on a momentary lapse of concentration, in the hope that the potential disciplinary procedure does not focus on driving skill or re-training. Hence, the investigation may be hampered by the individual who has been involved. This results in an incomplete report. This is borne out by the number of SPAD investigations that contain insufficient information to allow causal coding. More than double the percentage of SPAD investigations than Formal Inquiry reports contained insufficient information for causal codes to be assigned. It is concluded that SPAD investigations are therefore not as comprehensive as Formal Inquiries. There are a number of possible ways to improve the completeness of the SPAD reports. The first is to establish a training course for the Driver Managers to help them to understand investigative processes and how to interview Drivers regarding determining causal factors. The second is to provide guidelines and flow charts on the type of information required and the questions to ask to get that information. The third is to use information from data recorders to establish the sequence of events. A final way to increase the completeness of reports regarding SPADs is to change the disciplinary policy relating to three SPADs – although this may not be desirable.

5.2.3 CIRAS reports

CIRAS reports were made voluntarily by 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 the Driver wishes to report. Following the initial report a critical incident interview was performed by the author with each Driver. Critical incident interviews are derived from the critical incident technique developed by Flanagan (1954). These interviews include details about the 'what', 'when', 'where', 'why' and 'how' of the incident. The CIRAS reports all came from the same company as the Formal Inquiries and SPAD investigations. CIRAS reports can be differentiated into two main types, namely issues and incidents. All issues were excluded from the data set used to test the hypothesis. Issues are recurring problems and are defined as: a concern about personnel or equipment failing to operate as planned but with no specific example given or concern about the behaviour of the general public that has safety implications but with no specific example given. An incident is defined 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. As issues are about on-going problems and not about a particular event, all issues were discarded. A total of 555 CIRAS event reports were received during the five year period of 1996 - 2000 from the target company. Of these 361 were related to on-going issues and so were discarded. Of the remaining 194 incidents, reports were discarded for the following reasons:

- the cause of the incident was a result of public behaviour (e.g. vandalism, harassment of the Driver by passengers, or the behaviour of vehicle drivers at level crossings)
- the report was related to the actions of BTP (British Transport Police)
- reports were not first hand accounts (i.e. the reporter was not the individual who was involved in the incident)
- insufficient information for causal coding to be performed (e.g. the reporter was unable to provide information on the role of others)
- the report related to an incident which would not lead to a near miss (e.g. reports on complaints regarding being asked to account for late running; the nature of interactions with other staff; disciplinary matters relating to Human Resources policies such as sick time)

Of the 194 incidents identified from the database, 106 were selected use in the study. A total of 34 incidents were discarded as they involved public behaviour. A further three related to the response of BTP and so were discarded. Twenty-eight CIRAS reports were not first hand accounts and related to hearsay about the behaviour of colleagues and were therefore not selected. Seven reports did not contain enough information to allow for full causal coding. Sixteen reports did not relate to a near miss situation. Table 21 below provides a summary of reasons for discarding CIRAS reports from the analysis.

Reasons for excluding CIRAS incident reports	**Number of CIRAS incident reports**	**% of total CIRAS incident reports collected (n = 196)**
Public behaviour	34	17.5%
Actions of BTP	3	1.5%
Not first hand accounts	28	14.4%
Insufficient information for causal coding	7	3.6%
Not a near miss situation	16	8.2%

Table 21: Reasons for excluding CIRAS incident reports

From Table 21 it can be noted that the largest number of reports discarded were due to public behaviour (including trespass and vandalism; inappropriate behaviour of vehicle drivers at level crossings; and behaviour of travelling passengers). 3.6% of CIRAS reports did not contain sufficient information for causal coding to take place. In the majority of cases this was related to the complicated nature of the report and the number of staff involved. A Driver is able to give a good account of his own role in an incident, and also to describe interactions with the Guard and Signaller. However, where contractors are involved (e.g. in a 'possession' of the line), the Driver is unable to provide insight into the methods of working and the rationale behind why decisions have been made. Likewise if a Driver has alleged an error by a member of Signalling staff, causal codes are merely speculative and so cannot be applied. It is also interesting to note that 14.4% of CIRAS reports relating to actual incidents come from Drivers discussing the role of their colleagues (i.e. not first hand reports) – such reports may provide useful information for individual companies, but they have not been used to test the common cause hypothesis, as again the basis for causal coding is speculative.

5.3 Severity levels selected to test the hypothesis

This section discusses the assignment of data to the different severity levels in order to test the hypothesis. Heinrich's original 1931 triangle model posited "disabling injury", "minor injury" and "no injury", while Bird's 1966 triangle suggested "serious or disabling injury", "minor injury", "property damage", and "no visible damage". In order to test the hypothesis data were assigned to one of three groups: 'fatality and injury'; 'damage'; and 'near miss'. These three groups were chosen as the best comprise of the data available and to include the property damage level suggested by Bird.

Fatalities were not included in either Heinrich's 1931 original iceberg model or Bird's 1966 iceberg model - the top of the triangle was disabling injuries. However, in a validation of the ratio data hypothesis, Salminen et al (1992) added fatalities to the top of their triangle. There is no valid reason why the common cause hypothesis is not equally applicable to fatalities as it is to disabling injuries. Further, the fatal incidents included also resulted in injury to other staff or passengers. However, as predicted by the iceberg model, there were insufficient numbers of either fatalities or disabling injuries to enable a meaningful analysis to be performed on serious injuries as a separate group. Further, the incidents resulting in injury did not contain sufficient information to distinguish between ‘major’ and ‘minor’ injury or between ‘serious’ and ‘disabling’ injury. Therefore all injuries were amalgamated into one group, namely ‘fatality and injury’. Further a number of incidents contained both injuries and damages. Such cases were assigned to the class of ‘fatality and injury’ as this is the higher consequence level in the triangle and does not contaminate the damage data.

Damage incidents included incidents that only resulted in damages and not any other consequences (e.g. injury or fatality).

Near misses were incidents that could have resulted in either property damage or injuries had the incident not been avoided by recovery or barriers.

5.4 Assigning the severity (consequence) level

The data were assigned to the different severity levels (fatality and injury, damage, and near miss) based on the actual consequences of the investigated incident.

5.4.1 Assignment of severity level to Formal Inquiries

The consequences of incidents resulting in a Formal Inquiry were assigned based on the actual consequences as detailed in specifically designed Appendices contained in each report. The Appendices contained information about staff and passengers who were killed or injured (although not always about the extent of the injuries sustained) and about the type of damage that occurred as a result of the incident. If neither damage nor injury resulted from the incident then this was indicated in the appropriate Appendices by the word 'nil'. Accordingly, where no damage or injury was sustained the incident was assigned to the 'near miss' category. Where damage was sustained the incident was assigned to the 'damage' category and where injuries were sustained the incident was assigned to the 'fatality and injury' category. In the case of both damage and injury occurring as a result of the incident, the incident was assigned to the 'fatality and injury' category as this is the most severe consequence in terms of the iceberg or triangle models.

Of the total of 46 Formal Inquiries selected as being appropriate for the study 17 were assigned to the category of 'fatality and injury', 18 to the category of 'damage' and 11 to the 'near miss' category.

5.4.2 Assignment of severity level to SPAD investigations

On the SPAD form that is completed by the investigating Driver Team Manager, there are tick boxes for the allocation of the actual and potential consequences. The 'actual consequence' category is commensurate with the severity levels chosen to test the hypothesis: fatality, injury and various levels of damage (such as collision, derailment, overhead line damage), and was therefore used to assign the SPAD incidents to 'fatality and injury', 'damage' and 'near miss' severity levels.

Of the total of 88 SPAD investigations selected as being appropriate for the study seven were assigned to the category of 'damage' and 81 to the category of 'near miss'. None of the SPAD incidents resulted in either fatality or injury.

5.4.3 Assignment of severity level to CIRAS data

The actual consequences of CIRAS reports are related by the reporter. None of the selected incidents reported to CIRAS resulted in damage or fatality, therefore all 106 CIRAS incidents selected were assigned to the 'near miss' severity level.

5.5 Final data sets selected

The following sections provide a brief summary of the data sets selected to test the common cause hypothesis.

5.5.1 Fatality and injury data set

The 'fatality & injury' data set all come from Formal Inquiries. A Formal Inquiry was rejected from this data set if it failed to have a fatality or injury as a consequence of the incident. Of the 46 Formal Inquiries selected, 17 resulted in a fatality or an injury to a member of the travelling public or a railway employee. No SPAD investigation data resulted in an injury or fatality. No CIRAS data resulted in an injury or fatality.

5.5.2 Damage data set

The damage data set comes from a combination of Formal Inquiries and SPAD investigations. A total of 25 incidents resulted in damage. These comprised 18 Formal Inquiries and 7 SPAD investigations. No CIRAS reports resulted in damage.

5.5.3 Near miss data set

Near misses come from all three types of data collected to test the common cause hypothesis; namely Formal Inquiries, SPAD investigations and CIRAS reports. All incidents (198 in total) used for this data set resulted in a near miss situation: there were no injuries or damages associated with the incidents. Formal Inquiries yielded 11 near misses; SPAD investigations yielded 81 near misses and CIRAS reports yielded 106 near misses.

5.5.4 Summary of data sets

The table below illustrates the source of the data (i.e. Formal Inquiry, SPAD investigation or CIRAS report) and the number of cases in each of the levels of the iceberg model used to test the common cause hypothesis.

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

Table 22: Data source and level of severity

5.6 Assigning the Technical and Human factors codes

Technical and human factors codes (Proximal, Intermediate and Distal) were assigned to all incident reports regardless of source in accordance with the UoS (University of Strathclyde) CIRAS human factors model as outlined in Chapter 3, Figure 12. The causal codes were assigned to the individual incidents on the basis of reading them thoroughly, constructing a causal tree of the event and then assigning the appropriate code(s). An example of an event tree and the codes assigned can be seen in Appendix 2. Causal codes were only assigned when it was clear from the reading of the text that the particular code contributed to the propagation of the incident. The causal coding was performed by the author who devised the UoS model in conjunction with the implementation of the CIRAS project and developing the methodology with which to test the common cause hypothesis. A reliability trial was also performed and will be presented in Chapter 5. Analysis was stopped when contributing causal factors could no longer be identified or when the following situations were encountered:

- information was lacking

- information was ambiguous

Codes were not assigned as either present or absent. Rather the number of times a code occurred was assigned (i.e. if two pieces of equipment failed, and thereby both were causal factors in the development of an incident, this would be coded as Technical Change x 2). This provides frequency data for each possible code.

5.7 Data limitations

There are certain inherent limitations to the data being used for this study. These are acknowledged and discussed in this section.

5.7.1 Data collection

Due to the company procedure the data have not all been collected in the same way. Both Formal Inquiries and SPAD investigations were collected from the company concerned following completion of the investigation and submission of the final report. Due to the nature of the incidents and the process of investigation, the author was not involved in the initial process of data collection. In both cases, the reporting of incidents was mandatory. CIRAS reports were collected by the author via a confidential and voluntary reporting system. All the CIRAS interviews used for this thesis were performed by the author.

5.7.2. Depth, goal and scope of incident investigation

While the same method has been used to analyse all the different types of data (i.e. the CIRAS system taxonomy; see previous chapter), the depth and level to which the incidents have been originally investigated varied according to the seriousness of the incident and the method of collection.

Formal Inquiries are investigated in much more detail than SPAD reports. Formal Inquiries consist of a panel representing the company involved and the infrastructure controller (Railtrack Zone) plus Union representation. At Formal Inquiries witness statements are presented and the witnesses are called to answer a series of questions from all panel members. These question and answer sessions are fully transcribed and available in the final report of the inquiry. The opinions of experts in traction and scientific results of adhesion and braking tests are also presented. The aim of the Formal Inquiry is to discover the root and underlying causes of the incident.

SPAD incidents are investigated only by the Driver Team Manager unless the incident is considered of such seriousness that a Formal Inquiry is deemed necessary. The investigation usually consists of an interview with the Driver and the Driver's report of the incident. If necessary scientific testing of the railheads and train brakes may be carried out (usually only in cases where the Driver complains of poor brakes or poor adhesion conditions). Witnesses such as the Guard or Conductor may also be asked to provide a written report regarding the circumstances of the incident. SPADs are automatically detected on almost the entire railway infrastructure and are required to be reported and investigated. Due to the discipline associated with SPADs - three infractions and the Driver is removed from driving duties - Drivers are often defensive when reporting these incidents and tend to provide a minimum of evidence. This results in a cursory report. This is borne out by the number of SPAD investigations that contain insufficient information to allow causal coding. More than double the percentage of SPAD investigations than Formal Inquiry reports contained insufficient information for causal codes to be assigned. Therefore it is concluded that SPAD investigations are not as comprehensive as Formal Inquiries.

There is no fully-fledged method to detect or record near misses or unsafe acts by the railway companies. In fact, in railway terms, a near miss is synonymous with a near collision between a train and trespassers or vehicles on the line, and not to other types of near miss incident. CIRAS provides a channel whereby reports of near misses, unsafe acts and general safety issues may be voluntarily and confidentially provided by railway employees. Due to the confidentiality of the system, witnesses cannot be interviewed, nor are scientific tests of braking capacity or rail adhesion carried out. Hence the report is fully from the perspective of the individual making the report, and is therefore potentially subject to individual bias and understanding of the situation.

Such differences in investigation methods may have an affect on the assigning of causal codes. The deeper the investigation, the more likely it is that more causes will be found. This of course presents problems for testing the common cause hypothesis: all of the more serious incidents (fatality and injury data set) have been investigated to the greatest depth. It is therefore a limitation of the study that all data was not collected and investigated in the same way.

5.7.3 Number of incidents

The number of incidents comprising the data sets 'fatality and injury' and 'damage' are much smaller than the number of incidents comprising the data set 'near miss'. In an ideal situation there would similar numbers in each of the three data sets. Given that all the data come from one company the research was limited by the number of incidents that occurred during the period under study. The time period for the collection of Formal Inquiries was extended in an attempt to address this problem. However, this still did not yield a commensurate number of serious incidents.

5.7.4 Discussion of the limitations

Despite these limitations, the type of data used for this thesis reflect the reality of the situation in the UK railway industry. Differences currently exist in the level of investigation performed on different types of losses: the larger the loss, the deeper, longer and more costly the investigation. This situation is unlikely to change. It is not thought to be cost effective to spend resources, time and manpower investigating minor events in a manner similar to the investigation of more serious events. There are currently insufficient staff to increase the depth of investigations. There is no system in the railway industry under which different types of events can be classified using one over-arching taxonomy of 'causes', although SMIS (Safety Management Information System) does collect a variety of different data and causes. The data are provided by the individual companies comprising the Railway Group. The reliability and validity of SMIS data (in terms of underlying and root causes) has yet to be demonstrated. Therefore, whilst the data used here are not ideal from a scientific point of view, they are a reflection of the real world situation (a situation that is unlikely to change very soon), and as such it is justifiable to use these data which have been coded using the same framework, despite being collected in different ways.

