Precision and objectivity in the expression of hazard controls

Phil Cook and Neil Robinson
RGB Assurance Pty Ltd
236 Montague Rd, West End, QLD 4101, Australia

phil.cook@rgbassurance.com.au

neil.robinson@rgbassurance.com.au

Abstract

The process of hazard analysis for a system culminates in the definition of controls to suitably mitigate the hazards identified. These serve as a conduit through which system safety engineers communicate their intent for achieving a safe outcome to system designers, verifiers and validators, operators and maintainers, and others. As such, the manner in which such controls are expressed is important – a simple fact that is often overlooked. In this paper, we argue that hazard controls should be expressed in a way that leads to a shared understanding of how safety is to be achieved. Not only does this require a clear presentation of a control's goal or expected outcome, but also how the control serves to reduce safety risk. We motivate the discussion of these ideas with examples inspired by real world transgressions we have observed. We also discuss strategies for balancing the need for abstraction during early stages of hazard analysis versus the need for precisely and objectively specified hazard controls during later project stages.

Keywords: hazard controls; safety requirements

1 Introduction

System safety literature has no shortage of material addressing hazard identification, analysis, and management (e.g., Bahr (2015), Ericson (2016), Leveson (2012)). However, the elaboration of hazard controls, although generally agreed to be the major output of such activities, has been afforded relatively little attention. This paper attempts to make a step towards remedying this, by considering the issue of how hazard controls ought to be expressed.

We argue that, because proper implementation of a complete set of hazard controls involves a diverse range of stakeholders, those controls should be expressed in a manner that promotes a shared understanding of how safety is to be achieved. This requires considerable care in the language used to express controls so that they result in a single reasonable interpretation. We present some general principles that are applicable to meeting this goal. We also discuss the role of precision throughout the safety lifecycle and present a simple method for successively refining the expression of hazard controls throughout the hazard analysis process.

This paper is organised as follows. Section 2 provides some background on the role of hazard controls in system safety engineering, building to the key idea of this paper. Section 3 builds on this by presenting some general principles for the expression of hazard controls. We provide a brief discussion on the role of successive refinement of hazard controls in Section 4. Section 5 summarises and concludes the paper.

2 The role of hazard controls in system safety

This section provides some background on hazards and hazard controls. It then presents the key idea of this paper: that hazard controls should be expressed in a way which leads to shared understanding of how safety is to be achieved.

2.1 Preliminaries: hazards and related events

In this paper we adopt a definition of *hazard* given by Leveson (2012): "a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss)". Furthermore, we consider a hazard to be an undesirable event that is observable at the boundary of some system-of-interest.

Each hazard has a set of *causes*: events that, in some logical combination, give rise to it. Similarly, each hazard has a set of *consequences*: undesirable events that may arise from the occurrence of the hazard. Causes are events occurring within the system-of-interest. Consequences are either hazards of some larger (sociotechnical) system in which the system-of-interest is embedded, or an accident.

Both hazards and consequences are, in general, conditioned by further events external to the system-of-interest. We call such events *co-effectors*.

This model of hazards is illustrated in the following figure.

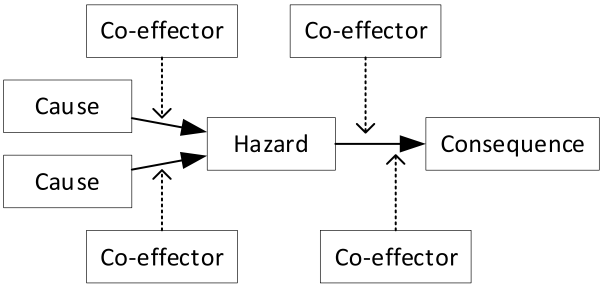

Figure 1: Hazards and related events

Example. Throughout this paper we will draw on a running example of a hazard associated with anti-lock braking systems (ABS) that are fitted to most modern automobiles. This example is adapted from Becker, et al. (2018).

