Risk Assessment in the Wild

Andrew Rae and Richard Hawkins
Department of Computer Science
University of York
Heslington, York, United Kingdom

ajrae@cs.york.ac.uk

Abstract
This paper introduces an empirical research method for systems engineering based on the examination of work products. To illustrate the method we describe an investigation of safety risk assessment as it is actually recorded, rather than the standards, forms and procedures used to guide risk assessment. A body of risk assessments was collected via a combination of public search, freedom of information request, and private request. The risk assessments are from multiple domains, for multiple purposes, and follow diverse formats – the one thing that they have in common is that they are genuine work products.

Due to the necessarily arbitrary selection process, the collection cannot resolve quantitative hypotheses about the distribution of phenomena. However, it provides an opportunity to explore assumptions and suspicions about the real-world conduct of risk assessment that cannot be examined by looking at academic literature or guidance documents.

The paper makes contributions in three areas:

- Our early findings about the characteristics of the risk assessment collection
- Our experiences with the exercise itself, and the lessons learnt which may be helpful in future similar research
- Observations on the relationship between theoretical and applied system safety, and the methods that may be applied to answer important questions in each sphere.

Keywords: Risk Assessment, Empirical, Research Methods

1 Introduction – The Empirical Approach to System Safety Research

System safety is a relatively young engineering discipline. Whilst concern about accidents is long-standing (ASSE 2011), the start of modern system safety engineering is generally dated to the invention of Fault Tree Analysis in 1965 (Dhillon 1982). The body of knowledge in system safety, typical of many young disciplines, is populated by beliefs and techniques drawn from experience or constructed based on plausible theories. As we showed in a study of current practice (Rae et al. 2010) there is a limited basis of evaluation evidence for supporting or pruning the body of knowledge. Progress in an academic discipline is characterised by refining and replacing knowledge in the light of new evidence, and system safety currently lacks strong mechanisms for testing the knowledge we have.

System safety is inherently a "soft science", populated by researchers drawn mainly from a "hard science" engineering background. Faced with questions requiring soft science research methodologies, the field has concentrated on activities not involving empirical study. It is possible that discomfort with research methods seen as "unscientific" has led to a failure to recognise the developing body of work in the traditionally soft sciences aimed at tackling empirical difficulties.

The traditional distinction between "hard" and "soft science is discussed by Howard under the criteria of empirical cumulativeness and predictive accuracy (Howard 1993). Empirical cumulativeness is the reliability with which experiments produce results which are consistent with each other (Hedges 1987). Predictive accuracy describes how well a theory can predict the outcome of a real-world interaction. For example, a theory in psychology might correctly predict the outcome of an event 70% of the time – this shows low predictive accuracy. On the other hand, different experiments to validate the theory might consistently produce this 70% result – this shows empirical cumulativeness. Howard argues that soft sciences may have low predictive accuracy due to the fact that the phenomena being studied have a large number of interacting causes, making it difficult to comprehensively account for variation in observations.

On the criterion of predictive accuracy, system safety engineering is inevitably a soft science. The safety of a system emerges from a large number of interacting causes, and the precise characterisation of these causes and interactions is far beyond the state of the art. Consider, for example, the measurement of "safety culture" (Guldenmund 2000) . Even if safety culture could be fully characterised (it cannot) or reliably measured (it cannot), culture would be only one among many factors determining accident rates. This does not make safety culture an unscientific concept. If we could establish empirical cumulativeness by finding a repeatable measure of safety culture, and establishing a reliable correlation between safety culture and accidents, then the fact that the correlation is not 100% does not make the relationship less real.

System safety engineering is growing in scope and importance (see for example the recent introduction of ISO 26262 (ISO 2011)). It is important that the body of knowledge continues to grow in parallel. This growth requires empirical foundations.

In Section 2 of this paper we discuss the ways in which system safety research can be supported by empirical research, and introduce risk assessment as a running example of a topic needing empirical support.

In Section 3 we introduce the "Safety Menagerie" as a research method. In Section 4 we show how the research method can be applied to questions about risk assessment, and provide indicative findings. The contribution of the paper is not these findings, but the conclusions reached about the research approach itself, which are provided in Section 5 and discussed in Section 6.