5.8 Possible ways to test the common cause hypothesis

This thesis aims to test the common cause hypothesis which proposes that the causes of near misses are the same as the causes of more serious events (property damage, injuries and fatalities). 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
- comparing the relative proportions of causal codes contributing to the various incident outcomes.

Each of these possible methods will be discussed in turn.

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. Such a method also does not enable weighting to be assigned to codes which occur most often.

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. Despite the fact that the analysis procedure may be the same for all consequence levels, it is clear that differences in the depth of investigation will impact on the results of the analysis (in terms of number of causes assigned). Incidents that are investigated in greater depth, are more likely to have more causal codes assigned than those which are less thoroughly investigated. Hence, unless all incidents being used to test the common cause hypothesis have been investigated to the same depth, using the actual frequencies of causes assigned is inappropriate at worst and misleading at best. In this thesis, incidents were investigated in three different ways: the more serious incidents were investigated via Formal Inquiries, which consist of a greater depth of investigation than either SPAD investigations or 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.

It is difficult to be sure whether a difference in number of codes assigned is an artefact of data collection (different styles of accident reporting and investigation) or whether it signifies a real difference in the number of causes that contribute to more serious events. As this cannot be determined, and has not been established by previous research, it is therefore misleading to use the actual frequencies, and instead it is more appropriate to test the common cause hypothesis by using the relative frequencies. By using proportions to test the common cause hypothesis, the problems mentioned above are thus avoided.

5.9 Summary and conclusions

Chapter 4 has provided an outline of the different types of data collected to empirically test the common cause hypothesis. The criteria for including data in the hypothesis testing procedure were provided and discussed (e.g. any incidents involving public behaviour – vandalism and trespass – were excluded). The criteria for each data type are comparable. A rationale for the severity levels ('fatality and injury', 'damage' and 'near miss') was provided. The data selected comprise the most robust set of fatalities and injuries, damages and near misses from the possible data available.

Inherent problems in the data (e.g. level and depth of analysis) are discussed, however it was concluded that such problems are a function and reflection of the 'real world' situation which pose problems for any research design using field accident data. Different data collection methods were used to collect the data which were used to test the common cause hypothesis (i.e. Formal Inquiries, SPAD investigations and data from CIRAS near miss reports), and it was demonstrated that different levels of information were gathered via these methods. The data suggested that SPAD investigations produce less complete reports than Formal Inquiries. It was suggested that this may be improved by training and aids for those performing SPAD investigations.

It was concluded that due to the confounding of data collection methods it is impossible to determine whether differences in the absolute frequencies of causal codes contributing to different severity levels is due to data collection methods or is a reflection of real differences in number of causes. It is therefore not applicable to test the common cause hypothesis using absolute frequencies where different collection and investigation techniques have been used as the basis for assigning contributory causes, but to use relative frequencies i.e. proportions.

Chapter 6 Results

6 Overview

This chapter provides details of the data analysis used to test the common cause hypothesis using causal factors data, which come from three data sets: 'fatality and injury', 'damage' and 'near miss'. The hypotheses are stated in full and a rationale is provided for the method of analysis.

6.1 Hypotheses

The UoS CIRAS human factors model is hierarchical. According to this model individual causal codes are subsumed under one of four top-level categories: 'Technical', 'Proximal', 'Intermediate' and 'Distal' which shall be called the 'macro' codes. The macro codes each comprise an exclusive set of individual causal codes, which shall be 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. For the sake of clarity, the macro and micro codes are shown again in Table 23 below.

Macro code	Micro code
Technical	Technical – Steady Technical - Change
Proximal	Attention Perception Knowledge based error Rule based error Skill based error Rule violation Communication between staff Fatigue
Intermediate	Communication from managers to staff Communication from staff to managers Maintenance Training
Distal	Communication from senior managers Communication to senior managers Procedures Design of equipment Staffing/ Rostering Objectives Culture

Table 23: An overview of the macro and micro causal codes of the UoS CIRAS model

6.1.1 Hypothesis one

The common cause hypothesis proposes that there are no significant differences between the causes of near misses and the causes of more serious incidents. At the general level of testing this hypothesis it is proposed that if the null hypothesis is true, the relative proportions of causal codes at the various macro levels (Technical, Proximal, Intermediate and Distal) contributing to the three severity levels will not be significantly different. The finding of significant differences would require the acceptance of the alternate hypothesis that the causes of near misses are not the same as the causes of more serious incidents.

6.1.2 Hypothesis two

Hypothesis two aims to test the common cause hypothesis at the specific level of individual causal codes. Hence, if the null hypothesis is true the relative proportions of micro codes contributing to the three severity levels will not be significantly different. The finding of significant differences would require the acceptance of the alternate hypothesis that the causes of near misses are not the same as the causes of more serious incidents.

6.1.3 Rationale for method of testing the hypothesis and the unit of analysis

Chapter 4 briefly discussed the two main ways in which the common cause hypothesis could be tested. It was argued that using the absolute frequencies of causal codes assigned to the number of incidents at each severity level is misleading. This is due to the confounding between methods of investigation of incidents and the number of causal codes determined to have contributed to the incidents. Differences in actual number of codes assigned (e.g. based on means) may highlight differences between 'fatality and injury', 'damage' and 'near miss'. However, it cannot be determined whether this is an artefact of data collection methods or a real difference in numbers of contributory causal codes. Further, such a difference does not necessarily disprove the common cause hypothesis – the validity of the hypothesis is not dependent upon the absolute number of causes to be the same at the various severity levels, but that the relative proportions are the same. Thus this thesis shall test the validity of the common cause hypothesis by using the proportion of causal codes at the various severity levels.

The unit of analysis in this thesis is not the incidents themselves at the various severity levels: the incidents are simply the data source for the causal codes. As the unit of analysis is the relative frequency of causal codes applied to each severity level, the actual number of incidents is not important. Chi-square will be applied to sets of proportions of causal codes that contribute to the various severity levels to determine if there are significant differences between groups (Fleiss 1981, Edwards 1972).

The Chi-square equation for comparing proportions is given by Fleiss (op cit) as the following:

$$\chi^2 = \frac{1}{pq} \sum_{i=1}^{m} n_{i.}(p_i - p)^2, \qquad \text{where } q = 1 - p.$$

Analysis will be performed at the level of both macro codes and micro codes.

6.2 Reliability

In Chapter 2 it was stated that reliability was a minimal requirement of any taxonomy aiming to identify the causal factors of incidents. Therefore, before analysis of the causal factors takes place, it is necessary to demonstrate the reliability of the coded data. To this end, two independent raters (experienced in using the coding scheme) coded a total of 14 incidents from various classes of event used in this study. Incidents were randomly selected from the three methods of investigations (Formal Inquiries, SPAD investigations and CIRAS reports). Four Formal Inquiries were coded, and 5 each of SPAD investigations and CIRAS reports. The reliability trial was performed under strict conditions. Raters independently read and coded the incidents in the normal way without discussing or comparing results until the trial was completed. Index of concordance was calculated using the following:

$$\frac{\text{Number of agreements} \times 100}{\text{Number of agreements} + \text{number of disagreements}}$$

The number of agreements between raters was 29 and the number of disagreements was 8. This resulted in an index of concordance of 78.4%. Reliability trial data are shown in Table 24 below.

Incident number and source	Coder 1	Coder 2	Agreements	Disagreements
20 (CIRAS)	Attention Communication between staff Objectives	Technical Steady Communication between staff	1	3
7 (Formal Inquiry)	Attention	Attention	1	0
146 (CIRAS)	Technical Steady Attention	Technical Steady Attention	2	0
13 (Formal Inquiry)	Technical Steady Rule violation Communication managers to staff Procedures Objectives Culture	Technical Steady Rule violation Procedures Objectives Culture	5	1
16 (Formal Inquiry)	Rule violation	Rule violation	1	0
2 (Formal Inquiry)	Perception Knowledge	Perception Knowledge Rule based	2	1
14 (SPAD)	Technical Steady Attention	Technical Steady Attention	2	0
23 (SPAD)	Perception	Perception	1	0
56 (SPAD)	Attention	Attention Rule violation	1	1
11 (SPAD)	Perception Communication between staff	Perception Communication between staff	2	0
96 (SPAD)	Attention	Attention	1	0
230 (CIRAS)	Perception Rule violation Maintenance Objectives	Perception Rule violation Maintenance Rostering Objectives	4	1
93 (CIRAS)	Perception Objectives	Perception Objectives	2	0
277 (CIRAS)	Perception Rule violation Communication between staff Culture	Perception Rule violation Communication between staff Training Culture	4	1

Table 24: Results of Reliability trial

6.3 Overview of the data

This section provides a brief overview of the data. A total of 78 individual causal codes were assigned as contributing factors in the 17 incidents that were classified as 'fatality and injury'. For the 25 'damage' incidents a total of 100 individual causal codes were assigned. The 198 'near miss' incidents were assigned 572 individual causal codes. Table 25 below provides a summary of this information.

Severity level	No. of incidents	Mean codes assigned	Mean proximal	Mean intermediate	Mean distal	Mean technical
Fatality & Injury	17	4.58	2.64	0.58	1.12	0.24
Damage	25	4	2.32	0.32	1.08	0.28
Near miss	198	2.88	1.58	0.16	0.73	0.41

Table 25:Mean number of codes assigned by severity level

Table 25 above shows that there are a mean of 4.58 causal codes assigned to the 'fatality & injury' data set, 4 to the 'damage' data set and 2.88 to the 'near miss' data set. There are therefore more causal codes in total assigned to fatality and injury incidents than to damage or near miss incidents, with fewest causal codes being assigned to near misses.

6.4 Type I errors

Due to the nature of the experiment, testing the hypothesis requires a large number of pre-planned comparisons using Chi-square. In order to reduce the incidence of Type I errors (i.e. erroneously detecting a significant difference) Bonferroni corrections were performed. A significance level of 0.05 was chosen and then corrected for each set of planned comparisons.

6.5 Testing hypothesis one: comparison of the proportions of macro causal codes

The graph below shows the distribution of the occurrence of Technical, Proximal, Intermediate and Distal factors (the macro level) across the different groups of 'fatality & injury', 'damage' and 'near miss' as a percentage of the total causes assigned to each of the individual severity levels.

As can be seen from the graph below the percentages of causal codes assigned to the various levels of fatality & injury, damage and near miss are very similar at the macro level of the coding hierarchy (i.e. at the Technical, Proximal, Intermediate and Distal levels). It can be noted that the percentage of Technical codes increase as the consequences become less severe: i.e. a greater proportion of the total number of codes assigned are assigned to technical codes for near misses than to either damage or fatality & injury incidents.

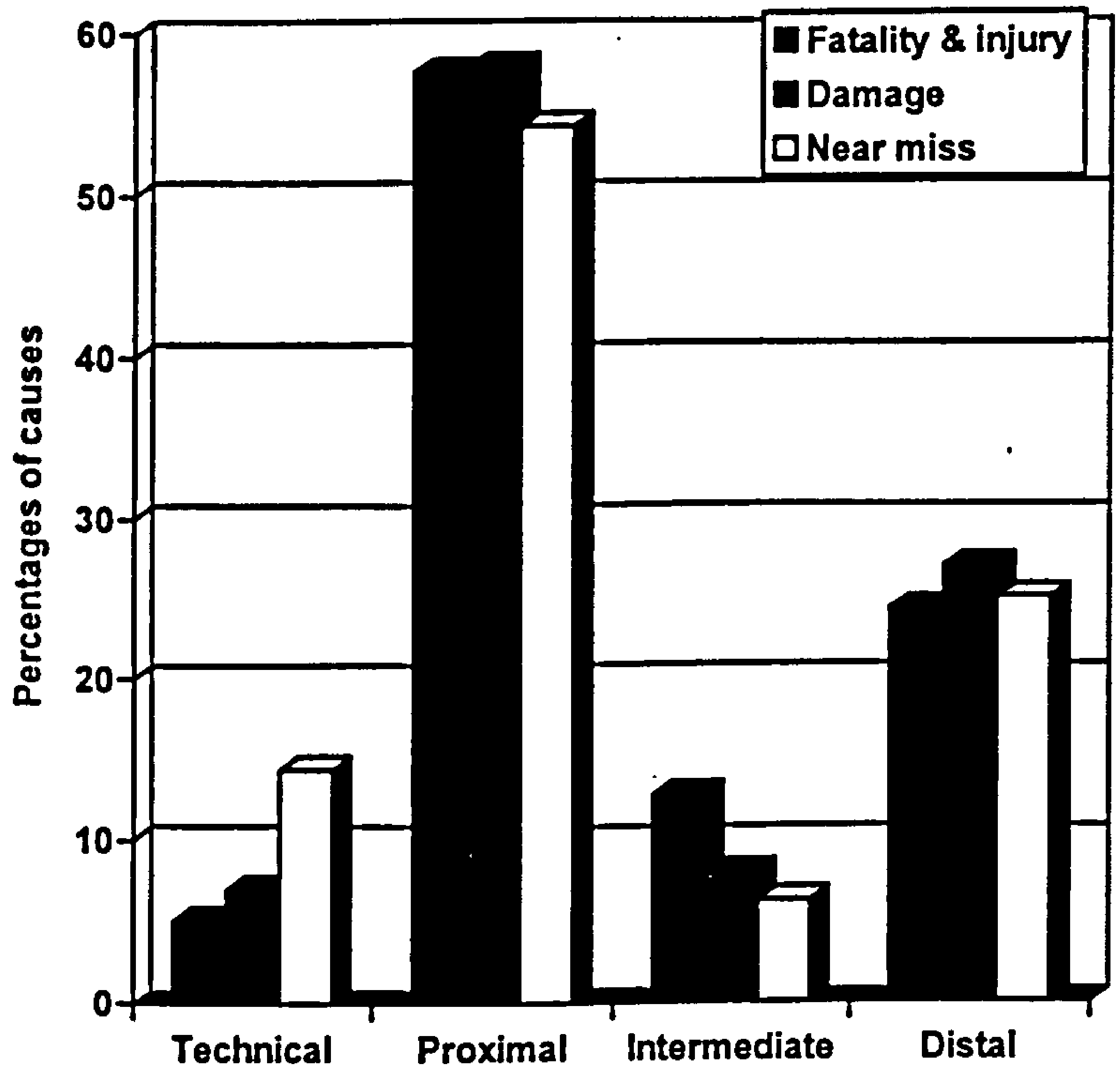


Figure 13: Macro causal factors by level of severity

Proximal and Distal codes appear to be fairly stable across the different consequence levels, while Intermediate codes show the opposite pattern from the Technical codes i.e. they occur less frequently as the consequences become less serious. Proximal codes are the most frequently occurring human factors causes regardless of the severity of consequence.

