MEASURING HAZARD IDENTIFICATION

P. R. Caseley,
Dstl, UK
Fax: (0)1684 771436
email: prcaseley@dstl.gov.uk

Dr. S. Guerra and Dr. P. Froome,
Adelard, UK
Fax: (0)20 7490 9451
aslg@adelard.com, pkdf@adelard.com

Keywords: Hazard Identification, Severity Measurement.

Abstract

This paper discusses an experiment that measured the effectiveness of a hazard identification process used to support safety in Defence Standard 00-56 project. The experimental case study utilised a Ministry of Defence project that assessed simultaneously two potential suppliers who were competing for a MOD equipment contract. The UK MOD Corporate Research Programme funded the comparison work and the MOD Integrated Project Team funded the project which included each contractor's project safety processes.

1 Introduction

MOD has developed a guidance lifecycle for acquisition of their equipment, services and systems (see Figure 1). During the first phase of this lifecycle (concept) a new MOD project explores the operational needs. When these are reasonably clarified, a User Requirement Document (URD) is developed.

On significant projects, contractors are invited to competitively bid to develop a URD into a System Requirements Document (SRD) in the Assessment Phase. During this phase technical solutions may be explored. To improve overall solutions, two or more competing contractors may be selected.

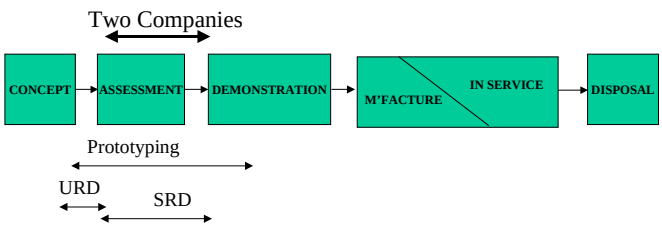

Figure 1 MOD's CADMID procurement lifecycle

If a project is safety related, the contractors will be required to analyse the safety aspects of the system in accordance with Def Stan 00-56 [1]. The competing contractors' Assessment Phase work is funded by the MOD Defence Procurement Agency sponsoring Integrated Project Team (IPT).

Comparing contractors in this phase provides an excellent opportunity to assess the effectiveness and efficiency of some of the early safety processes as both contractors are using the same URD and safety standard.

This paper discusses the results of such a comparison on one project and gives some insight to the effectiveness of the hazard identification process.

The specific for this project equipment will not be discussed but the following should add to the understanding of the system:

- The equipment was fitted to a land based mobile platform.
- It is a medium sized procurement (cost).
- The safety risk of the initial baseline system to be delivered by the Project was assessed as medium-low during the Concept Phase. That is judged to be risk class C, tolerable risk with endorsement of project safety committee, in accordance with Def Stan 00-56.
- Some potential customer enhancements were also proposed for the Assessment Phase and these were to be considered by the contractors. If progressed, they would probably increase the safety risk to class B.

Note: This is just one of eight case studies investigating measuring safety process; the others are discussed in [2].

2 Comparing two sets of hazard identification

The identification of system hazards is one of the most challenging tasks of safety engineering. There is a wide choice of identification methods and techniques, some of

which appear to be more effective in particular industrial domains than others, e.g. chemical process control. Empirical evaluation work of safety analysis techniques within industrial applications by Suokas [3] compared a number of techniques used for hazard identification, including HAZOP, a MOD preferred method. The AIChE has also compared HAZOP against some other popular hazard identification techniques [4] providing effort comparison metrics based on the size of the project.

Most hazard identification techniques depended on the input factors summarised in Figure 2, and usually result in a unique set of hazards documented in a database or list, sometimes referred to as the hazard log. Other products of the hazard identification processes could be accident scenarios and causal factors.

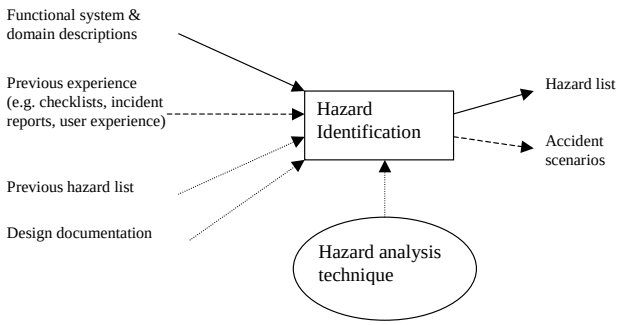

Figure 2 Hazard Identification safety process (Def Stan 00-56 issue 2 describes the initial HI process as "preliminary hazard listing")

2.1 The measurement method

For this MOD project a comparison was carried out on the project deliverable hazard logs. The hazard logs were compared twice, at the mid and end points of the Assessment Phase. Separate Independent Safety Auditors (ISAs) carried out the two successive comparisons.

