Reliability Engineering and System Safety 125 (2014) 67–81

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment

Andrew Rae, Rob Alexander*, John McDermid [1]

Department of Computer Science, University of York, Deramore Lane, York YO10 5GH, United Kingdom

ARTICLE INFO

Available online 3 October 2013

Keywords:
Quantitative risk assessment
Probabilistic risk assessment
Evaluation
Validity
Process maturity
QRA
PRA

ABSTRACT

Quantitative risk assessment (QRA) is widely practiced in system safety, but there is insufficient evidence that QRA in general is fit for purpose. Defenders of QRA draw a distinction between poor or misused QRA and correct, appropriately used QRA, but this distinction is only useful if we have robust ways to identify the flaws in an individual QRA. In this paper we present a comprehensive maturity model for QRA which covers all the potential flaws discussed in the risk assessment literature and in a collection of risk assessment peer reviews. We provide initial validation of the completeness and realism of the model.

Our risk assessment maturity model provides a way to prioritise both process development within an organisation and empirical research within the QRA community.

1. Introduction

Quantitative risk assessment (QRA) is widely used to support system safety decision making by industry, regulators and government. QRA has been heavily criticised by academics [1–4], working engineers [5,6] and public welfare campaigners [7,8]. Attempts to rebut the criticisms (e.g. Apostolakis [9]) rely on distinguishing “good” QRA from “bad” QRA, a review task which is itself a black art. Our perception is that QRA remains widely practiced because of inertia and uncertainty (benefit of the doubt), not because of justified belief in its validity. Our aims in this paper are to argue in support of this perception and to provide a maturity model that shows a path to justifiable QRA practice.

QRA is used in many domains for many different purposes; in this paper we are solely concerned with its use in safety. We do not make claims about the utility or otherwise of QRA in other domains. Due to our focus on safety we are concerned with risk associated with hazards – states of the system which can cause harm, especially loss of life or injury, without anything else needing to go wrong. In some domains the term “risk scenario” is used; because of our focus on safety we use the term “hazard” throughout.

In terms of years, QRA is a mature discipline – fault tree analysis has been in use since the Minuteman Missile and Boeing 747 development projects in the 1960s [10], and probabilistic nuclear safety assessment began at a similar time [11]. Actual maturity of science, however, does not come from age; it comes from revision and correction as weaknesses in theories and methods are identified and resolved. In this respect, QRA has led a charmed life – it has been subject to little empirical evaluation and little critical review. Empirical studies do not seem to have influenced the actual conduct of QRA [12]. Many instances of QRA are never tested: QRA is most important for assessing safety-critical behaviour in circumstances where we are very unwilling for the top event to happen even once (we do not tolerate nuclear meltdowns). When a QRA predicts that an accident will occur only every million years, not having that accident in a plant's 40-year lifetime is negligible evidence that the QRA was correct. We thus have the combination of little empirical study with little natural feedback – a situation which leaves us in almost total darkness as to the validity and efficacy of QRA.

An important question is thus “how can we make things better?” A key first step is to understand the flaws – to systematically understand the full breadth of ways in which QRA can go wrong. A second step is to understand the relative importance of those flaws so that we can prioritise research into them.

In this paper, we first summarise the empirical evidence on the validity of QRA, and we show that it is inadequate given the strong claims that QRA users are making. We then present a comprehensive classification of possible flaws in QRA, drawing on those described in a wide range of published sources. We have assessed the validity of this error set by noting whether the flaws occur in a set of peer reviews of real-world risk assessments, or in our own industrial experience. While there are many previous “most important errors” or “most common error” lists (e.g. [7,13–15]), we are aware of no previous classification that even claims to be comprehensive.

Our goal in this paper is not to add to the criticism of QRA. We provide instead a constructive way forward – a maturity model for assessing and improving QRA. In order to qualify for a given level,

* Corresponding author. Tel.: +44 1904 325 474, +44 7813 134 388.
E-mail address: rob.alexander@york.ac.uk (R. Alexander).

[1] John McDermid has an OBE and is a Fellow of the Royal Academy of Engineering.

0951-8320/$ - see front matter
http://dx.doi.org/10.1016/j.ress.2013.09.008

a QRA process and report must be free of the flaws in all levels below. By organising the flaws in this way, our maturity model provides a roadmap both for organisations to develop their QRA practices and for researchers to target the most important real-world problems.

2. There is inadequate empirical support for quantitative risk assessment

2.1. We believe that QRA is valuable, but we cannot show it

We will start by setting out some of our (the authors') views on QRA, in order to set our assessment of practices in context. We believe that QRA, done well, probably helps people design safer systems. Furthermore, assessing the absolute size of risk is a necessary part of a well-regulated safety process, and by definition this activity is "QRA". Therefore we cannot say "do not do QRA"; we can, however, say "be sceptical about the validity and hence value of any instance of QRA".