The macro level of the coding hierarchy comprises a total of 21 individual causal codes. Technical comprises two codes, Proximal eight codes, Intermediate four codes and Distal seven codes. The fact that by far the greatest percentage of codes are Proximal in all levels of severity, may be a reflection of the greater number of individual codes which comprise that category, rather than a reflection of Proximal codes being more frequent in contributing to incidents.

In order to determine if there are any differences in the proportions of causal codes which contribute to the severity levels a number of chi-square tests were performed on the proportions of codes in the levels of Technical, Proximal, Intermediate and Distal. As four comparisons were performed, the significance level is corrected using the Bonferroni procedure thus: 0.05 / 4 = 0.0125. The results of these chi-square analyses are provided in Table 26 below.

Macro code	**Severity levels being compared**	**Chi-square**	**d.f.**	**Significance level**
Technical	Fatality & Injury, Damage and Near miss	6.26	2	P > 0.025, N.S.
Proximal	Fatality & Injury, Damage and Near miss	0.536	2	P > 0.05, N.S.
Intermediate	Fatality & Injury, Damage and Near miss	6.159	2	P > 0.025, N.S.
Distal	Fatality & Injury, Damage and Near miss	0.189	2	P > 0.05, N.S.

Table 26: Results of Chi-square analysis for proportions of macro causal codes by severity levels

The results of this first set of Chi-square tests show that there are no significant differences in the proportions of codes for Technical, Proximal, Intermediate and Distal codes for the three levels of severity. As no significant differences were found, no further follow-up analysis will be performed

6.5.1 Summary of results for hypothesis one

Hypothesis one aimed to test the null hypothesis that the contribution of proportions of the macro codes (Technical, Proximal, Intermediate and Distal) was not significantly different for each of the different severity levels. The results identified no significant differences in the proportions for any of the macro causal codes between the different severity levels.

6.6 Testing hypothesis two: comparison of the proportions of micro causal codes

The following section tests hypothesis two. This analysis compares the proportion of causes which contribute to the different severity levels from the total number of causal codes assigned per consequence type. These analyses are performed separately for the various sets of micro causal codes which comprise the macro categories.

6.6.1 Comparison of proportions of Technical micro codes

Below the macro category of 'Technical', a further classification of 'steady state' (i.e. equipment operating as it should) or 'change of state' (i.e. technical failure) can be assigned. The graph below provides the distribution of the two individual technical causal codes for fatality & injury, damage and near miss. The numbers shown are the percentages of the total codes assigned to each of the consequence levels.

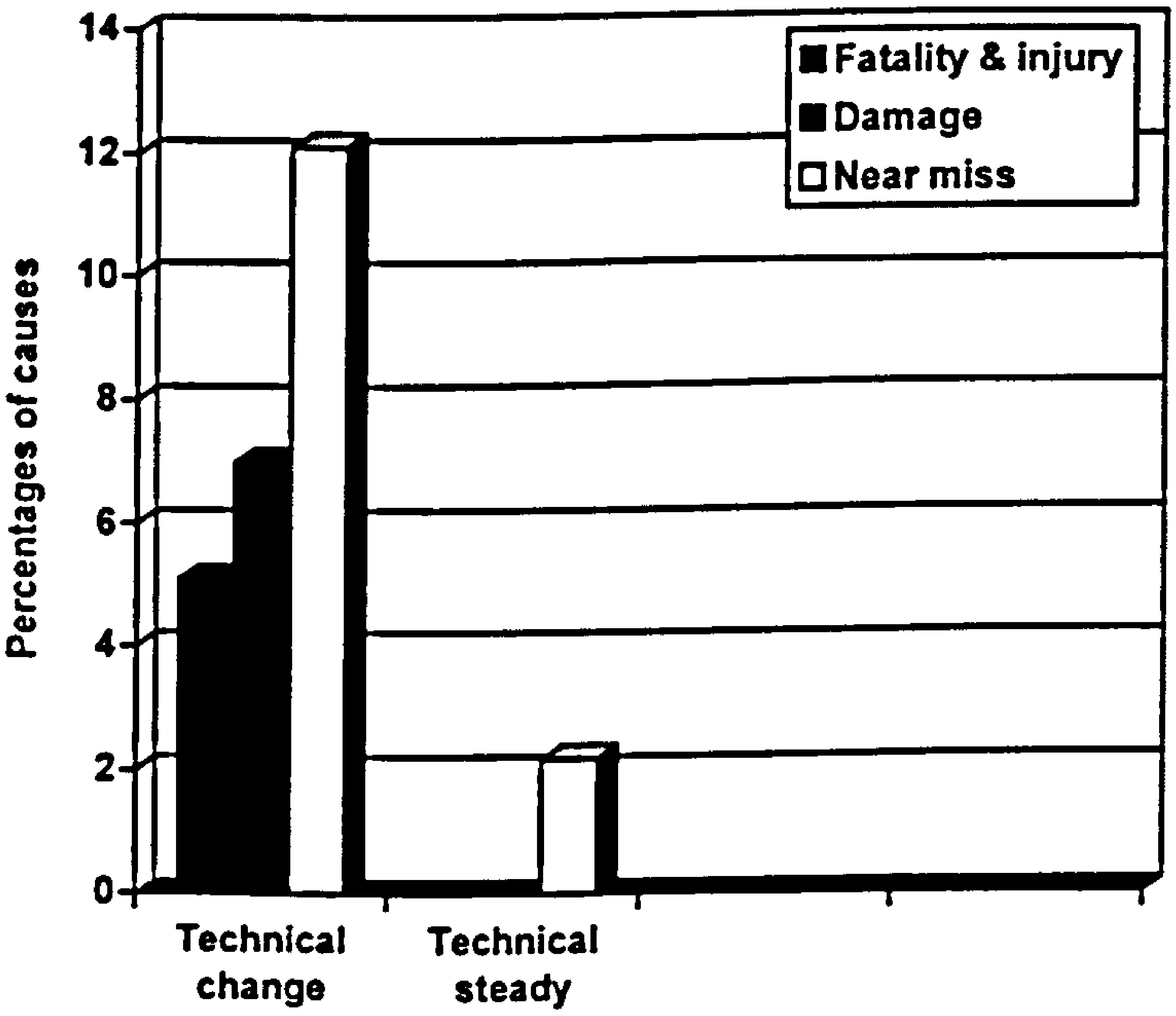


Figure 14: Micro technical codes by level of severity

As can be seen from the above graph, the technical change codes assigned increase as the consequences become less serious. Further, only the 'near miss' incidents have equipment coded as 'steady state'. 'Near miss' incidents have more than twice as many equipment codes as 'fatality & injury' and almost twice as many as 'damage' incidents.

The table below presents the results of Chi-square analysis for the difference in proportions of individual technical 'change' codes contributing to fatality & injury, damage and near miss. As there were no codes assigned for technical 'steady' for 'fatality and injury' and 'damage' Chi-square could not be performed. Therefore, Fisher's Exact test was used for three paired comparisons ('fatality and injury' with 'damage'; 'fatality and injury' with 'near miss'; and 'damage' with 'near miss').

Technical micro level codes	**Severity levels being compared**	**Chi-square**	**d.f.**	**Significance level**
Technical change	Fatality & injury, damage and near miss	5.26	2	p > 0.05, N.S.

Table 27: Chi-square analysis of Technical change micro code

Table 27 above shows that there are no significant differences between the proportions of the technical 'change' code that contribute to fatality & injury, damage or near miss.

Technical micro level codes	**Severity levels being compared**	**Two-tailed significance level (Fisher's exact)**
Technical steady	Fatality & injury, and damage	No data
Technical steady	Fatality & injury, and near miss	P = 0.383, N.S.
Technical steady	Damage and near miss	P = 0.233, N.S.

Table 28: Fisher's exact analysis of Technical steady micro codes

Table 28 above, shows that there are no significant differences between 'fatality & injury' and 'near miss' or between 'damage' and 'near miss' for the technical 'steady' code. Neither 'fatality & injury' nor 'damage' had the technical 'steady' code assigned, and therefore comparison between these two severity levels was not necessary.

6.6.2 Comparison of proportions of Proximal micro codes

There are eight possible factors that comprise the Proximal macro category. These are 'attention', 'perception', 'fatigue', 'knowledge based errors', 'skill based errors', 'rule based errors', 'rule violation' and 'communication between staff'. The distribution of these eight individual Proximal causal factors across the three levels of fatality & injury, damage and near miss is shown in the graph below. The numbers shown are percentages of the total number of codes assigned to all incidents in each of the consequence levels.

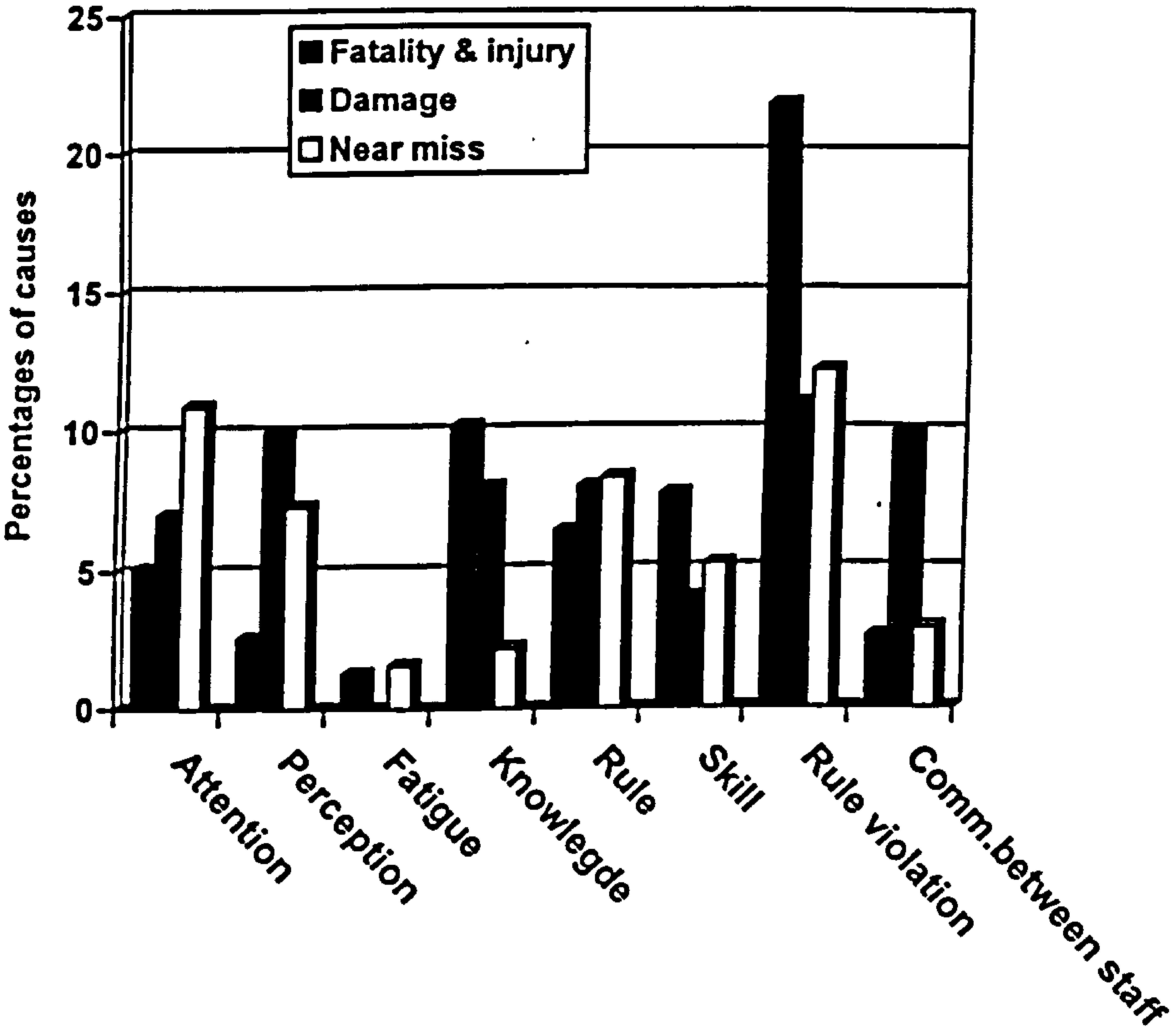


Figure 15: Micro proximal codes by level of severity

From the graph above it can be seen that the contribution of 'attention' codes to incidents increases as the severity level decreases. Knowledge based errors are more frequently found in incidents resulting in 'fatality & injury' and least frequently in 'near miss' incidents: this is the opposite pattern as for the attention category. Perception codes peak for 'damage', and are least for 'fatality & injury'. This is the not the only factor which follows this pattern. Communication between staff follows a similar pattern. Rule violations peak for 'fatality & injury'. Fatigue is only present in 'fatality & injury' and 'near miss' and not 'damage'; the percentages are similar for 'fatality & injury' and 'near miss' incidents.

The table below presents the results of Chi-square analysis for the difference in proportions of individual proximal codes contributing to fatality & injury, damage and near miss incidents. Given that eight comparisons have been performed, the significance level is 0.05 / 8 = 0.00625.

Proximal code	**Severity levels being compared**	**Chi square**	**df**	**Significance level (2 tailed)**
Attention	Fatality & injury, damage and near miss	4.237	2	P > 0.025, N.S
Perception	Fatality & injury, damage and near miss	3.65	2	P > 0.05, N.S.
Fatigue	Fatality & injury, damage and near miss	1.64	2	P > 0.05, N.S.
Knowledge based error	Fatality & injury, damage and near miss	16.854	2	P < 0.001
Rule based error	Fatality & injury, damage and near miss	0.237	2	P > 0.05, N.S.
Skill based error	Fatality & injury, damage and near miss	1.249	2	P > 0.05, N.S.
Rule violation	Fatality & injury, damage and near miss	6.076	2	P > 0.025, N.S.
Communication between staff	Fatality & injury, damage and near miss	3.9	2	P > 0.05, N.S.

Table 29: Results of Chi-square for proximal causal codes

At the level of individual Proximal causal codes only Knowledge based errors are significantly different. In order to determine between which two groups the proportions differ for knowledge based errors, further Chi-square analyses were performed between the pairs 'fatality & injury and damage', 'fatality & injury and near miss' and 'damage and near miss'. Given that three comparisons have been performed, the significance level is 0.05 / 3 = 0.0166. The results of these analyses are provided below.

Proximal code	Severity levels being compared	Chi square	df	Significance level (2 tailed)
Knowledge based	Fatality & injury and damage	0.282	1	P > 0.05, N.S.
Knowledge based	Fatality & injury and near miss	14.629	1	P <0.000
Knowledge based	Damage and near miss	9.2	1	P < 0.001

Table 30: Results of chi-square for combinations of severity level by Knowledge based causal code

Based on the results above the proportion of knowledge based codes contributing to the pair 'fatality & injury' and 'near miss', and the pair 'damage' and 'near miss' are significantly different. There are no significant differences between the pair 'fatality & injury' and 'damage'.