ISAs are utilised by MOD IPTs and suppliers to assess and audit safety products. They are normally qualified safety professionals with significant experience as safety practitioners. To enable their role, MOD normally negotiates with the contractor access to all design documentation. Generally they are regarded as a trusted third party independent of the contractor and the MOD IPT.

Importantly, the results from the first, midpoint, ISA assessment were withheld from the second ISA comparison assessment. This ensured that their analysis was independent.

The comparison measurement method was necessarily judgmental and relied on the ISA's professional safety engineering knowledge and competence.

2.2 Issues in matching hazards

In the experiments the first step in comparing the two hazard logs was to try and match up equivalent hazard descriptions. The comparisons made by the ISAs could be grouped into four classes:

- Class 1. Easy comparisons, where the wording and meaning were considered equivalent.
- Class 2. More judgmental assessments but where the intent or meaning was clearly the same.
- Class 3. Comparison showed one to many, or many to many, relationships.
- Class 4. Hazards that could not be matched.

The following examples highlight some of the difficulties in matching and classing the two contractor hazard descriptions:

Class 1:

Contractor 1: Inadvertent xxx operation.

Contractor 2: XXX inadvertently activated.

In this example XXX and xxx were easily identified synonyms describing a specialised function traceable to a user requirement of the system.

Class 2:

Contractor 1: Exposure to excessive noise.

Contractor 2: High audible noise level.

This example highlights the problem faced by safety engineers on whether to describe the hazard or a causal factor. In this case the Contractor 1 description could be viewed as a causal factor whereas it could be argued that Contractor 2 describes the hazard. Also, there is a more subtle variation in this case to the hazard identity description. The description used by Contractor 2 could also imply a secondary military hazard, i.e. the exposure of a friendly force position to an enemy due to high noise levels. Contractor 1's description implies only an operator's exposure to high noise levels.

Class 3:

Contractor 1: (three hazards grouped together)
Exposure to stressful working environment.
Inadequate training.
Inadequate Interactive Electronic Technical Publications.

Contractor 2: Human error.

In this case Contractor 1 has defined more specific hazards, which may improve identification of hazard controls and mitigation.

Class 4: These were of great interest as they may be a common hazard that was missed by the other contractor. Two examples were:

Contractor 1: Exposure to excessive vibration.

Contractor 2: Inadequate ventilation.

It would be easy to criticise both contractor teams for missing what appear to be obvious hazards. But it highlights how difficult even identifying simple hazards can be.

Comparing the hazards, particularly Class 1 and 2, took very little time and was not a great overhead for an experienced safety practitioner. Not surprisingly Class 3 hazard matching caused the greatest difficulty.

2.3 Comparing hazards sets using Capture-Recapture

Having matched the hazards it is possible to compare the two sets of hazards by applying a simple Capture-Recapture (CR) analysis. The CR analysis has been successfully exploited by biologists to estimate animal populations. It uses the principle of tagging captured animals and then releasing them into the wild and subsequently attempting to recapture the animals. From the proportion of recaptured animals with tags and those captured without tags an estimate of the total population can be made - an example of this technique used in the software engineering domain can be found in [5].

The same principle can be used for hazard identification if there are at least two independent hazard identification activities. In this case, one hazard identification group is used for "tagging"; the second identification represents the "catch", e.g.:

Hazard identification set 1:	20	(released)
Hazard identification set 2:	30	(captured)
Common Hazards to both sets:	15	(i.e. those recaptured)

Using Hazard identification set 1 as the "tagged" and released, and Hazard identification set 2 as the "catch":

Fraction in "catch" 15 / 30 = 0.5

Total Hazards 20 / 0.5 = 40

The choice of "tag" and "catch" is arbitrary as the formula for estimating the total number of hazards is simply:

$$N_{total} = N_1 .N_2 / N_{12} \quad (1)$$

where N_1 and N_2 are the number of hazards found in the two hazard identifications, and N_{12} is the number of common hazards.