2 Observation and Measurement

Alexander (2010) discusses the range of system safety research goals, and the suitability of various research methods for addressing these goals. Broadly speaking, there are two main categories of goals which can be best supported by empirical research:

1. Evaluation of methods and techniques; and
2. Observation and measurement of current practices.

Each of these goals requires knowledge to be shared across what Alexander refers to as the "research/practice boundary". In evaluation research, the techniques are transferred to industry, with information about efficacy returned across the boundary. In observation and measurement research, the challenge is to gain an accurate view of industry practice, to provide grounding for development of new theories and techniques. Trevelyan (2007) describes important discrepancies between the way engineers describe their work and the actual practice of that work. Thus, instruments such as surveys are seldom suitable for acquiring the necessary insight.

The use of social science methods in studies of engineering practice has received considerable recent attention (Ahmed 2007). This is particularly the case in software engineering, which has many features in common with safety engineering. The performance of a software project is influenced by many poorly understood factors, making it difficult to isolate single factors for systematic study (Wohlin et al. 2003).

Where there is a close relationship between the effect to be studied and the environment in which it occurs, case studies are considered the most appropriate research method (Creswell 2007). The most common forms of engineering case study are participant observation and action research. These methods have produced interesting results, but are limited in scope to a small number of workplaces, creating external validity problems for many questions of research interest.

In order to refine existing safety engineering methods, design new methods, and improve education, it is desirable to understand how safety engineering is currently practiced. This is particularly the case if there is divergence between "best practice" as described in the literature, and "industry practice" occurring in real-world organisations.

This paper focuses on the practice of risk assessment. For this practice, there are research questions of interest that can only be answered by observing the real world. The high level questions concern the value-add of risk assessment as an activity.

1. How often does risk assessment lead to implementation of improvements to a system or operations?
2. To what extent are the outcomes of risk assessment predetermined or expected before the assessment is conducted?
3. Are the hazards of a system better understood after risk assessment?

We may also be concerned with what makes a good risk assessment.

4. Are there elements of a risk assessment currently considered important which do not influence the outcomes?
5. Are some methods of risk assessment more effective than others?
6. How does the practice of risk assessment vary between industries? Between different types of risk? Between different system technology?

In designing guidance and education, we may be interested in the practical shortcomings of risk assessment.

7. To what extent do risk assessments document their assumptions?
8. What types of uncertainty are treated well or poorly in risk assessments?
9. Are internal inconsistencies common in risk assessment?
10. Do risk assessments commonly cite evidence in support of estimates used?
11. Is the theoretical division between risk assessment and risk acceptance preserved in practice?
12. Are mitigations selected systematically or arbitrarily?

Further, risk assessment may reveal beliefs and attitudes held by those who perform the assessment.

13. Is risk aggregated, or is each source of risk treated atomically?
14. What types of risk are considered in scope and out of scope?
15. Is risk identification part of risk assessment, or are the important risks considered to be already identified?
16. What language is used in talking about risk and in drawing conclusions?

Finally, we may be interested in what risk assessment reveals about those performing the risk assessment.

17. Do risk assessments for systems involved in accidents look different from risk assessments for other systems?
18. Is the style and language of risk assessment an indication of safety culture?

3 The Safety Menagerie Method

3.1 Purpose

In the work reported in this paper we are trialling a method of observational research based on work products. Document analysis is frequently used in ethnography to extend and add detail to interviews and observations (Creswell 2007). Such analysis typically focuses on the way culture is revealed through features of the document. For our present research we are interested in documents as records of practice. This is not as direct as actually observing practice, but covers many more situations for the same research effort. Rather than examining one work situation through a clear lens, we observe many situations through a foggy window.

The results reported in this paper are preliminary. Throughout the work we were as much concerned with testing and improving the research methods as we were with the research questions. The main question addressed by this paper is "Can existing safety work products be used as research objects to learn about and improve the practices of system safety?"