We will spell out in detail, through the rest of this section, what the evidence for and against QRA says. In summary, it suggests that

(a) Effort should be placed on improving the quality of QRA performance, in particular on providing tools and mechanisms for review of QRA.
(b) QRA should be used in ways that match its strengths – the focus of analysis should be on finding ways to manage risk.
(c) QRA should not be used in ways that place excessive reliance on its accuracy.
(d) We need more research into the quality and attributes of QRA as it is practiced.

Many of the flaws in QRA are equally applicable (and probably prevalent) in non-quantitative (qualitative) risk assessment. It is not our intent here to promote a qualitative alternative to QRA.

In order to investigate QRA, we need to distinguish between several different concepts.

1. The properties QRA has as a set of practices.
2. The properties of any specific instance of QRA (e.g. "the QRA performed on a proposed expansion of chemical plant X by safety consultancy Y between June and September 2009").
3. The properties shared by sets of specific examples of QRA (e.g. "large-scale QRAs in the UK chemical process industry between 2008 and 2013").

In order to investigate QRA, we will need to examine instances of QRA. Some QRA instances will be better than others, and specific criticisms levelled at QRA may be true for some instances but not for others.

QRA, however, has a very definite existence beyond individual instances. There are frequent arguments about the benefits and drawbacks of QRA as a practice. QRA may be mandated or not mandated by standards and regulations. Empirically, there may be patterns and trends in the properties of QRA examples. There are statements which, if not necessarily true for "QRA" can be shown to be true for most or all instances of QRA.

In these cases, the main question of concern from a big-picture perspective is "If people set out to do a QRA, are they likely to do it well in terms of the claims they will want to make?" If a claim cannot be shown to be true for QRA generally, or to be predominantly true in an identifiable set of QRA examples, then it is an unsupported claim.

QRA practices vary according to the competency of QRA practitioners and between industry sectors. At present there is no research which examines properties common to QRA instances in specific sectors. One of the goals of this paper is to facilitate such research by allowing QRA instances to be measured against a common framework.

2.2. There are strong claims made about QRA

There is no academic or industrial source which spells out all the properties that QRA as a tool must have in order to be fit for purpose. Guidance is available on the form and content of that individual QRA examples should follow, and on errors to be avoided, but this falls short of stating clearly what QRA should achieve. However, we can infer the required properties by looking at how QRA is used and what people say about QRA, and then capture these properties in the form of the claims that users make about QRA. Sometimes these claims are explicit, but often they are implicit. For example, if a risk assessment report states that a QRA was presented to a public focus group and that the group found the risk distribution acceptable, then it is making the implicit claim that the focus group could understand the QRA well enough to make that judgement. Once we have a clearly-defined set of claims, we can examine whether they are supported or refuted by the available evidence, and thereby assess whether the use of QRA is actually justified.

To identify suitable claims we surveyed a collection of real-world risk assessment reports [16] and noted the self-identified purpose of the reports. Activities making use of QRA include

1. Classifying risk (usually for the purpose of regulating a substance or technology).
2. Reacting to public concern regarding a known or suggested risk.
3. Identifying ways to improve a design.
4. Selecting between competing designs.
5. Comparing risk with pre-defined targets.
6. Trading-off risk against other concerns.
7. Tracking changes in risk over time.
8. Accepting or declining risk as a public policy decision.

Of course, QRA is rarely used alone – they are typically used in tandem with qualitative analyses, operational experience reports, and expert judgement. The value of a given QRA partly depends on this context; it also depends on its suitability to benefit from the context – for example, if a QRA has weaknesses that other methods can compensate for, that potentially enables that QRA to be used effectively, but it needs to be clear from the QRA artefacts and activities that it does indeed have those weaknesses. If the weaknesses are not clear, the complementary activities may never be performed, or their results never checked for those issues.

For example, in activity 4 ("Selecting between competing designs"), analysts might proceed by building a quantitative model (such as a set of fault trees) for each competing design, then calculating the safety risk posed by each of them. These quantitative models would seldom be the sole basis for the decision. For example, a qualitative common-cause analysis could explore issues of inter-dependence and systematic failure, and design reviews could examine compliance with standards and regulations. These analyses would feed into the decision-making process, where safety would be one factor to be weighted and considered against competing concerns such as cost, time to market, and non-safety performance.

Similarly, in activity 5 ("Comparing risk with pre-defined targets"), analysts would model in an appropriate quantitative