6.6.3 Comparison of proportions of Intermediate micro codes

There are four factors that comprise the Intermediate macro category. These are 'maintenance', 'communication from manager to staff', 'communication from staff to managers', and 'training'. The distribution of these factors across the three levels of 'fatality & injury', 'damage' and 'near miss' is shown in the graph below.

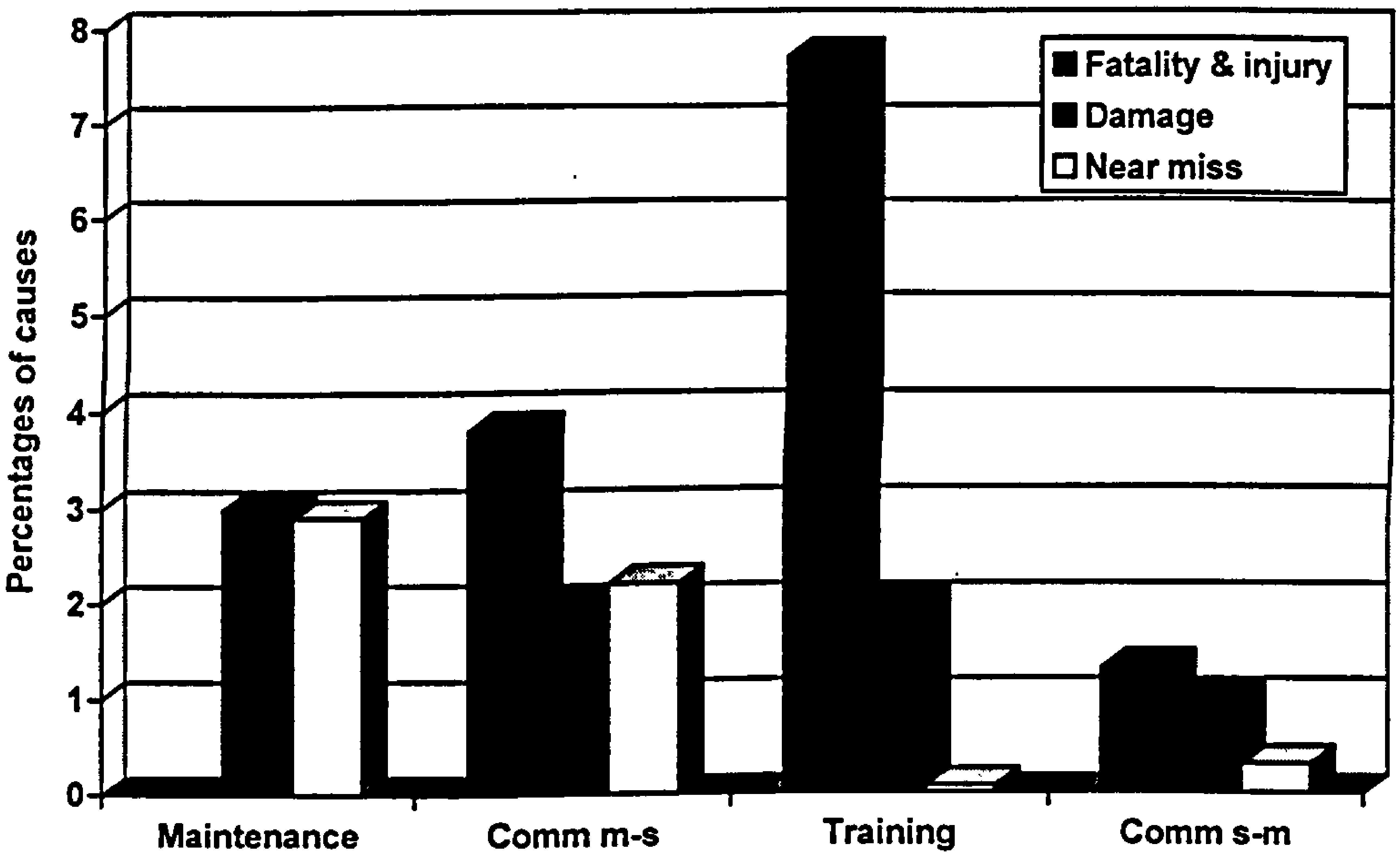


Figure 16: Micro intermediate codes by level of severity

The numbers shown are percentages of the total number of codes assigned to all incidents in each of the consequence levels. The graph below shows that the causal code maintenance has not contributed to any of the incidents resulting in 'fatality & injury', although similar percentages of codes contribute to both damage and near miss incidents. It can be also seen that training causal codes contribute to 'fatality & injury' incidents more frequently than to either 'damage' or 'near miss'. This is the same pattern as for rule violations at the micro Proximal level.

The table below presents the results of Chi-square analysis for the difference in proportions of individual Intermediate codes contributing to 'fatality & injury', 'damage' and 'near miss' incidents. Given that four comparisons have been performed, the significance level is 0.05 / 4 = 0.0125.

Intermediate code	**Severity levels being compared**	**Chi square**	**df**	**Significance level (2 tailed)**
Maintenance	Fatality & Injury, Damages and Near misses	0.233	2	$P > 0.05$, N.S
Communication manager to staff	Fatality & Injury, Damages and Near misses	0.759	2	$P > 0.05$, N.S
Communication staff to manager	Fatality & Injury, Damages and Near misses	5.38	2	$p > 0.05$, N.S
Training	Fatality & Injury, Damages and Near misses	17.89	2	$p < 0.001$

Table 31: Results of Chi-square analysis for Intermediate causal codes

Only the causal code training is significantly different in terms of the proportions for 'fatality and injury', 'damage' and 'near miss'. In order to determine which pairs of consequence outcome ('fatality & injury' and 'damage', 'fatality & injury' and 'near miss', and 'damage' and 'near miss') are significantly different, further Chi-square analyses were performed. Given that three comparisons have been performed, the significance level is 0.05 / 3 = 0.0166. These comparisons are shown in Table 32 below.

Intermediate code	**Severity levels being compared**	**Chi square**	**df**	**Significance level (2 tailed)**
Training	Fatality & injury and damage	3.39	1	P > 0.05, N.S.
Training	Fatality & injury and near miss	22.73	1	P < 0.000
Training	Damage and near miss	1.64	1	P > 0.05, N.S.

Table 32: Results of chi-square for combinations of severity level by Training causal code

The proportions of training causal codes contributing to fatality & injury and near miss incidents are significantly different.

6.6.4 Comparison of proportions of Distal micro codes

There are seven possible factors that comprise the Distal macro category. These are design, procedures, objectives, staffing/ rostering, top-manager to staff communication, staff to top-manager communication and culture. The distribution of these factors across the three severity levels ('fatality & injury', 'damage' and 'near miss') is shown in Figure 17 below. The numbers shown are percentages of the total causal codes assigned to all incidents in each of the consequence levels.

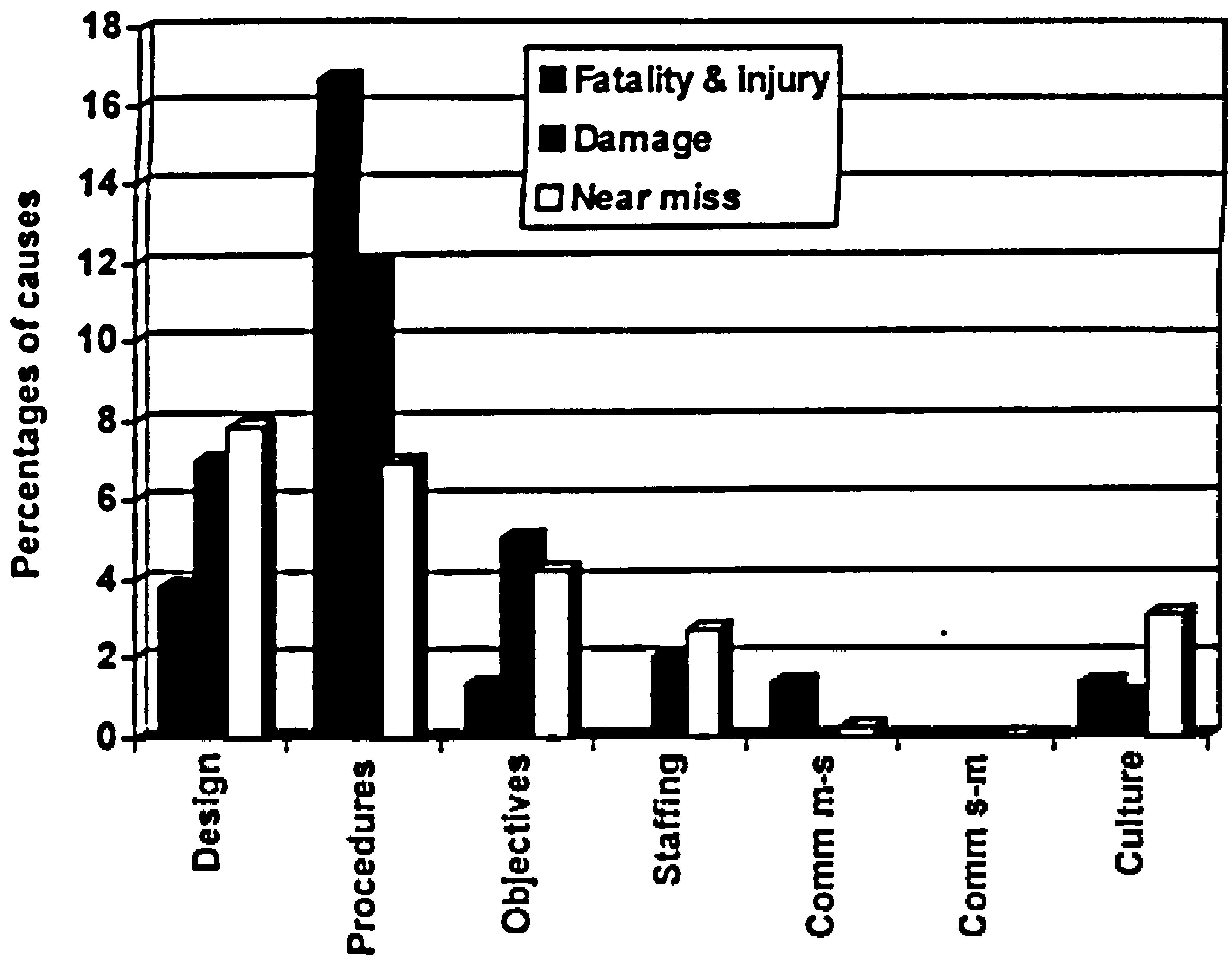


Figure 17: Micro distal codes by level of severity

From the graph above it can be seen that design causal codes increase in frequency as severity level decreases (i.e. more design causes occur at the level of 'near miss' than at either 'damage' or 'fatality & injury'). The opposite is observed for procedural codes – as severity increases so does the frequency of contribution of procedures to incidents. Objectives peak for 'damage' incidents. Staffing is present for 'damage' and 'near miss' incidents though not for 'fatality & injury' incidents. Communication from top-managers to staff is only present in a small percentage of 'near miss' incidents and not for either 'damage' or 'fatality & injury' incidents, while communication from staff to top-managers is not present in any of the severity levels. Culture is most frequent in 'near miss' incidents.

The table below presents the results of Chi-square analysis for the difference in proportions of individual distal causal codes contributing to injuries, damages and near misses. As five comparisons were performed, the significance level is 0.05 / 5 = 0.01.

Distal code	Severity levels being compared	Chi square	df	Significance level (2 tailed)
Design	Fatality & Injury, Damages and Near misses	1.73	2	p > 0.05, N.S
Procedures	Fatality & Injury, Damages and Near misses	9.97	2	p < 0.01
Objectives	Fatality & Injury, Damages and Near misses	1.78	2	p > 0.05, N.S.
Staffing	Fatality & Injury, Damages and Near misses	2.419	2	p > 0.05, N.S.
Communication from managers to staff	Fatality & Injury, Damages and Near misses	No data	No data	No data
Communication from staff to managers	Fatality & Injury, Damages and Near misses	No data	No data	No data
Culture	Fatality & Injury, Damages and Near misses	1.918	2	P > 0.05, N.S.

Table 33: Results of Chi-square for distal causal codes

Only procedures have significantly different proportions contributing to 'fatality & injury', 'damage' and 'near miss' incidents. Further Chi-square analyses were performed between the various paired combinations of consequences to determine where these differences lie. As three comparisons were performed, the significance level is 0.05 / 3 = 0.0166. The results are provided in Table 34 below.

Distal code	Severity levels being compared	Chi square	df	Significance level (2 tailed)
Procedures	Fatality & injury and damage	0.772	1	P > 0.05, N.S.
Procedures	Fatality & injury and near miss	8.83	1	P < 0.005
Procedures	Damage and near miss	3.11	1	P > 0.05, N.S.

Table 34: Results of chi-square for combinations of severity level by Procedures causal code

From Table 36 above, the results demonstrate that the proportion of procedure causal codes contributing to 'fatality & injury' and 'near miss' incidents is significantly different.

6.6.5 Summary of results for hypothesis two

Hypothesis two aimed to test the null hypothesis that the contribution of proportions of the micro codes (i.e. the individual codes that comprise the superordinate categories of Technical, Proximal, Intermediate and Distal) was not significantly different for each of the different severity levels. The results identified no significant differences in the overall proportions of individual Technical causal codes. There were significant differences at the level of Proximal, Intermediate and Distal codes. At the level of individual Proximal causal codes differences in proportions were found for knowledge based errors: these differences existed between the severity levels 'fatality & injury' and 'near miss' and 'damage' and 'near miss'. At the level of individual Intermediate causal codes proportions were significantly different for training between the severity levels of 'fatality & injury' and 'near miss'. While at the individual Distal causal codes the only significant difference identified was between 'fatality & injury' and 'near miss' for the causal code procedures.

6.7 Summary of results

At the more general level of testing the common cause hypothesis there were no significant differences found (i.e. for the Technical, Proximal, Intermediate and Distal codes).

From a total of 21 individual causal codes, only three were found to be significantly different: these were knowledge based errors, training and procedures. Knowledge based errors were significantly different between 'fatality & injury' and 'near miss', and 'damage' and 'near miss' at the individual level of Proximal codes. At the individual Intermediate level of coding significantly different proportions were found for the causal code 'training' for 'fatality & injury' and 'near miss' only. At the Distal level, significantly different proportions were found for the causal code 'procedures' between the pair 'fatality & injury' and 'near miss'.

Chapter 7 Discussion

7 Overview

This chapter provides a discussion of all the issues that have been raised throughout this thesis and of the results of the hypothesis testing.

The chapter begins by recapitulating the general problems of testing the common cause hypothesis and the failure of research to date to have adequately tested the hypothesis.

Issues relating to analysing data using a taxonomy are presented and discussed. Relevant aspects for designing and implementing reporting systems are summarised and discussed. Recommendations relating to the objectives of collecting and analysing near miss data are presented.