The specific work products the paper is concerned with are risk assessment reports. Risk assessment is a natural starting point for exploring real world safety engineering practices because:

- it is widely practiced;
- it is typically well documented in a single report; and
- risk assessment reports are often treated as public or non-confidential documents.

3.2 Data Collection

For convenience of reference, the data set for this project is referred to collectively as the "System Safety Zoo – Risk Assessment Reports" (SafeZoo-RAR). Each item in SafeZoo-RAR is self-described as a report of a risk assessment activity. Exactly what is meant by "risk assessment" varies between items, as is discussed further below. SafeZoo-RAR has been assembled by a non-systematic search process combining solicitation and search-engine approaches. The items have been made public by a number of mechanisms:

- regulations which require publication of risk assessments;
- Freedom of Information requests;
- government or local authority information policy;
- publication in support of press-releases;
- provision in response to informal requests for information; and
- publication for no apparent deliberate purpose.

The combination of search method and publication methods means that SafeZoo-RAR is not systematically representative of all risk assessment reports. It is likely to be biased towards industries and organisations with an interest in public disclosure, and in many cases the knowledge that the reports could become publicly available may have influenced their content.

A known bias in the sample is that it excludes industry groups with policy directly requiring secrecy of risk assessments. Specific examples are major hazardous facilities (where revealing risk assessments is considered to compromise national security) and medical devices (where risk assessments are considered proprietary information).

3.3 Composition of SafeZoo-RAR

SafeZoo-RAR consists of approximately one hundred risk assessments. The exact size is fluid – new risk assessments are added to the collection as they are obtained. A permanent method of open access to SafeZoo-RAR has not yet been found. Most of the reports have not been formally published, so there is no reliable method for other researchers to recreate the data set from the names of the reports. However, we do not have license to redistribute the individual reports.

Access to data is an important issue for the research methods we are trialling. Unlike case study research, where replication can be achieved through comparable case studies, researchers attempting to replicate any of our results will need access to the original data set. Whilst in theory a new data set could be assembled, this will only be possible if there are many reports which are readily findable but not found by the SafeZoo-RAR search. For the purpose of current publication we have summarised the reports described in this paper in **Error! Reference source not found.** and will provide access on request.

Thirty of the reports within SafeZoo-RAR have been classified according to questions of interest. The details captured for each report are:

1. Title or identifier
2. Purpose
3. Jurisdiction
4. Source of Harm
5. Size
6. Whether the report includes a quantitative assessment of risk
7. Whether the report includes risk identification, or is based on previously identified risks
8. Whether the report discusses uncertainty
9. Whether the report discusses risk acceptability
10. Whether the report documents assumptions
11. Whether the report recommends actions
12. Whether the report discusses rejected actions
13. What targets of risk the report considers

The full SafeZoo-RAR has not been classified, to allow tentative conclusions drawn from this initial set to be tested on further reports.

3.4 Data Analysis

The process of analysis is based on iterative test and improvement of models. Firstly, models are created. These models reflect how we as researchers "expect" that risk assessment is conducted. From these models testable hypotheses are formed – questions that can be asked of the SafeZoo-RAR and answered in the affirmative or negative. These questions are then applied to a subset of SafeZoo-RAR. The result is a set of surviving hypotheses, as well as insights gained from the falsified hypotheses. These are used to form new models and the process is repeated. Several illustrations of this process are provided in Section 4.

4 Application of the Method

Section 2 discusses a range of questions which can only be addressed with real-world data about risk assessment.

Among these questions is the relationship between theoretical models of risk assessment and actual risk assessment practices. In this section we show how this relationship can be explored using the SafeZoo method. We take two theoretical models for risk assessment, identify measurable features of the models, and search for those features within a subset of SafeZoo-RAR.