The system boundary of our example is the vehicle to which the ABS is fitted. The goal of an ABS is to automatically regulate the pressure of brake fluid to prevent the vehicle's wheels from locking. The hazard we will focus on is "Unintended Vehicle Lateral Motion". This hazard could lead to consequences such as "Collision with traffic" and would have a variety of causes, including the following.

Ca1: Over-correction of pressure, reducing braking force.

Ca2: Under-correction of pressure, resulting in lock-up.

Either of these events could lead to the hazard. However, we also assume that the ABS provides an indication of its own availability to the vehicle's driver, leading to an additional cause:

Ca3: Failure to indicate unavailability of ABS.

When the ABS fails to properly control brake fluid pressure, the vehicle's driver may have a part to play in avoiding the hazard. Similarly, if the hazard does materialise, the driver has a role to play in avoiding or limiting the potential accident, by avoiding over-steering. These lead to the following two co-effectors.

Co1: Driver braking actions fail to compensate for loss of ABS function, resulting in lock-up.

Co2: Driver over-steers.

These causes and co-effectors, together with other events we have not elaborated on, combine in various ways to result in the hazard and its consequence. We leave exploration of these combinations as an exercise. □

2.2 Hazard controls

A *hazard control* (or, simply, *control*) is an action that is to be (or has been) taken to reduce the safety risk associated with a hazard by eliminating or reducing the likelihood of the hazard's causes, co-effectors, or consequences. Controls are also variously referred to as "mitigations", "countermeasures", or "risk reduction measures".

We classify controls according to following three types.

System safety requirements – these are generally applied to technical systems and serve to define the functional and non-functional properties of such systems. They come in two forms: invariants that must apply at all times (including system states or behaviours that must be avoided) and requirements defining the system's reaction to external events. System safety requirements also include requirements on documentation and associated material that is to be delivered as part of the system.

Constraints on how the system is to be engineered – these include, for example, that certain processes or standards should be applied in the development of the system.

Conditions on how the system is to be operated and maintained – these serve to codify assumptions about the behaviour of system users that are important to achieving and maintaining safety.

Example. For the ABS hazard above, we would expect several controls to be established. We will defer elaboration of these controls to later in the paper, but in broad terms, the controls would include:

- a set of system safety requirements that seek to directly limit the occurrence of causes Ca1, Ca2, and Ca3;
- additional system safety requirements that describe the documentation to be provided with vehicle, especially as it pertains to warning the vehicle's driver of the limitations of ABS;
- constraints dictating the use of established standards for the design and implementation of systems like ABS; and
- conditions relating to the assumed behaviour of the vehicle's driver (which seek to address Co1 and Co2). □

2.3 Controls should establish shared understanding

Elaboration of controls is often viewed as the culmination of the hazard analysis process. But in the larger story of implementing a safe system, it is a step nearer the beginning than the end. Hazard controls are only effective at reducing risk insofar as they can be implemented. Since they play a crucial role in the achievement of system safety, the bar for claiming "proper implementation" of them is necessarily high, and involves a considerable variety of audiences, including:

- system safety engineers;
- system designers and implementors;
- verifiers and validators;
- documentation authors;
- operators and maintainers;
- system owners; and
- assessors and regulators.

These groups will each be interested in some set of controls, each of which overlap with each other in general. Controls are an imperfect medium through which their authors (system safety engineers) express their intent for achieving a safe outcome to each of these groups. Furthermore, these groups are all comprised of fallible humans, who inevitably interpret controls through the filters of their own biases and assumptions.

This leads us to the central thesis of this paper: for a hazard control to be adequate, not only must it be something that is effective at reducing risk, *but it must do so under all reasonable interpretations of those required to act under it*. That is to say, a "good" control is one which, among other things, establishes a *shared understanding* of how safety is to be achieved. Not only does this require a clear presentation of a control's goal or expected outcome, but also how the control serves to reduce safety risk.

3 Principles for the expression of hazard controls

This section builds on the key idea presented in the preceding paragraph by presenting general principles for expressing controls in a manner that helps to build a shared understanding of how safety is to be achieved.