The results presented in Chapter 6 are discussed in detail and possible reasons for the significant differences found are presented. It is concluded that the common cause hypothesis is more supported by the results of this thesis than refuted. However, as this thesis was restricted to testing the hypothesis in one domain (the railway industry) subsequent research should address the validity of the common cause hypothesis across different domains. The limitations of the data sets used to test the common cause hypothesis are presented and their implications discussed.

7.1 Problems of testing the common cause hypothesis

Although Heinrich proposed the common cause hypothesis in 1931, his only confirmation for the proof of the hypothesis was based upon evidence relating to percentages of causes (this is shown in graphical form in Figure 3). As Heinrich was a proponent of a rather simple model of single cause, only the 'immediate cause' (which could be remedied) was actually used in supporting the case of the common cause hypothesis. It is unclear who assigned the causes or upon what basis that one cause which could be reversed (via remedial actions) was chosen. Nevertheless, the hypothesis has gained support. Despite Heinrich's original thoughts on similar causal pathways for serious accidents and near misses, the idea was not developed further (Chapter 1 shows that in fact Heinrich and his co-authors went to great lengths to convince readers to the contrary in the final edition of Industrial Accident Prevention). It is unfortunate then that Heinrich did not further develop the theory or discuss ways in which the hypothesis may be proven or refuted. The theory has been misunderstood, misinterpreted and confused with the ratio relationship between serious accidents, property damage incidents and near misses. Therefore research to date has concentrated on testing a confounded view of the common cause hypothesis or has tested the hypothesis using inadequate/ inappropriate data.

Furthermore, research using appropriate data to test the hypothesis has concentrated on the absolute numbers of causes assigned to serious accidents and less serious accidents or near misses. There are a number of problems associated with using the actual frequencies of contributory causes: these are related to the type and depth of investigation performed on the different levels of consequence. Chapter 2 discussed the issue of depth of investigation and biases inherent in investigation and coding of causes. If different numbers of causes are assigned to serious accidents, property damage incidents and near misses and these causes are based upon different depth or type of investigation, then it is impossible to disentangle the two. It cannot then be stated with confidence that injuries and fatalities have more contributory causes than near misses: such differences may merely be an artefact of the investigation technique rather than an absolute difference in contributory causes between the different outcomes. Hence, where different investigation techniques are used (even when the same taxonomy is applied) the use of absolute frequencies to test the common cause hypothesis is invalid. Any taxonomy is only as good, valid and meaningful as the data (via investigation, interviews with those involved and scientific evidence) to which the taxonomy is applied. In cases where different investigation methods are used to collect data from different severity outcomes, the relative frequencies should be used (this means comparing each level of severity based on the total number of contributory causes assigned to that particular consequence).

In order to use absolute frequencies to test the common cause hypothesis, the same method of data collection and investigation, and the same taxonomy must be used. It is recognised that such a process of ensuring the same level of investigation for major and minor events is unlikely to be a realistic process: it is time consuming to investigate minor incidents with little obvious consequence to the same depth as catastrophes or major accidents. However, such a process would enable researchers to determine if there are more or different causes involved in more serious incidents than in near misses.

7.2 The goals of accident analysis

In Chapter 2 the possible goals and methods of accident analysis were presented. It was concluded that while accident analysis could be used for many purposes (e.g. apportioning blame, fixing liability and determining compensation) that the main purpose of analysis in the context of organisational learning should be to determine the causes of the incident in order to implement preventive measures. Within the auspices of organisational learning there are three purposes behind collecting and analysing near miss incidents. These are given by van der Schaaf (1991) as:

1. Modelling and learning about new failure modes
2. Monitoring the frequency of known failures
3. Maintaining alertness and motivating staff

7.3 The use of taxonomies to analyse incidents

The use of a taxonomy to determine root causes allows the first two objectives to be fulfilled. The third objective follows automatically when a feedback system to and for staff is in place, or when the reported incidents are used for learning purposes.

The use of a taxonomy to determine causes is necessary if comparisons are to be made across multiple incidents and learning is not to take place from individual incidents. However, any taxonomy used should conform to a common set of criteria to ensure that the taxonomy is appropriate and fulfills the purposes for which it has been designed. As yet in both research and industry there is slow progress in providing a definitive set of criteria. Using an analytical approach, this thesis has provided a minimal set of requirements that can be used to judge the appropriateness of any taxonomy. These are currently being researched further. At present they provide clear guidance for researchers and industry as a 'rule of thumb' to evaluate taxonomies currently used or being considered for implementation. The minimal requirements for a taxonomy are provided below.

1. Based on a suitable underlying theory of human behaviour
2. Should include technical components, human behaviour, and organisational and managerial causes.

3. Be reliable i.e. that various independent analysts should reach the same conclusions
4. Should be comprehensive in a qualitative sense (i.e. should be able to handle accidents, damages and near misses; cover both failures and recoveries)
5. Be quantitative, enabling results to be computed and accumulated across many accidents
6. Be consequential. Suggest ways of influencing and rectifying the root causes identified
7. Be revealing i.e. distinguish between events and underlying causes

There are additional aspects which are viewed as beneficial and optional (though not necessary to ensure that the taxonomy will aid in the process of assigning cause and helping to improve the safety of a system) and may be used in determining whether to institute a system (e.g. cost factors, training). The implementation of a method will, of course, depend to some extent upon cost. However, turning this into a minimal requirement may detract from the efficacy and applicability of the analysis process selected, as currently all minimal requirements are viewed as equally important. The validity aspect is as yet an undetermined measure for taxonomies which retrospectively assign causes to incidents and accidents. As such it is not included in this list of minimal requirements. However, as robust methods develop to check the validity of analysis methods then this item could be added. It is viewed here as an attractive addition but not as a requirement. It is suggested that taxonomies should at least meet the minimal requirements (and preferably some of the optional ones) before being used in a 'real life' setting. Optional criteria for taxonomic approaches are therefore outlined below.

1. Validity
2. Auditability
3. Resource efficiency (training)
4. Resource efficiency (usage)
5. Usability
6. Life-cycle applicability
7. Flexibility

These criteria provide guidance for anyone about to implement an analysis system or a way to evaluate a system that is currently operating.

7.3.1 The CIRAS taxonomy evaluated

The above criteria were used to evaluate the CIRAS taxonomy. The taxonomy was compared against the minimal proposed criteria and was found to be satisfactory. The CIRAS taxonomy was then used to analyse and code the data to test the common cause hypothesis.

7.4 Problems for accident investigation and analysis

There are a number of problems that are encountered by researchers and industry alike when investigating and analysing accidents and incidents.

The first problem is one of data comparison. Data collected by industry differs in terms of factors identified, depth of analysis and causal models adhered to. This leads to similar industries being unable to compare their data. It also can lead to different types of incident within one industry being incomparable (for example within the UK railway industry SPADs are analysed and coded in a way that is different to other types of incident such as station overshoot or signallers errors.) In order to ensure that data are comparable (at least within companies if not between them) it would be prudent for all incident types to be analysed using the same method and root cause taxonomy.

Other problems for investigators and analysts relate to the depth of investigation, the detection of incidents and the biases inherent in individuals. These are briefly summarised in the table below (for a full description of these problems see Chapter 2), along with possible solutions to the problems.

Problem	**Causes**	**Possible ways to counteract**
Incomplete investigation	Investigator's view of causation	Training Counterchecking with colleague Reliability trials
	Significance of consequences	Establishing rules relating to analysis (start and stop rules) Analysis performed on potential rather than actual consequences
	Time contraints	Realistic expectations Build in additional time for reporting back Dedicated safety analysts
	Eye witness evidence	Gather as much eye witness evidence as possible Establish no-blame policy or Just Culture approach to enable eye witnesses to report without fear
Incomplete detection/ collection of incidents	Technology not advanced enough or not appropriate	Automatic detection is currently used in some industries for certain types of incident (e.g. SPAD). Automatic detection will never be able to be used in all circumstances, thus reporting systems need to be employed.
	Individuals unwilling to report	Institute no-blame/ confidential/ just culture reporting schemes
Biases in analysis	Confirmation & frequency bias	Training Reliability trials Team coding rather than individual Data checking
	Recognition bias	Range of causal factors in Taxonomy Training Reliability trials
	Author bias	Identify areas of possible conflict of interest
	Political, sponsor and professional bias	Identify areas of possible conflict of interest
	Counter-factual reasoning	Include all causal factors and context conditions in the analysis

Table 35: Problems in accident investigation and some possible solutions

7.5 Issues for the design of incident reporting systems

Chapter 3 provided detailed discussion and analysis of the issues surrounding designing and implementing incident reporting systems. The main learning points are discussed here. Van der Schaaf (1992) outlines 7 modules that are necessary for the design of a successful event reporting and analysis system. The use of these seven modules is dependent upon the goals and purposes of the system. The goals of the system should be decided upon prior to the design of the system. The seven design modules are outlined below:

1. Detection (recognition and reporting)
2. Selection of events for deeper analysis
3. Detailed description and deeper study
4. Classification of root causes
5. Computation to recognise patterns and/ or priorities
6. Interpretation and implementation of recommendations
7. Evaluation

7.5.1 Management support

Other issues that must be taken into consideration when designing and implementing a reporting system are the level of management support and understanding. Are management fully behind the principles and aims of the system? Will managers seek to punish individuals who report their own violations? Is the system to be used for learning only or for discipline as well? Management can be supportive by ensuring anonymity, forgiveness and feedback (Lucas 1991).

7.5.2 Anonymity versus confidentiality

It was discussed in chapter 3 that anonymity at the point of reporting has a number of disadvantages that do not hold for confidential or named reporters. Confidential reports can be made anonymous once the reporter and the incident has passed through the analysis system (this ensures re-contact and follow-up where necessary while still maintaining trust in the system). The majority of third party reporting systems e.g CHIRP, ASRS, MARS and CIRAS all provide confidentiality rather than anonymity. Confidential reporting has the following advantages (Harrison 1991):

1. Improves the completeness of accident and incident reports
2. Helps overcome the barriers associated with near miss reporting
3. Increases the information to build up a human error database
4. Increases suggestions for improvements

In terms of forgiveness, reports are more likely to be made if reporters are not punished for their reports. This requires an approach of either a no-blame policy or a policy of just culture (Marx 2001), not a policy of automatic immunity. Immunity from discipline can be counter-productive to the aims of a confidential reporting system as it distorts the number and quality of reports received (duplicate reports are counted separately). This results in statistical evidence that is based not on the number of near misses actually occurring but on the number of individuals who witnessed or were involved in the same incident. Thus immunity from discipline should not be an integral part of any near miss reporting system. Confidential reporting is one to stimulate a reporting culture – a just culture will hopefully follow.

7.5.3 Feedback

In order to maintain the impetus for making reports, any reporting system requires to have a thorough, sophisticated system of feedback. Reporting schemes may be readily accepted and embraced by staff – reports may be forthcoming in the start up phase. However, if staff are to continue to use the system and keep making reports then they have to see concrete results and receive feedback on what actions their reports have generated. There are a variety of ways in which staff can (and must) receive feedback. Feedback may be performed on an individual basis, where the reporter is provided with an update on what has happened to his/ her report (or even made a temporary member of the investigation team). However, while this may satisfy the individual, it does not ensure that others are informed about the types of report made and what is being done. Therefore feedback via a publication which all potential reporters receive is most appropriate and reaches the widest possible audience. This however, does not preclude individual feedback. Feedback to all potential reporters should also detail and show system changes that have been implemented as a direct result of issues that have been raised through the reporting scheme. Feedback to staff thus ensures that reports are not seen as entering a 'black hole' and disappearing without trace.

7.5.4 Reporting routes

Any reporting system must decide upon the route via which reports can be made. Reports may be made via a form (paper based system), a telephone call, over the internet via either e-mail or an electronic form, or in person. There is no best method – the method selected should suit the staff, the aims and goals of the reporting system and the type of system i.e. confidential, anonymous, open. Table 36 below provides the advantages, disadvantages and points for consideration of various reporting routes.

Reporting route	**Advantages/ Strengths**	**Disadvantages/ Weaknesses**	**Points for consideration**
Paper (form) based	Written description of incident. Provides time to gain understanding before interviewing reporter.	Description may be short. Form may be misunderstood. Fields may not be completed. If name and contact details not provided then no further information can be gained.	Design of form. Fields for inclusion. Space for contact details and name. Availability of forms in the workplace. Reply paid envelopes.
Telephone	Immediate discussion of incident. Can take details and call back if necessary.	May not leave contact details on answering service and so may lose reports.	Freephone line. Enough staff to man telephones. Publicity of phone numbers.
Electronic	Easy to use and immediate receipt.	Potential reporters need to have access to computers.	Publicity of e-mail address. If form based then design of form, and space for contact details and name.
Face to face	Immediate discussion.	Potential reporters may not want to be seen in the safety office. Availability of analysts.	Room for reporting not near managers offices. Enough staff to man the system.

Table 36: Reporting routes and points for consideration

7.5.5 Failures vs recoveries

Any reporting system must decide upon the type of incidents to be reported. A near miss reporting system obviously aims to collect incidents that did not result in an accident in order to learn from them and implement preventive measures. The collection of not only failure modes (i.e. description and analysis of what went wrong) but also of how events were successfully recovered is equally important for a reporting system. Further, the fact that successes (i.e. successful recoveries) are also the focus of reports rather than only failures (what went wrong, what errors were made) helps to lower the threshold for reporting and remove the blame and shame associated with reporting errors.

7.5.6 Incentives

Incentives for encouraging staff to make reports may well result in an increased number of reports coming through a near miss reporting system. However, the increase in reports should be balanced against a number of disadvantages. Firstly, incentives for reporting may lead to staff making a number of similar reports in order to gain rewards, or to staff falsifying reports to gain rewards. Secondly, once the incentive scheme is discontinued reports may fall dramatically. In either case this confounds the incentive system with number and cause of reports, and the evaluation of the effectiveness of measures cannot be disentangled from the incentive system. A such incentives to encourage reporting should be avoided.

7.5.7 Preparation and planning

Before a reporting system (confidential or otherwise) can be launched in a company/ organisation there must adequate planning and preparation in all of the issues discussed in the sections above. Preparation and planning should not be underestimated, and should be undertaken sequentially, but all elements should be planned and prepared before the system is launched.

7.5.8 Evaluation

Once the reporting system has been designed and implemented review and evaluation of the system is necessary to ensure that the system fulfils the goal(s) that were the impetus for initiating the system. However, evaluating the effectiveness of a reporting system is difficult and complicated for the following reasons:

- Multiple interventions usually take place in parallel (ideally to test the effectiveness of a particular corrective measure it would be implemented in isolation and the effect could be measured. As corrective measures also take place at the same time as many other company initiatives and rule changes it is very difficult to assess how successful any one corrective measure is).
- The time lag for changes to occur and take effect is unknown – hence multiple interventions may have taken place and it is difficult to be sure which intervention is responsible for the changes measured.
- Organisations and their culture and rules change over time.