4.1 The Red Book Model

"Risk Assessment in the Federal Government: Managing the Process" (Committee on the Institutional Means for Assessment of Risks to Public Health, National Research Council 1983), informally known as the "Red Book", describes risk assessment as a scientific process which is conceptually and managerially distinct from the political process of risk treatment or acceptance. This is a theoretical model of how risk assessment is conducted. If the model matches reality, there are several hypotheses which would be confirmed. For example

a) Risk assessments would not contain statements of risk acceptability
b) Risk assessment conclusions would be reported in a way which did not imply acceptability or unacceptability
c) Risk assessment would contain statements about uncertainty which indicate whether the conclusions are certain enough to allow decisions about acceptability

These hypotheses can be tested to see whether the model fits each item in a subset of SafeZoo-RAR. Because of the inherent bias in the set, we cannot draw quantitative conclusions, but the overall usefulness of the model can be explored.

When the hypotheses were applied to twenty-three risk assessment reports, ten contained explicit statements of risk acceptability. A further two reports strongly implied acceptability in their conclusions. Three reports quantitatively compared risk to pre-determined benchmarks. Of the remaining reports, six recommended actions in response to the risk, implying that residual risk would be acceptable if the actions were taken. In only two cases was the risk assessed without any implied judgement of the acceptability of the risk.

Nine of the reports discussed uncertainty. In three cases it was explicitly stated that the conclusions could or could not be relied upon based on the amount of uncertainty. In the other cases causes of uncertainty were discussed without making judgements on the acceptability of the uncertainty.

From these findings, two conclusions can be tentatively reached. Firstly, the model of separation between assessment and acceptability is not generally applicable. Secondly, where the model might apply, knowledge about levels of acceptability is often available, informing (and arguably influencing) the risk assessment.

4.2 ALARP Model

"As Low as Reasonably Practicable" (ALARP) is the principle applied in order to meet the United Kingdom (Health and Safety Executive 2001) legal benchmark for risk reduction. At the heart of any practical application of ALARP is consideration of alternative risk reduction strategies (Redmill 2010). Whilst ALARP is only required for certain legal jurisdictions, it is applied more widely, and it is appropriate to consider the extent to which it is used within SafeZoo-RAR.

For assessments which include recommended actions, the ALARP model predicts that the reports would include discussion of risk control measures that are not recommended. This is because in order to determine that risk is ALARP the report must explain why further risk reduction is not practicable.

To test this hypothesis, thirteen reports containing recommendations were considered. Of these reports, only two discussed rejected options. One of these reports was written for the primary purpose of making a selection from several options.

From this result, it can be concluded that ALARP is not a generally applied method of choosing which mitigations to recommend. It would not be appropriate due to the small number of UK reports in the sample (five) to conclude that ALARP is generally not correctly applied in jurisdictions where it is a legal requirement – this would require a larger set of reports all from the same jurisdiction.

5 Strengths and Limitations of the Approach

The approach described in Section 3 and 4 has some inherent strengths and weaknesses described here. Whilst the strengths and weaknesses are apparent in our use of the approach so far, we have insufficient evidence to support or reject claims about the overall efficacy or efficiency of the research method.

For any given risk assessment report, there are objective questions which can be answered. We can explore the methods used to conduct the assessment, the scope of the assessment, whether the report contains common features that undermine risk assessments, and the way the assessment is reported. We may also be able to explore more subjective questions about the values and attitudes reflected in the language of the report and its conclusions.

There are also questions which we cannot answer about each report. Unless specifically mentioned, we cannot know about preparation or training for the risk assessment, and context such as procedures or norms that guided the assessment. We cannot know what decisions were supported by the assessment, or even if the report is an accurate representation of the assessment itself.

To extend the validity of findings beyond the scope of a single report, it is necessary to find patterns within the reports, and then to test these findings on further reports. Without evidence that the data set is representative, there will be a need for more systematic investigation of models that have passed this initial attempt at falsification.

The main strength of the approach is that it provides insight into safety methods as they are practiced rather than as they are academically described. As researchers who are heavily engaged in safety teaching, we are equally interested in evaluating what constitutes good practice, and the weaknesses of current practice.