The detection efficiency of each hazard identification E_i is therefore:

$$N_i / N_{total} \text{ , i.e.: } E_i = N_{ij} / N_j \quad (2)$$

Applying the CR method depends on an assumption that hazard identification processes are independent and unbiased. If some hazards are easier to find than others (analogous to the tagged animals being easier to re-catch) the results can be biased. For example if:

$N_1 = 30$, $N_2 = 20$, $N_{12} = 20$

then Hazard Identification set 2 would be a subset of Hazard Identification set 1. This may be because Hazard Identification 1 was perfect (E_1=100%), but it is also possible that Hazard identification 2 was fairly cursory and only picked up the "easy" hazards and these are more likely to be common to the hazards in the Hazard Identification set 1.

2.4 Alternative analysis - Suokas

Suokas [3] compared a number of techniques used for hazard identification, including a method often used by MOD contractors - HAZOP study. Suokas attempted to derive measurements for reliability, coverage and validity of safety analysis techniques.

The experimental case studies research results in this work indicate that the effects of competency of analysts, diverse analysis techniques and complexity of systems impact the quality of safety product. The case studies showed that:

- Some hazards cannot be easily identified by some hazard identification techniques.
- Competency and experience can affect the analysis, although not always as predicted – in one case the most experienced practitioner found the least hazards.
- The nature of the system probably influences the results – in one case over 40% of hazards of one sub-system were not identified.

For the most part, the studies depended on multiple teams using diverse techniques as well as using accident and incident data as a feedback mechanism to confirm existing hazards or identify new ones. Suokas also identified that the quality of the data about the system or work practice is also a factor in determining hazards. To experienced safety practitioners, these are perhaps unsurprising findings. What is important is that the work attempted to measure the effect, and proposed that formula for inter-analyst metrics of reliability, coverage and validity could be derived:

Reliability = (N_{HI} by a test person or Team)/(TN_{HI} in the experiment) (3)

Coverage = (N_{HI} by Method examined)/(TN_{HI} in the experiment (and belonging to the search pattern of the method examined)) (4)

Validity = (N_{HI} by the method examined and in accident reports)/(TN_{HI} in accident reports) (5)

N_{HI}, TN_{HI} = Number and Total number of hazard identified

The reliability measure (3) is very similar to efficiency calculation (2) except the denominator N_{total} is an estimate of the total number of hazards. Ideally a prediction from either (2) or (3) of hazard identification process should closely match the value of the validity measure (5). However, it is very difficult to give a true validity measure (5) as, even after decommissioning a system, the accidents may not have revealed all hazards. As with all measures these metrics are only indicators not absolutes so should be used cautiously.

3 Success of Hazard Identification

Tables 1 and 2 show the number of hazards identified by the two contractors. Table 1 is the comparison performed at mid point and Table 2 that performed at end point of the Assessment Phase. The results in the tables indicate a large discrepancy and potentially many missing hazards. Further

analysis and additional contextual information alleviated much of the concern, for example some of the discrepancy could be explained by the different design solutions. Nevertheless, at the end of the Assessment Phase, the MOD IPT advised the successful contractor of some of their potential omissions.

	Num hazards (2)	**Num hazards (1)**
Con 1	46	45
Con 2	40	33
Common	22	22
Found Tot	64	56
Suokas Reliability (3)	(60.6%-62.5%)	(80.3%-58.9%)
Est Tot (1)	*83.6*	*67.5*
Eff (2)	*(48%-55.4%)*	*(48.9%-66.6%)*

Table 1 estimate of the total number of hazards and efficiency using CR at the mid point of Assessment Phase. Note: hazards in column (2) are higher because they include the additional customer enhancements, which were later abandoned.

	Number of hazards
Con 1	40
Con2	41
Common	35
Found Tot	46
Suokas Reliability (3)	86.9%-89.1%
Est Tot (1)	*46.86*
Eff (2)	*85%-87.5%*

Table 2 estimate of the total number of hazards and efficiency using CR at the end of the Assessment Phase. Carried out by the second ISA.

An improvement in efficiency and reliability is indicated when the two tables are compared. This is possibly a result of the hazard identification process maturing and is supported by the fact that the number of hazards fell by 5 for Contractor 1 but rose by 8 for Contractor 2. The subjective nature of hazard matching must be considered when using these results, nevertheless they were very encouraging.

3.1 Comparing results with other experiments

The efficiency metric compares favourably with the inter-analyst reliability metric. If we assume that Table 2 corresponds to a maturing of the hazard identification process, the efficiency metric (2) gives a more conservative estimate of effect of the hazard identification process than the reliability metric (3).