3.1 Relation to requirements engineering

The problem of writing statements in a manner that avoids misinterpretation is one that is faced daily by requirements engineers (both for safety critical systems and at large). The practice of this art has led to various lists of "quality attributes" of "good requirements" that have become conventional wisdom. Authors differ on the exact set of attributes to apply, but they generally share a common core. For example, Young (2004) states that each requirement should be:

- necessary;
- feasible;
- correct;

- concise;
- unambiguous;
- complete;
- consistent;
- verifiable;
- traceable;
- allocated;
- design independent;
- nonredundant;
- written using the “standard construct” (i.e., stated as an imperative using “shall”);
- assigned a unique identifier; and
- devoid of escape clauses.

All of these are desirable attributes that we believe apply equally to the expression of hazard controls.

In the case of system safety requirements, this would seem to be an uncontroversial position. In our experience, however, we have found that all too often, the collective wisdom of requirements engineering good practice is not brought to bear on the articulation of system safety requirements. We have even observed cases where an organisation has applied such principles diligently to the definition of system and subsystem requirements, but left the corresponding system safety requirements to languish in a puddle of ambiguity and subjectivity. This may be attributable to differences in skill sets among the individuals responsible for the respective sets of requirements, but this explanation provides no excuse.

If hazard controls are able to persist in this kind of state, this is an indication that the system safety engineering function, and its results, are not being taken seriously by the wider project, and that the rest of the project is simply “getting on with the job” in the absence of meaningful safety engineering. This has obvious risks. If system safety engineering is not performed and documented properly, then safety requirements could be missed, and an unsafe system might be produced. On the other hand, if the safety engineering work does not stand up to scrutiny, then this could prevent the system from being accepted into service.

While on this topic, we also wish to address a common point of debate: should hazard controls of the “system safety requirement” variety be viewed as actual requirements or are only as “sources” of requirements? We find the latter view to be revealing of an attitude that the responsibility to express a control in a manner that is concise, unambiguous, verifiable, etc. rests with “someone else”. We reject this view for the reasons explained in Section 2.3: a control can only be effective if it admits a single reasonable interpretation that is shared by all. We acknowledge that there are also varying viewpoints on where system safety requirements should be captured (e.g., in a dedicated specification or intermingled with system requirements) and offer no particular opinion on this matter.

While we generally think that the attributes listed above are also good properties to demand of constraints and conditions, it is sometimes acceptable to relax these attributes somewhat in those cases. This is because such controls are typically subject to fewer layers of articulation and requirements management compared to that applied to a system safety requirement.

Example. In our ABS example, we might expect to encounter the following two, closely related, controls. The first is a system safety requirement that acts as a mitigation against Ca1, while the second is a condition that acts against Co1.

Rqt1: The System shall present an alert to the driver when it detects that the brake fluid pressure is high.

Cnd1: When presented with an alert indication, the driver shall compensate by pumping the brakes.[1]

The system safety requirement proposed here is obviously both ambiguous and unverifiable because it does not define what it means for the pressure to be “high”. Nor does it include any constraint on the timing of the alert. Similarly, we might criticise the condition as being incomplete because it does not state for how long the driver should carry out this compensating action, at what cadence it should be performed etc. Ultimately, however, as this condition is one which is to be implemented by a human being (who can be expected to apply their own judgement, for better or worse), it may be acceptable to cast the condition in this incomplete manner. □

In addition to the application of requirements quality attributes to the expression of hazard controls, there is another key idea from the world of requirements engineering that we think control authors should abide by: the principal of a system boundary. Just like in the definition of a hazard that we offered in Section 2.1, controls should be expressed as statements about phenomena or properties observable at the boundary of the system of interest. An obvious prerequisite to doing this (and to effectively analysing hazards in the first place) is to clearly define what the system of interest is (and is not), and those things with which it interfaces, influences, and is influenced by. This is another “obvious” dictum which we have observed to be ignored by practitioners.

3.2 Avoiding dangerously different interpretations

In this section, we will present several examples of poorly expressed controls. These are cast in terms of our running ABS example but are each inspired by real world transgressions we have observed. In each case, we try to highlight the potential for differences in interpretation among different stakeholders.