Beyond individual techniques, we have the opportunity to examine a snapshot of the decision making processes of organisations attempting to manage risk. The existing body of work on sociology of organisations in the lead-up to accidents (Pidgeon 1991) suggests that leaders are forced to apply a form of `bounded rationality' when they think about risk. They cannot pay attention to everything, so it makes sense to devote resources to what they see as important. If the resources are mis-allocated, it appears as if the leaders were wilfully blind to some hazards. Through study of the risk assessments we can see what different organisations consider to be important risks, and how they discuss risks of different types. We can see the basis on which they choose to filter risks, prioritise risks, and determine the adequacy of risk mitigation.

6 Discussion and Observations

There is a large volume of safety work products held within organisations. Each item taken separately may seem of limited research value, but together they provide a cost-effective way of examining safety engineering practice. One fault tree is just a fault tree, but ten fault trees may provide a description of the way fault trees are used, and twenty fault trees may explain the mistakes commonly made in fault trees, and lead to better guidance.

Throughout this work we have been pleasantly surprised by the amount of material we have been able to access. Freedom of Information enquiries have been on occasion refused, and more often simply ignored, but most direct requests for examples or documents referred to in the media have met with positive responses.

7 Further Work

The research approach has proved practical, but has not yet yielded significant results. It is reported here for peer review of the method, and to provide encouragement to others to engage with empirical system safety research.

Our immediate ongoing work is exploring the representation of uncertainty in risk assessments. Initially, we were surprised by the fact that more than half of the reports, including all of those which present quantitative risk data, discuss uncertainty. Prior to this finding we expected that uncertainty would be ignored in most reports. Uncertainty, however, is invariably discussed only in terms of source data. Methodological uncertainty, including fallibility of the risk assessors themselves, is invariably omitted. This is only a tentative conclusion, but we are working on further comparisons of ideal treatment of uncertainty with the sample of reports.

8 References

Ahmed, S., 2007. Empirical research in engineering practice. *J. of Design Research*, 6(3), pp.359 – 380.

Alexander, R.D., Rae, A.J. & Nicholson, M., 2010. Matching Goals and Methods in System Safety Engineering. In *IET System Safety*.

ASSE, 2011. A Brief History of the American Society of Safety Engineers. *American Society of Safety Engineers*. Available at: http://www.asse.org/about/history.php [Accessed May 17, 2011].

Committee on the Institutional Means for Assessment of Risks to Public Health, National Research Council, 1983. *Risk Assessment in the Federal Government: Managing the Process*, Washington, D.C.: The National Academies Press.

Creswell, J.W., 2007. Qualitative inquiry & research design: choosing among five approaches, Sage Publications.

Dhillon, B.S., 1982. Systems safety: A survey. *Microelectronics Reliability*, 22(2), pp.265–275.

Guldenmund, F.W., 2000. The nature of safety culture: a review of theory and research. *Safety Science*, 34(1-3), pp.215–257.

Health and Safety Executive, 2001. *Reducing Risk Protecting People*, HSEBooks.

Hedges, L.V., 1987. How hard is hard science, how soft is soft science. *American Psychologist*, 42(2), pp.443–455.

Howard, G.S., 1993. When psychology looks like a 'soft' science, it's for good reason. *Journal of Theoretical and Philosophical Psychology*, 13(1), pp.42–47.

International Organization for Standardization, 2011. *ISO 26262 Functional Safety*

Pidgeon, N.F., 1991. Safety Culture and Risk Management in Organizations. *Journal of Cross-Cultural Psychology*, 22(1), pp.129 –140.

Rae, A.J., Nicholson, M. & Alexander, R.., 2010. The State of Practice in System Safety Research Evaluation. In IET System Safety. Manchester.

Redmill, F., 2010. *ALARP Explored*, CS-TR 1197. Available at: http://www.cs.ncl.ac.uk/publications/techreports.

Trevelyan, J., 2007. Technical Coordination in Engineering Practice. *Journal of Engineering Education*, 96(3), pp.191–204.

Wohlin, C., Höst, M. & Henningsson, K., 2003. Empirical Research Methods in Software Engineering. In *Empirical Methods and Studies in Software Engineering*. pp. 7–23. Available at: http://www.springerlink.com/content/UFKGYWVMMVBTPC4M [Accessed August 5, 2010].