Nevertheless, evaluation is necessary and can be achieved by monitoring the types of incidents and causes that occur over time and whether such causes increase or decrease. Evaluation can also take place at the level of reporting system and whether reports are dealt with in a timely fashion.

The above analytical and methodological results were used to devise a method whereby the common cause hypothesis could be properly tested. This chapter now turns to a discussion of the results of testing the hypothesis.

7.6 Testing the common cause hypothesis - data limitations

The limitations of the data used for this thesis are discussed in Chapter 5, section 5.7. A brief summary will be presented here and the implications for the results discussed.

Firstly, the data were collected by different means. Formal Inquiries were all selected by the company concerned and investigated according to the procedures which are in place. The reporting of incidents that resulted in Formal Inquiries was mainly mandatory. These were collected following completion of the investigation. SPAD reporting is mandatory. Investigations were performed by Driver Team Managers and collected following completion of the investigation. CIRAS reports were made voluntarily by Drivers and investigated by the author.

Secondly, the depth, goal and scope of the incident investigation procedures were different for the three data collection methods. Formal Inquiries were investigated in greater depth than SPAD investigations. Both Formal Inquiries and SPAD investigations were supplemented by evidence from witnesses, while the CIRAS reports were not (due to confidentiality).

Thirdly, the data comprises a much smaller number of serious incidents than of near misses. In an ideal situation, the underlying causal factors would have come from a similar number of incidents for each class of severity.

Finally, the data come from only one domain – the UK railway industry.

The data analysis attempts to take most of these limitations into consideration. The method of testing the common cause hypothesis is based upon proportions of causal factors rather than absolute frequencies. This therefore reduces the problems introduced by different data investigation methods (see section 4.8) and also the problems associated with unequal numbers of incidents in the different severity levels. As the data come from only one domain, any claims as regards the validity, or otherwise, of the common cause hypothesis is not generalisable to all industry types. The testing of the hypothesis is specific to one company in the UK railway industry which is typical of a TOC (train operating company) and interpretation of the results must take this into account.

7.7 Discussion of results of the testing of the common cause hypothesis

The results generated by this empirical test of the common cause hypothesis pertain only to one domain (the railway industry in the UK) and mainly to one group of staff (Drivers). The results are therefore limited to this one group and it remains to be proven if they can be replicated using different domains and a variety of staff groups. The data used in this thesis were investigated in a variety of ways (Formal Inquiries; SPAD investigations and CIRAS reports) and therefore the approach taken was to compare the proportions of causal codes applied to the different incident consequences rather than to compare the absolute frequencies.

7.7.1 Macro results

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

7.7.2 Micro results

Each of the macro codes comprises a number of individual causal codes. The Technical macro code contains two individual codes: namely Technical steady and Technical change. The Proximal macro code has eight constituent individual codes: attention, perception, fatigue, knowledge based, skill based, rule based, rule violation, communication between staff. The Intermediate macro code has four individual codes: maintenance, communication from (middle) managers to staff, communication from staff to (middle) managers and training. The final set of codes is grouped under the Distal macro code: these are design, procedures, objectives, staffing, top manager to staff communication, staff to top manager communication and culture. The results of each of the individual micro codes will be discussed in turn.

7.7.2.1 Micro results – Technical causal factors

At the level of micro Technical causal codes (Technical Steady and Technical Change), there were no significant differences in the proportions for the various consequence levels. Thus for the technical causal factors, near misses and more serious incidents do not have significantly different proportions.

7.7.2.2 Micro results – Proximal causal factors

At the Proximal micro level, one of the eight codes was significantly different in terms of the proportions of causal codes. The significant difference was for Knowledge based errors. Further analysis revealed that significant differences in proportions of Knowledge based errors occurred for two of the three paired comparisons. There were no significant differences between 'fatality & injury' and 'damage'. However there were significant differences in the proportions of Knowledge based errors for 'fatality & injury' when compared with 'near miss', and also for 'damage' when compared to 'near miss'. Thus for knowledge based errors near miss incidents are not indicative of more serious consequences.

In order to provide further insight into these differences, the proportions shall now be discussed. For 'fatality & injury' incidents the proportion of codes assigned to Knowledge based errors was 0.102. For 'damage' incidents the proportion was 0.08. For 'near miss' incidents the proportion was 0.023. Thus, 'fatality & injury' and 'damage' incidents had a greater proportion of knowledge based causal factors than 'near miss' incidents. There are a number of possible explanations for this finding. Firstly, knowledge based errors may occur more frequently in incidents that have a more serious consequence. Secondly, the methods of investigation may have a bearing on the more frequent assignment of knowledge based errors. It may be that more in-depth investigation has resulted in the recognition of knowledge based errors. Embrey and Lucas (1988) 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. Reason (1990) discusses the detection of skill, rule and knowledge based errors and concludes that knowledge based errors are only detected over a longer time scale than either skill or rule based errors, and that the success of knowledge based decisions cannot be clearly recognised in advance. Further studies by Rizzo et al (1986) and Bagnara et al (1987) found that knowledge based errors were detected mainly as the result of standard check behaviour. Both studies found that between 75 and 85% of errors were detected.

Further research in both the railway industry and other industries is necessary to determine if this finding can be replicated and is generalisable across industry.

7.7.2.3 Micro results – Intermediate causal factors

At the Intermediate level of coding, only one of the four Intermediate micro causal factors, Training, was significantly different in terms of the proportion of training codes assigned across the three severity levels. Further analysis revealed that only 'fatality & injury' incidents were significantly different from 'near miss' incidents.

A more detailed insight into the proportions reveals that training causal codes are assigned more frequently for 'fatality & injury' incidents than for 'near miss' incidents. The actual proportions are 0.077 for 'fatality & injury' and 0.0069 for 'near miss' incidents.

There are a number of possible explanations for this finding. Firstly, it may be the case that training causal factors are more prevalent in incidents with a more serious outcome. However, as training was high on the company agenda, it is more likely that training issues have been more frequently recognised by managers during the Formal Investigation and SPAD investigations than have been revealed by staff during interview after submitting CIRAS reports. Issues such as training 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.

7.7.2.4 Micro results – Distal causal factors

There are seven possible causal codes at the micro Distal level. Of these one (communication from staff to top managers) was not a causal factor in any of the incidents. Thus of the six causal factors used at the Distal micro level only Procedures was significantly different when the proportions were compared for the three severity outcomes. On further examination it was revealed that the proportions of procedure codes assigns was significantly different for 'fatality & injury' incidents and 'near miss' incidents: there were no significant differences between 'fatality & injury' and 'damage' incidents or between 'damage' and 'near miss' incidents. The proportion of procedures assigned to 'fatality & injury' was 0.166, for 'damage' the proportion was 0.12 and for 'near miss' the proportion was 0.069. Thus it can be seen that a greater proportion of such causal codes was assigned to 'fatality & injury' than to 'near miss' incidents.

There are a number of possible explanations for this finding. It may be that more procedural causal factors are underlying causes in 'fatality & injury' incidents than in 'near miss' incidents. It may also be possible that procedures are standard items that managers select as causes when incidents happen – a failure in procedures is fairly easy to remedy (new procedures are compiled). It is highly likely that causes that have traditionally dominated management thinking such as procedures and training are more prevalent when management carry out the investigation.

7.8 Conclusions

The thesis has outlined the problems with testing the common cause hypothesis and developed a method for properly testing it. The analytical findings in relation to collecting and analysing data were discussed. The analysis of the common cause shows that at the general level of macro codes, there is support for the common cause hypothesis as there are no significant differences between severity levels. At the more specific level of individual codes only three of 21 causal factors are significantly different, namely knowledge based errors, training and procedures. In all cases three cases 'fatality & injury' and 'near miss' incidents have different proportions assigned. Also in all three cases the proportions are greater for the more serious incidents. These findings provide qualified support for the common cause hypothesis within the railway domain .

Chapter 8 Key conclusions and recommendations

8 Introduction

This chapter provides an overview of the key conclusions from the thesis including the analytical and methodological insights as well as the results from the testing of the common cause hypothesis.

8.1 Goals of accident and incident analysis

The goals of accident and incident analysis should be to determine the causes in order to implement effective remedial measures thus mitigating the possibility of a similar incident recurring. This goal is incompatible with apportioning blame or determining liability. Investigations that are performed to determine liability are often focussed on one single incident, one cause or individual and so fail to capture or use pertinent and useful information.

8.2 Accident and incident analysis

The goal of organisational learning as described above, can only be achieved by the analysis of multiple events and not from single events. Thus, a taxonomy should be used to perform the analysis as this provides the opportunity to systematically analyse and compare incidents. Any taxonomy that is used should conform to a common foundation and thus enable comparisons between incident types and eventually across companies. A set of minimal requirements for any taxonomy are presented and discussed in Chapters 2 and 7.

Data collected by industry differs in terms of factors identified, depth of analysis and causal models adhered to. This leads to similar industries being unable to compare their data. It also can lead to different types of incident within one industry being incomparable (for example within the UK railway industry SPADs are analysed and coded in a way that is different to other types of incident such as *station overshoot* or signallers errors.) In order to ensure that data are comparable (at least within companies if not between them) it would be prudent for all incident types to be investigated and analysed using the same method and root cause taxonomy.

8.3 Investigation problems

The problems of accident investigation may lead to incomplete information or reports containing less information than is useful. This is often the case in the railway industry especially with regard to SPAD investigations. The problems in relation to accident investigation are discussed in Chapters 2 and 7. However, training, counter checking with colleagues, discussion of missing information, designing a data collection checklist, performing reliability trials, establishing rules relating to analysis (start and stop rules), and establishing a reporting culture may help to counteract the problems associated with investigating incidents.

8.4 Issues for the design of incident reporting systems

Issues for designing and implementing incident reporting systems were discussed in Chapter 3 and the example of designing and implementing CIRAS was provided in Chapter 4. The CIRAS methodology was then used to test the common cause hypothesis.

In this section, the main points to be considered when designing a reporting system are summarised. When designing and implementing a reporting system, decisions should be taken as to whether the system to be implemented should be confidential, anonymous or open. The type of system implemented may depend to some extent on the culture of the organisation. Anonymous reporting has many disadvantages (inability to re-interview reporter or follow-up report subsequently) and so should only be instituted as a last resort, where neither confidential nor open reporting would work. Confidential reporting provides all the reassurances of anonymous reporting but enables follow-up investigation and interviewing. Open reporting may not be successful in a company where management are distrusted and a culture of discipline for minor mistakes is prevalent. Confidential reporting is thus preferred. Further, decisions need to be made on whether reporters to the system are guaranteed immunity from discipline, whether the policy is no-blame for reports via the system, or whether a policy of just culture will be implemented. Immunity from discipline should be avoided as this may result in data that does not reflect the actual situation (as witnesses, bystanders and those involved may all report the same incident in order to take advantage of the immunity clause). Either a no-blame policy (like that of CIRAS which enables discipline to take place should the incident be reported or detected via normal channels) or a policy of just culture should be instituted. This however means that the policy in place should be clear and defined and publicised to staff.

Incentives designed to increase reporting rates may well succeed, however they are to be avoided. This is because incentives may disguise the actual rate of occurrence of incidents and when the incentives cease reports may also cease.

Van der Schaaf's (1992) modules for designing a reporting system are shown in the table below along with examples of the decisions necessary at each module. All modules should be prepared and planned prior to the launching of the system.

Van der Schaaf's 7 modules	**Decisions necessary at each module**
Detection (recognition and reporting)	1. What type of reports are to be collected: e.g. near misses, unsafe practices, deviations, recovered events. 2. Publicity and training to ensure staff are made aware of the type of events to report. 3. Designing of the reporting route.
Selection of events for deeper analysis	1. As it is unlikely that all reported incidents can be analysed to the same level, the system needs to devise rules that clearly identify the events that should be analysed deeper, and those which have a cursory analysis. 2. Based on potential rather than actual consequences.
Detailed description and deeper study	1. Select a method which enables detailed description and deeper analysis)e.g. Fault Tree analysis) which leads to a detailed description of the event. 2. This relies on detailed investigation and interviewing and therefore adequate training for analysts.
Classification of root causes	1. Select taxonomy for the classification of root causes. The taxonomy should satisfy the minimal requirements for a

	taxonomy, and possibly some of the additional ones. 2. Analysts require training in the methodology 3. Inter-rater reliability performed regularly. 4. Results in a set of classifications and not simply the 'main cause'.
Computation to recognise patterns and/ or priorities	1. Design data base to enable statistical analysis. Can database handle text and numbers or numbers alone? 2. Perform analysis to determine reliable patterns which enable structural factors to be identified.
Interpretation and implementation of recommendations	1. Statistical results used to generate recommendations that can be implemented. 2. Design group that meets to discuss the feasibility of implementing changes. 3. Feedback channel to provide route for organisational learning (e.g. publications, input from system to 'toolbox talks' or briefing and training sessions.
Evaluation	1. Decide on ways in which the above recommendations to improve the system will be evaluated, and time frame (bearing in mind that it may take some time to build up a sufficiently large database).

Table 37: The near miss modules and the decisions at each stage

8.5 Testing of the hypothesis

All of the issues above were addressed in order to develop a methodology to test the common cause hypothesis. The statistical method of testing the common cause hypothesis is based upon proportions of causal factors rather than absolute frequencies. This therefore reduces the problems introduced by different data investigation methods (see section 4.8) and also the problems associated with unequal numbers of incidents in the different severity levels. As the data come from only one domain, any claims as regards the validity, or otherwise, of the common cause hypothesis is not generalisable to all industry types. The testing of the hypothesis is specific to one company in the railway industry and interpretation of the results must take this into account.

8.5.1 Results

The results from the testing of the hypothesis were fully discussed in Chapter 7 and are therefore summarised here.

At the macro (or general) level of testing the hypothesis there was no significant differences between the consequence levels at the Technical, Proximal, Intermediate or Distal causal codes.

The micro level (individual) codes are shown in the table below along with those which were significantly different. As Table 38 below shows, only three of 21 codes were significantly different. All the differences were in the same direction, with 'fatality and injury' having a greater proportion of codes that the 'near miss' incidents.

Causal codes at the micro level	Significant difference found: Yes/ No
Technical steady	No
Technical change	No
Attention	No
Perception	No
Knowledge	Yes: between 'fatality & injury' and 'near miss' and between 'damage' and 'near miss'.
Rule violation	No
Rule based error	No
Skill based error	No
Communication between staff	No
Fatigue	No
Communication from staff to middle management	No
Communication from middle management to staff	No
Maintenance	No
Training	Yes: between 'fatality & injury' and 'near miss'.
Top down communication	No
Communication from staff to senior management	No
Procedures	Yes: between 'fatality & injury' and 'near miss'.
Design of equipment	No
Rostering and staffing	No
Objectives	No
Culture	No

Table 38: Causal codes and significance

The analysis of the common cause hypothesis shows that at the general level of macro codes, there is support for the common cause hypothesis as there are no significant differences between severity levels. At the more specific level of individual codes only three of 21 causal factors are significantly different, namely knowledge based errors, training and procedures. In all cases three cases 'fatality & injury' and 'near miss' incidents have different proportions assigned. Also in all three cases the proportions are greater for the more serious incidents. These findings provide qualified support for the common cause hypothesis within the railway domain .