Example. Proper maintenance of the sensors and actuators employed by ABS is likely to be essential to ensuring the continued safety of the system. We would therefore expect to find a condition (or set of conditions) in relation to this. Suppose we had the following condition.

Cnd2: The vehicle maintainer shall follow the maintenance process manuals.

[1] The effectiveness of conditions placed on the behaviour of users of consumer products, such as an automobile, where there is little in the way of on-going training, competency assessment, etc. is dubious. However, that issue is not central to the topic of this paper.

What does the author of this control think this control is saying? It may be referring to a specific set of manuals that already exist, or to manuals yet to be written, or a combination of the two.

What might a maintainer think this control is saying? He or she may be thinking of a different set of manuals entirely; e.g., some generic manuals that apply to make of vehicle.

Clearly, the possibility exists that these two people might be thinking of entirely different maintenance process manuals. The author may have had a specific set of manuals in mind, in which case they should have been specified in the control. Alternatively, they may be manuals that are yet to be written. In this case, there should be corresponding system safety requirements to prepare those manuals, and again a more specific reference included in this control. □

Example. Following on from the previous example, suppose that a maintenance manual specific to the model of vehicle is to be developed. There should be a system safety requirement calling for this manual to be produced as part of the overall vehicle system. Suppose we had the following requirement.

> **Rqt2:** The system shall include a maintenance process manual.

What does the author of this control think this control is saying? Obviously, the author thinks that it reduces risk somehow.

What does the documenter think this this control is saying? It could be almost anything! Although absurd, the documenter could deliver a copy of the King James Bible with “Maintenance Process Manual” written on the cover, and that would be a valid implementation of this requirement.

Clearly, this requirement should be elaborated with significantly more detail before it can act as an effective hazard control. (This might entail it being unfolded into a suite of related requirements.) As we explain in Section 3.3, below, the main thing that needs to be added to this requirement is a sense of precisely what qualities the maintenance process manual should have in order to reduce risk. □

Example. In the absence of complete ABS function, “over-braking” can lead to lock-up (this is a form of co-effector Co1). We might therefore expect to find a condition like the following.

> **Cnd3:** The driver shall depress the brake pedal as far as it is safe to do so.

What does the author of this control think this control is saying? He or she probably intended to say “shall only…” in this instance. Beyond that, however, it is not clear what the author means by “safe to do so”, nor even if the author had a specific notion of this in mind.

What does the driver think this control is saying? He or she may have a completely different understanding of what “safe to do so” means.

This condition is obviously ambiguous due to the inherent subjectivity of its wording. Prima facie, it appears that the author of the control is deferring to the mythical “someone else” to determine what is and is not safe in this instance. □

Example. A modern ABS implementation draws on a variety of sensor measurements, which it needs to process faithfully. A system safety requirement like the following might therefore be written.

> **Rqt3:** The system shall correctly process measurements of wheel speed, brake fluid pressure, steering wheel angle, etc.

What does the author of this control think this control is saying? Is there a specific list of measurements that need to be processed correctly? Or was this simply a case of “fuzzy thinking”? Does the author have a specific view of what “correctly” means in this context?

What does the designer think this control is saying? Perhaps he or she will simply ignore the “etc”. Possibly, the designer will ignore the requirement altogether, since there would likely be other, more precise requirements that cover the same functions.

What do verifiers and validators think this control is saying? They may have yet another interpretation of the measurements to be considered. They will likely be confounded, however, by the term “correctly” here.

Although the ambiguity of this requirement is apparent and clearly problematic, a deeper issue is that this requirement is *incomplete*. The reason that getting a sense of what “correctly” means in this context is rooted in the fact that the requirement fails to specify something that is observable at the system boundary. Rather than specifying something unobservable about the system’s internal data processing, the requirement should be recast into a set of requirements that relate the sensor measurements to actions taken by system in response. □

3.3 Relating hazard controls to risk reduction

The overall goal of expressing hazards and their controls is to build a shared understanding of system safety risks and how they are to be mitigated.