Suokas's experimental results indicated that measured inter-analyst reliability reached 100% for identified hazards when using HAZOP in only one of the three experimental studies. This reliability dropped to as low as 50% when factors such as deviations (events leading to possible accidents) and determining factors (constant safety pitfalls) were taken into account. Corresponding coverage and validity for HAZOP were measured at 89% and 71% respectively. Unfortunately, there is no further published evidence of Suokas's measures being used in industry.

Suokas's coverage measure is modified dependent on the process used for identification and the validity measure is dependent on accident data, which is often only available towards the end of a system's life. The CR method produces an estimated value and gives a higher estimated total than simply counting the identified hazards, so it could be used as a predicted validity measure.

Table 2 gives very similar results for reliability and efficiency as the estimated total of hazards is 46.86 and the counted total is 46. This estimated figure relies on class 3 comparison issues being minimal.

It is not possible to apply the coverage measure (4), in this case, as it is aimed at techniques rather than number of hazards identified. The contract teams used HAZOP along with checklists and past experience of other similar projects.

All the described measures do indicate that the hazard identification process and associated technology is not perfect. The measures can be used as a basis for quantified hazard identification process improvement. For organisations who assess their maturity using schemes such as +Safe [9] then both (2) and (3) may be useful predictors of process improvement. The major disadvantage is that both measures need at least two independent analysis teams. This is not a major problem for many projects following the MOD CADMID lifecycle, as the element of competition facilitates duplication of analysis and design.

For procurers and developers without this luxury, it may be possible to estimate the effectiveness of hazard identification by comparing the results of previous projects or by tracking the identification of hazards against the project schedule. For

example, Figure 3, reproduced from Watt [7], tracks the progress of hazard from identification to closure.

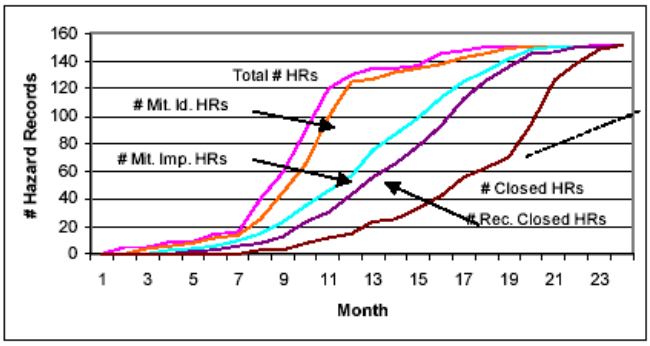

Figure 3 tracking of hazards

Watt found that tracking this indicator was a useful predictor in managing project safety, but it could also be used to see the effects of specific analysis meetings such as a HAZOP on hazard identification.

3.2 Assessing severity analysis

The identification of a common set of hazards enables comparison of other aspects of risk assessment and safety analysis processes. The severity hazard comparison results below allow some assessment of the next phase of safety analysis. They indicate how much reliance a project can place on the accuracy of risk assessment.

Using Contractor 1 as the base for the 35 hazard matches (Table 2) and comparing the base hazards against the matched Contractor 2 hazards:

- 23 hazards could be traced to matching severity (65.7% agreement)
- 7 were off by 1 degree e.g. catastrophic = critical (20%)
- 3 were off by 2 degrees e.g. negligible = critical (8.6%)
- 2 were off by 3 degrees negligible = catastrophic (5.7%)

The discrepancies between the severities of the accidents identified by the two contractors prompted investigation into the significance of deviations, in particular those off by 3 degrees severity. In most cases there was a reasonable explanation, for example one contractor solution used more toxic materials than the other hence the severity difference was greater for a common hazard.

3.3 Other factors effecting hazard identification (skill and effort)

When comparing the teams carrying out hazard identification, factors such as resource, effort and competence should also be considered. A contractor expending more effort and with greater experience should produce better results.

At the end of the assessment phase (Figure 1) both contractors were assessed to see who would progress contractually. The safety element of both contractor submissions exceeded the MOD minimum requirement when judged at the end of the Assessment Phase by an IPT nominated MOD Assessor. There was some turnover of safety management personnel in Contractor 2, however, the key safety analysts for both contractor teams were retained throughout the Assessment Phase.