8.6 Recommendations for further research

The following recommendations for further research have all been generated from the research undertaken to produce this thesis. They relate not only to the common cause hypothesis but also to the analytical and methodological issues addressed throughout the thesis.

The first set of recommendations relate directly to the UK railway industry.

1. In order to determine if traditional causal factors such as procedures and training are identified and selected more frequently by management a research study should be performed to compare the investigation techniques and findings of managers with human factors experts. In order to use the same set of incidents for a direct comparison, the incidents could be scenario based.

2. To capture a greater depth of information and therefore *have more complete reports* on SPAD events it is recommended that training in investigation techniques be performed with Driver Team Leaders/ Managers. Such training should also include material on the biases that can occur during incident investigation.

The following set of recommendations are general to industry and academia.

1. As this study was limited to one domain, it is recommended that a further empirical test of the common cause hypothesis be performed for a number of different domains (each comparing near misses and more serious incidents using the same taxonomy). This would provide more robust evidence of the applicability of the theory.

2. In order to fully test the theory it is recommended that any further empirical test of the hypothesis should also be based upon more than one taxonomy (i.e. a comparison of different taxonomies would help to establish the validity of the hypothesis and the utility of the taxonomies).

3. Further research should be undertaken to establish if serious accident consequences do have more contributory causes than near misses, or if this proposition is merely a function of the data collection and investigation process. This would require an adequate number of serious incidents, property damage incidents and near misses to be collected and investigated in the same manner and to the same depth of investigation.

4. Research should be undertaken to establish the minimum criteria for a taxonomy. A set of *minimum* criteria based on various suggested criteria from the literature is described in Chapter 2. Once established such criteria could be used to evaluate the various taxonomies that are currently available – thus enabling users to establish which taxonomy best meets their needs.

5. Near miss reporting systems should be extended to include not only the causes of failures but also the causes of successful recoveries. This additional data collection enhances system safety by enabling not only the prevention of failure factors but also the enhancement of recovery factors. Such a focus for reporting (i.e. recovery factors) may also help lower the threshold for reporting as it concentrates on positive aspects of incidents rather than negative aspects. When recovery factors are collected and analysed then the taxonomy used for analysis should also include recovery factors.

6. The combination of retrospective safety management methods (such as incident reporting) with prospective (prediction-oriented) techniques (e.g. FMEA) in the same domain would lead to an integrated system that is better equipped to fully understand, predict and therefore prevent the occurrence of accidents.

References

Adams, E. (1976) Accident causation and the management system. *Professional Safety, October, 1976,* pp. 26 – 29.

Adams, N.L. and Hartwell, N. M.(1977) Accident reporting systems: a basic problem area in industrial society. *Journal of Occupational Psychology,* 50, 285-298

Andersson R & Menckel E.(1995) On the prevention of accidents and injuries. A comparative analysis of conceptual frameworks. *Accident Analysis and Prevention,* Vol 27, No. 6, 757 – 768, 1995.

Bagnara, S., Stablum, F., Rizzo, A., Fontana, A., & Ruo, M. (1987) Error detection and correction: A study on human-computer interaction in a hot strip mill production planning and control system. In *Preprints of the First European Meeting on Congnitive Science Approaches to Process Control,* Marcoussis, France, October, 1987.

Baram, M. (1997) Shame, blame and liability: why safety management suffers organisational learning disablities. *In* Hale, A., Wilpert, B., and Freitag, M. (eds) *After the event: from accident to organisational learning* Pergamon, Oxford.

Benner, L. (1985) Rating accident models and investigation methodologies. *Journal of Safety Research* 16, 105 – 126.

Benner, L. (2000) Accident Investigations – A case for a new perspective. Available online at http://home.cox.rr.com/lbjr/papersa/SAE80.html

Berman, J. (1996) Confidential event reporting (HF/GNSR/5009). Final report. Document reference: 112/C/03/il 1996 – A report submitted under the 1995/1996 IMC GNSR Programme. *Commercial-in-confidence.*

Billings, C. E., & Reynard, W. D. (1984) Human Factors in Aircraft Incidents: Results of a 7 – year study. *Aviation, Space and Environmental Medicine,* October 1984, 960 - 965

Bird, F. E. (1966) *Damage control.* Insurance Company of North America. Philadelphia

Bird, F. E. (1976) *Loss Control Management,* USA, Institute Press

Blalock, H. M., Jr (1972) *Social Statistics.* (2nd edition) New York: McGraw-Hill

Brown, G. R. (1991) Use of traffic conflicts for near miss reporting. *In* Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (eds) *Near miss reporting as a safety tool.* Oxford: Butterworth-Heinemann Ltd

Caro, T. M., Roger, R., Young, M., Dank, G. R. (1979) Inter-observer reliability. *Beahviour* 69:303-315.

Clarke, S. (1998) Safety culture on the UK railway network. *Work and Stress* Vol 12 No.1, 6 – 16

Davies, J. B., Wright, L., Courtney, E., & Reid, H. Confidential incident reporting on the UK railways: The CIRAS system. *Cognition, Technology and Work* Vol:2 Number:3 2000, Springer-Verlag London Limited

Department of Transport (1987) *Investigation into the Kings Cross Underground Fire.* Desmond Fennell QC, HMSO, London

Department of Transport (1989) *Investigation into the Clapham Junction Accident.* Sir Anthony Hidden QC, HMSO, London.

DeSalvo, F. J., & Zurcher, L. A. (1984) Defensive and supportive parental communication in a discipline situation. *Journal of Psychology*, 117: 7 – 17

Edmondson, A. C. (1996) Learning from mistakes is easier said than done: Group and organizational influences on the detection and correction of human error. *Journal of Applied Behavioral Science*, 32: 5 - 28

Edwards, A. L. (1972) *Experimental design in psychological research.* (4th edition) New York: Holt, Rinehart & Winston.

Embrey, D. E., & Lucas, D. A. (1988) The Nature of Recovery from Error. *In* L. H. J. Goossens (ed.) *Human recovery: Proceedings of the COAST A1 Working Group on Risk Analysis and Human Error*, Delft University of Technology, 13 October 1987

Fahlbruch, B., and Wilpert, B. (1997) Event analysis as problem solving process. *In* Hale, A., Wilpert, B., and Freitag, M. (eds) *After the event: from accident to organisational learning* Pergamon, Oxford.

Ferry, T. S. (1988) *Modern accident investigation and analysis.* 4th edition, J. Wiley & Sons, New York.

Feyer, A. M. & Williamson, A. M. (1991) A classification system for causes of occupational accidents for use in preventive strategies. *Scand. J. Work Environ. Health,* 17: 302-311

Fischoff, B., Slovic, P. and Lichtenstein, S. (1978). Fault trees: sensitivity of estimated failure probabilities to problem representation. *Journal of experimental Psychology: Human Perception and Performance,* 4,330 – 344.

Flanagan, J. C. (1954) The critical incident technique. *Psychological Bulletin,* Vol 51, No 4, 327 - 358

Fleiss, J. L. (1981) *Statistical methods for rates and proportions.* New York, John Wiley & Sons

Fletcher J A & Douglas H M (1970) *Total environmental control.* National Profile Ltd. Toronto.

Frese, M., & van Dyck, C. (1996) Error management: Learning from errors and organizational design. *Paper presented at the annual meeting of the Academy of Management,* Cincinatti

Gibb, J. R. (1961) Defensive communication. *Journal of communication,* 11: 141 – 148

Groeneweg, J. (1988) *Controlling the controllable: the management of safety.* DSWO Press, Leiden, 4th Edition

Hale, A. R. & Hale, M. (1972) *A review of the industrial accident research literature.* Research paper to the Committee on Safety and Health at work. HMSO London.

Hale, A. (1997) The goals of event analysis. *In* Hale, A., Wilpert, B., and Freitag, M. (eds) *After the event: from accident to organisational learning* Pergamon, Oxford.

Hale, A.R. (2000) Conditions of occurrence of major and minor accidents. *2me Séance du seminaire "Le rsique de defaillance et son controle par les individus et les organisations.* 6 - 7 novembre. Gif sur Yvette

Harrison, P. I. (1991) Harnessing operational experiences and learning lessons: the value of confidential incident reporting schemes. In: *Proceedings of the 11th Annual Safety and Reliability Society Symposium, September 18 –19, Sutton Coldfield, England, pp. 42 – 56.*

Heinrich, H. W. (1931) *Industrial accident prevention.* McGraw-Hill, New York.

Heinrich, H. W. (1941) *Industrial accident prevention.* (Second Edition) McGraw-Hill, New York.

Heinrich, H. W. (1950) *Industrial accident prevention.* (Third edition) McGraw-Hill, New York.

Heinrich, H. W. (1959) *Industrial accident prevention.* (Fourth edition) McGraw-Hill, New York.

Heinrich, H. W., Petersen, D. & Roos, N. (1980) *Industrial Accident Prevention.* (Fifth edition). McGraw-Hill, New York

Heybroek, R. (1995) *Improving safety training: Human factors discrepancies report.* Vosper Thornycroft (UK) Limited, MSC Division. Commercial-in-confidence.

Van der Horst, R. (1991) Video analysis of road user behaviour at intersections. *In* Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (eds) *Near miss reporting as a safety tool.* Oxford: Butterworth-Heinemann Ltd

Hyden, C. (1987) The development of a method for traffic safety evaluation. *The Swedish Traffic Conflicts Technique.* Bulletin 70. University of Lund.

Hyos, C. G. & Zimolong, B. (1988). *Occupational safety and accident prevention.* Elsevier, Amsterdam

Johnson, C. (2002) Reasons for the failure of incident reporting in the healthcare and rail industries. In F. Redmill and T. Anderson (eds) *Proceedings of the 10th Safety-Critical Systems Symposium;* Springer-Verlag, Berlin, Germany.

Johnson, W. G. (1975) MORT – the management oversight and risk tree. *Journal of Safety Research,* 7, pp. 4 – 15.

Johnson-Laird, P. N. & Wason. P.C. (1970) Insight into a logical relation. *Quarterly Journal of Experimental Psychology,* 22, pp 49 – 61

Kaplan, H. S., Battles, J. B., & Mercer, Q. S. (1998) The Hazard Analysis Action Decision Table, a Means to Limit Unnecessary Change in a Medical Event reporting

System. In proceedings of *Enhancing Patient Safety and Reducing Errors in Healthcare,* Rancho Mirage, CA., November 8 – 10, 1008.

Kirwan, B. (1992) Human error identification in human reliability assessment. Par t 2: detailed comparison of techniques. *Applied Ergonomics*, 23, (6) 371 – 378

Kjellen, U. (1984) The deviation concept in occupational accident control – I, definition and classification. *Accident Analysis and Prevention,* 16, pp. 289 – 306.

Ladkin, P. B. (2000) Causal reasoning about accidents. In F. Koorneef and M. van der Meulen, (editors) *SAFECOMP 2000, Lecture notes in computing science* No. 1943, pages 344 355. Springer-Verlag, Berlin, Germany, 2000.

Laflamme L (1990) A better understanding of occupational accident genesis to improve safety in the workplace. *J. Occup. Accid.* 12: 155 – 165

Lekberg, A. (1997) Different approaches to accident investigation: how the analyst makes the difference. In *Proceedings of the 15th International Systems Safety Conference.* Sterling, VA: International Systems Safety Society.

Lortie, M. & Rizzo, P. (1999) The classification of accident data. *Safety Science* 31 (1999), 31 - 57

Lozada-Larson, S. R. & Laughery, K. R. (1987) Do identical circumstances precede major and minor injuries? In *Rising to new heights. Proceedings of the 31st annual meeting of the Human Factors Society.* New York. 1. 200 - 204

Lucas, D. A. (1991) Organisational aspects of near miss reporting. *In* Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (eds) *Near miss reporting as a safety tool.* Oxford: Butterworth-Heinemann Ltd

Lundberg, A. (1992) Dialysmalet – ett oppet sar I svensk rattskipning. Unpublished paper.

Makin, P. & Sutherland, V. (1991) A fatal inversion? *Occupational Safety and Health,* Nov. 1991, 40 - 42

Marx, D (2001) Patient Safety and the 'just culture': A primer for health case executives. Columbia University, New York

Meister, D. (1971) *Human factors – theory and practice.* John Wiley, New York.

National Patient Safety Agency: *Doing Less Harm* (2001) Available online at http://www.npsa.org.uk/admin/publications/docs/draft.pdf

NCCAN (1998) Study for National Center for Child Abuse and Neglect (Grant 90-CA-1550). Available online at http://www.nccd-crc.org/crc/nccan.pdf

O'Leary, M. (2001) The British Airways human factors reporting programme. *Reliability Engineering and System Safety* 75 (2002) 245 – 255

Petersen, D., C. (1971) *Techniques of safety management.* New York. McGraw Hill

Petersen, D., C. (1978) *Techniques of safety management.* (2nd edition) New York. McGraw Hill

Petersen, D. C. (1989) *Techniques of safety management. A systems approach.* (3rd edition). Aloray. Goshen. New York.

Railway Safety 1996, Department of Transport (London: HMSO)

Rasmussen, J. (1997) Risk management in a dynamic society: a modelling problem. *Safety Science* 27 (2/3) 183 – 213.

Rasmussen, J. (1986) *Information processing and human-machine interaction.* North-Holland, Amsterdam

Reason, J. (2001) *Human Error Reduction Operation.* (for Railway Safety Limited) The Vision Consultancy, England

Reason, J. (1997) *Managing the risk of organisational accidents.* Ashgate Publishing Limited, Hampshire, England.

Reason, J. (1990) The contribution of latent failures to the breakdown of complex systems. *Philosophical Trans R Society of London.* B. 327, 475 – 484

Reason, J. (1990) *Human error.* Cambridge University Press, Cambridge.

Rizzo, A., Bagnara, S., & Visciola, M. (1986) Human error detection processes. *In* G. Mancini, D. Woods & E. Hollnagel (Eds.) *Cognitive Engineering in Dynamic Worlds.* Ispra, Italy: CEC Joint Researcg Centre, 1986.

Robinson, J. & Shor, G. (1989) Business-cycle influences on work related disability in construction and manufacturing. *Milbank Quarterly* 67, 92-113

Rosenthal, I. (1997) Major event analysis in the United States chemical industry: Organisational learning vs. liability. *In* Hale, A., Wilpert, B., and Freitag, M. (eds) *After the event: from accident to organisational learning* Pergamon, Oxford.

Saari, J. T. & Lahtela, J. (1979) Characteristics of jobs in high and low accident frequency companies in the light metal working industry. *Accident Analysis and Prevention* 11: 51-60

Salminen, S., Saari, J., Saarela, K. L. & Rasanen, T. (1992) Fatal and non-fatal occupational accidents: identical versus differential causation. *Safety Science* 15 (2) 109 – 118

Salomiemi, A. & Oksanen, H. (1998) Accidents and fatal accidents – some paradoxes. *Safety Science* 29 (1998) 59 – 66

Van der Schaaf, T., W. (1988) Critical incidents and human recovery: some examples of research techniques. *In* L. H. J. Goossens (ed.) *Human recovery: Proceedings of the COAST A1 Working Group on Risk Analysis and Human Error*, Delft University of Technology, 13 October 1987

Van der Schaaf, T., W. (1991) A framework for designing near miss management systems. *In* Van der Schaaf, T. W., Lucas, D. A. and Hale, A. R. (eds) *Near miss reporting as a safety tool.* Oxford: Butterworth-Heinemann Ltd

Van der Schaaf, T. W. (1992) *Near miss reporting in the Chemical Process Industry.* Eindhoven: Technische Universiteit Eindhoven. PhD Thesis

Van der Schaaf, T. W. & Kanse, L. (2000) Errors and error recovery. *In* Elzer, P. F. Kluwe, R. H. & Boussoffara, B. (Eds.), *Human error and system design and management.* (Lecture notes in control and information sciences, 253, pp. 27 – 38). London: Springer-Verlag

Van der Schaaf, T. W. & Kanse, L. (2002) Not reporting successful recoveries from self-made errors? An empirical study in the chemical process industry. *In CW Johnson (editor) Proceedings of the Workshop on Investigation and Reporting of Incidents and Accidents,* GIST Technical Report G2002-2, pp. 180 – 183 University of Glasgow, Scotland.

Shanning, H. S. & Manning, D. P. (1980) Differences between lost-time and non-lost-time industrial accidents. *Journal of Occupational Accidents*, 2 (1980), 265 - 272

Shappell, S. A., and Wiegmann, D., A. (1997) A human error approach to accident investigation: the taxonomy of unsafe operations. *The International Journal of Aviation Psychology*, 7 (4), 269 - 291

Shappell, S. A., & Wiegmann, D. A. (2001) Unravelling the mystery of general aviation controlled flight into terrain accidents using HFACS. *Presented at the 11th International Symposium on aviation psychology.* Columbus, OH: The Ohio State University, 2001.

Sheahy, J. E. (1979) Impact of accident causation on intervention strategies. *Proceedings of the Human Factors Society – 23rd Annual Meeting,* The Human Factors Society, Inc.

Sheen, Mr Justice. (1987) *M.V. Herald of Free Enterprise*. Report of Court No. 8074, (Department of Transport, London: HMSO, 1987)

Shorrock, S. (2002) Error classification for safety management – Finding the right approach. *Presented at Investigation and Reporting of Incidents and Accidents*, University of Glasgow, Glasgow, 17th – 20th July 2002.

Skiba, R. (1985) *Taschenbuch Arbeitssicherheit*, Erich Schmid Verlag, Bielefeld

Sutherland, V. J., Makin, P. J., & Cox, C. J. (2000*) The management of safety. The behavioural approach to changing organizations.* Sage Publications, London

Svenson, O., Lekberg, A., and Johansson, A. E. L. (1999) On perspective, expertise and differences in accident analyses: arguments for a multidisciplinary integrated approach. *Ergonomics*, 1999, Vol. 42, No. 11, 1561 – 1571

Swain, A. D. (1974). *The human element in systems safety: a guide for modern management.* InComtec Ltd, Camberley

Swuste, P., Hale, A. R., & Guldenmund, F. (1999) Change in a steelworks: learning from failures and partial successes. *Preliminary paper for the NeTWork workshop on "Achieving successful safety interventions"* Bad Homburg, June, Safety Science Group, Delft University of Technology

Tarrants, W. E. (1980) *The Measurement of Safety Performance.* London: Garland.

Tinline, G. & Wright, M. S. (1993) 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. Four Elements

Tversky, A. & Kahneman, D. (1982) Judgements of and by representativeness. *In* D. Kahneman, P. Slovic & A. Tversky (eds) *Judgements under uncertainty: Heuristics and Biases*, Cambridge University Press

Tye & Pearson (1974/5) Analysis of almost 100,000 accidents in British Industry. In HSE (Health and Safety Executive) (1992) *Successful Health and Safety Management.* HS(G)65. London:HMSO

Wagenaar W. A., Hudson, P. T. W., & Reason, J. T. (1990) Cognitive failures and the cause of accidents. *Applied Cognitive Psychology, 4, 231 –252*

Wagenaar, W. A., and van der Schrier, J. (1997) Accident analysis the goal, and how to get there. *Safety Science* Vol. 26 No.1/2, pp. 25 – 33

Wallace, B., Ross, A., Davies, J. B., Wright, L. & White, M. (2002) The creation of a new minor event coding system. *Cognition, Technology & Work*, 4:1-8

Webb, G. R., Redmand, S., Wilkinson, C. and Sanson-Fisher, R.W. (1989) Filtering effects in reporting work injuries. *Accident Analysis and Prevention*, 21, 115-123.

Williamson, A. M., Feyer, A., & Cairns, D. R. (1996) Industry differences in accident causation. *Safety Science* Vol 24, No. 1, pp 1 – 12

Wright, L. (2000) Towards an empirical test of the iceberg model. *7th European Conference on Cognitive Science Approaches to Process Control*, 2000. Conference proceedings.

Wright, L. B., Davies, J. B., Courtney, E., & Reid, H. (2000) CIRAS: Collecting and analysing human factors data from the UK rail industry. *Proceedings of ESREL 2000, SARS and SRA-Europe Annual Conference, Foresight and Precaution.*

Appendix 1

The CIRAS follow-up interview form

INTERVIEW FOLLOW-UP FORM

Date : ______________ **Region:** _____ **Initials:** _____ **Report number:** _________

1. Brief description of report:

Notes/Sketch

CONDITIONS (relating to incident or issue where appropriate)

2. Location where incident took place or issue occurs (place name, location etc.):

____

2b. Nature of site:

Train Depot	Maintenance Depot
Signalbox	On route
On board train	Platform
Level crossing	Loop
Lineside	On the line
Tunnel	Freight yard

Other ______________________

3. Does the report concern work activity associated with Railtrack's West Coast Route Modernisation Programme? Yes No

What WCRM Project does the work concern?______________________

4. Rail conditions:
Dry
Wet
Greasy
Contaminated by previous train
Leaf affected
Flooding

5. Lighting Conditions:
Day
Night
Dawn
Dusk
Tunnel

6. Weather Conditions:
Sunshine
Fine
Rain
Snow
Fog/ Mist
Wind
Other ______________

7. Underfoot Conditions:
Dry
Wet
Stable
Uneven
Slippery
Other__________

8. Day of the week:
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday

9. Time of day:
Exact time: ________
Midnight – 4am
4am – 8am
8am – 12pm
12pm – 4pm
4pm – 8pm
8pm – Midnight

INCIDENTS ONLY

10. When were you last on duty before the incident? (in days) ____________________

11. Did you have a 12 hour rest period? YES/NO

12. Time into shift when incident occurred? (in minutes) ____________________

13. Are you aware of similar incidents? Yes No
If Yes: Self Colleague

Description: ____________________

INCIDENT OR ISSUE

14. Is the situation ongoing? Yes No

15. What could have happened?

Potential fatality
Potential SPAD
Potential overshoot
Potential discipline
Potential injury to staff
Potential injury to passengers
Potential under reporting
Other
Please state: ____________________

Potential delay to service
Potential delay to work
Potential delay in emergency
Potential derailment
Potential collision
Potential damage to infrastructure
Potential damage to train

16. In your opinion why did the consequences just mentioned not happen?(incident only)
In your opinion what prevents the possible consequences from happening? (issue only)

____

____

17. What action did you take at the time of the incident? (incident only)
What action do you normally take when this situation occurs? (issue only)

____

____

18. Who do you think is responsible for this situation?

____

19. Why have you made your report to CIRAS?

____

SPECIFIC INFORMATION

QUESTIONS 20 TO 23 FOR TRAIN STAFF ONLY

20. What type of train was being driven? (e.g. 303, 158 not the head code) __________

21. How busy was the train?

Empty/ Nearly Empty Fairly Quiet Fairly Busy Crowded Don't know

22. Driver only operations? Yes No

23. Was the driver accompanied in the cab? Yes No

If Yes: Authorised personnel (please specify)____________________
Unauthorised personnel (please specify)____________________

QUESTIONS 24 TO 28 FOR SPAD / NEAR SPAD ONLY

24. AWS on unit :	**25.** AWS at signal:	**26.** DRA on unit:
Fitted and working	Fitted and working	Fitted and working
Fitted, not working	Fitted, not working	Fitted, not working
Not fitted	Not fitted	Not fitted

27. Type of signal passed? ______________________________

28. Type of signalling system? ______________________________

QUESTIONS 29 TO 38 FOR INFRASTRUCTURE STAFF ONLY

29. Equipment / Machinery being used (i.e. tools and on-track machines - specify types/names):______________________________

30. Was the work in a: **T(ii)** Type: _____ **T(iii)** Type: _____

31. Was the system of work: **Red Zone** **Green Zone**

CERTIFICATES

32. Do you hold a PTS certificate?	Yes	No
33. Are you a certified Lookout?	Yes	No
34. Are you a certified IWA?	Yes	No
35. Are you a certified COSS?	Yes	No
36. Are you a certified PC?	Yes	No
37. Are you a certified PICOP?	Yes	No
38. Any other certificates? (Please specify)		

GENERIC INFORMATION

39. Staff category. Please state job title and/or type of work:

____

40a. Time worked at this grade/ passed as driver: ______________________

b. Length of service in the railway: ______________________

41. Age: __________

42. Company currently employed by: ______________________

43. Company sub-contracted to (if applicable): ______________________

44. Do you have a regular "booking on point"? Yes No

State where: ______________________________

CONFIDENTIALITY

45. Did you report this through normal channels?	Yes	No
If Yes was this reported:	verbally	in writing
46. Have you told anyone you have put this report to CIRAS?	Yes	No
47. Would you be easily identified from this report?	Yes	No

If yes, state how: ______________________________

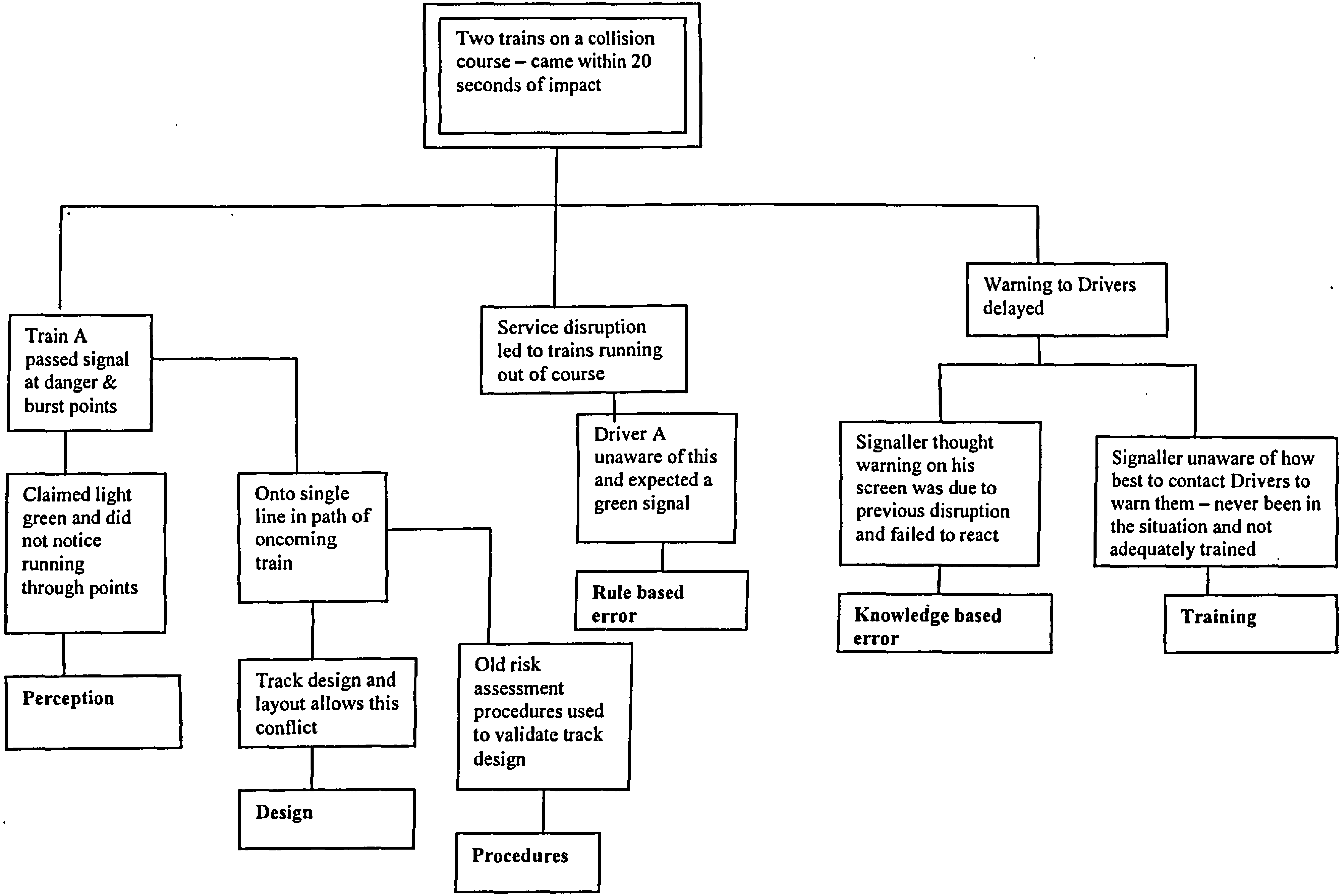
48. Would others be easily identified from this report? Yes No

If yes, state how: ______________________________

49. Is it OK to put this report in the journal?	Yes	No

Appendix 2

Example of event tree and coding


Two trains on a collision course – came within 20 seconds of impact
Train A passed signal at danger & burst points
Claimed light green and did not notice running through points
Perception
Onto single line in path of oncoming train
Track design and layout allows this conflict
Design
Old risk assessment procedures used to validate track design
Procedures
Service disruption led to trains running out of course
Driver A unaware of this and expected a green signal
Rule based error
Warning to Drivers delayed
Signaller thought warning on his screen was due to previous disruption and failed to react
Knowledge based error
Signaller unaware of how best to contact Drivers to warn them – never been in the situation and not adequately trained
Training