For both contractors, these analysts fulfilled dedicated safety engineering roles for their companies and had demonstrable safety experience on other development projects. Although not assessed by a competency scheme, both teams provided comprehensive CVs and work histories of their principal safety engineers who were judged by the MOD assessor to be at least at the competency level of safety practitioner [8] for their functional roles.

The expended efforts of both teams for hazard identification were also similar. Unfortunately, the effort measures lacked granularity to identify the hazard identification tasks specifically. However, the effort spent on three tasks related to hazard identification were recorded, namely: producing Preliminary Hazard List, setting Safety Criteria and documenting the results in the initial Safety Case, which totalled to 344 hours (Contractor 1) and 350 hours (Contractor 2) – a difference of less than 2%.

Another comparison effort point was the initial set up of a hazard log, which was 165 hours for Contractor 1 and 200 hours for Contractor 2. This difference could be attributed to the use of different hazard management tools.

4 Comparing and Utilisation

The analysis comparing the two hazard log data sets was relatively simple and proved to be very useful. It predicted the efficiency of hazard identification and confirmed the findings of other research. The severity analysis indicated that the risk assessment is not consistent but some of the inconsistency could be explained by detailed analysis of the design. The subsequent additional analysis, highlighted by the discrepancies, led to improved confidence and validity of the hazard identification and assessment.

A number of common factors also added to confidence and validity, they were:

- The common user requirement. The design solutions may have been different but these are more likely to influence mitigation or severity of the hazard rather than the identification – the hazard toxic material is an example.
- The application of a single standard. The techniques and processes applied by both teams were similar and they were contractually bound to deliver specified safety documents.
- The use of similar standard personnel. Both teams were judged as being of a similar standard and the final selection marks showed compliance with the project safety requirement.

- The close correlation of effort and resources. This indicates that the results were not influenced by budget.

The results above are from only one case study, but other work [6,7] discussed indicates that achieving identification of 85-86% hazards early in a project may not be an unreasonable result. This imperfect result justifies the continuation and refinement of research into improved safety analysis. It also leads to questions on what the accuracy level for hazard identification should be at various phases of a development lifecycle. This is important to all aspects of a system development as a hazard identified later in the development may need to have its mitigation defined (requirement), implemented (design) and demonstrated (test and procedures). Therefore for safety critical systems, these measures/metrics would be useful predictors of remaining derived safety requirement growth, and design and validation effort.

Overall this case study has given MOD and the contractors involved insight into safety processes in the areas of effectiveness of hazard identification, accuracy of severity assessment, and predicting costs of initial hazard analysis processes.

The case study experiment indicates how effective hazard identification processes are when using the same standard and requirements. The analysis of hazard logs is neither difficult nor costly, and the measures can highlight inconsistency for procurers and developers and demonstrate the application of a standard to the regulator.

We believe that the efficiency and reliability indicators can be applied in almost every circumstance where two teams are identifying hazards for the same system. They are useful to managers, designers and safety engineers.

References

[1] Def Stan 0056, *Defence Standard 0056 "Safety management Requirements for Defence Systems"*. Issue 2, 1996. Note this standard is now Issue 3, Dec 2004.

[2] Caseley, P. R., Clark, G., Murdoch, J. and Powell, A, TR11459 V1, The Measurement for Safety Processes and ALARP Volume 2 dated 26 November 2004.

[3] Suokas, J. *On the reliability and validity of safety analysis*, 1985, Espoo, Technical Research Centre of Finland.

[4] American Institute of Chemical Engineers (AIChE), *Hazard Evaluation Procedures*, Second Edition with Worked Examples, 1992.

[5] Barnard, J., Emam, K. E., Zubrow, D., *Using Capture-Recapture Models for the Reinspection Decision.* Quality Management, 2003. **5**(2).

[6] Suokas, J., Rouhiainen, V., *Quality control in safety and risk analyses*. J. Loss Prev. Process Ind., 1989. **2**(April): p. 67-77

[7] Watt, G. *Metrics for Assessing Safety Program Effectiveness in Hazard Identification and Resolution*. in *21st International System Safety Conference*. 2003: System Safety Society.

[8] IEE/BCS, *Safety, Competency and commitment, competency guidelines for Safety related systems practitioners.* 1999, IEE.

[9] Bofinger, M., et al. Experience with Extending CMMI for Safety Related Applications. in_12th International Symposium of the International Council on Systems Engineering (INCOSE'02). 28th July - 1st August 2002 . Las Vegas, Nevada: INCOSE. 2002.