When a system safety engineer specifies a control, he or she generally has some notion of risk reduction in mind. This may be either qualitative or quantitative, but in either case, a control carries with it some sense of “strength”. This is related to the hierarchy of controls: some controls (typically system safety requirements) act as the primary line of defence by eliminating or reducing a hazard, while others, such as administrative controls, provide “defence in depth” (e.g., controls that seek to reduce the demand rate on a safety function).

This sense of risk reduction should be communicated with the set of controls. It provides insight into the degree of reliance that must be placed on a given control.

Sometimes this can be communicated as part of the control itself. For example, in our proposed improvements to Rqt2, above, we emphasised that the specific risk reduction measures that come about from maintenance should be specified. This would give documenters the information they need to effectively fulfil the requirement, but it would also enable them to convey this important information to the eventual reader of the process manual.

In most cases, however, the way in which a given control serves to reduce risk is better articulated in rationale or commentary that accompanies the controls. This is another example of a practice that is common in the requirements engineering world, but for some reason is often absent from the specification of hazard controls.

The overall relationship between hazards and controls is also important to express clearly. In the model of hazards and related events that we described in Section 2.1, we were careful to allow for a general “logical structure” of causes and co-effectors (in essence, something that could be recorded in a Fault Tree). Establishing such a structure, and carefully relating controls to specific events within that structure, provides a clear exposition of the ultimate relationship between the controls and risk reduction.

A well-structured hazard log is the place where all this information should be recorded and maintained. With that in mind, however, it is not reasonable to expect everyone who is presented with a hazard control to reach for the hazard log every time they need to understand the context of that control. Controls need to stand on their own to some extent, and carefully curated rationale or commentary is an essential ingredient of this.

4 Precision considered harmful

The principles we outlined in the preceding section relate mainly to the “end state” of hazard controls (not necessarily the end of the project, but rather the point at which control implementation begins). However, they are not necessarily appropriate to apply during the early stages of the hazard analysis process.

The process of getting to this “end state” can often be messy. There is an inherent “cone of uncertainty” at play in relation to hazard identification, analysis, and mitigation: comparatively little is known about these at the beginning of the project, but this uncertainty narrows as time marches on.

Consider, for example, a Preliminary Hazard Analysis (PHA). When a PHA is conducted, there is usually little known about the design of the system of interest. Sometimes, even the functionality of the system is under-specified at this stage. Imposing conditions on operation and, especially, maintenance at this stage is unlikely to be done well (the components that will exist in the final design will probably not even be known, much less how they ought to be maintained). Overall, imposing detailed hazard controls at this stage will be hard, and likely counter-productive. This can actually have a deleterious effect, as later hazard control efforts can get “locked in” to what was thought about the system in this early stage.

In general, we believe that hazard controls are appropriate (indeed, essential) within a hazard log, but not so within a hazard analysis.

In some cases, it is possible to conduct a hazard analysis without addressing controls at all. However, when attempting to ascertain an overall degree of safety risk for a hazard or set of hazards, it will generally be necessary to articulate some measures by which hazards are, or are to be, controlled. In such circumstances, we advocate for expressing higher-level control concepts, which we refer to as *control strategies*.

A control strategy seeks to outline how a particular hazard, cause, co-effector, or related group of events, is to be controlled. It should describe a combination of *goals* to be achieved by the eventual controls, the kinds of controls these will generate (e.g., system safety requirements, constraints, conditions), and how they will combine to (satisfactorily) reduce risk.

Example. A control strategy for the “Unintended Vehicle Lateral Motion” hazard might look like the following.

> The following types of controls will be developed for this hazard.
>
> - System safety requirements that limit the combined frequency of events Ca1 and Ca2 to no more than 10^{-3} /hr[2].
> - System safety requirements to present alerts to the driver whenever either Ca1 or Ca2 is detected to occur.
> - Conditions on driver behaviour that specify that the driver:
>   - reacts to ABS unavailability by pumping the brakes; and
>   - does not depress the brake pedal so far as to cause over-pressure when ABS is unavailable.
> - etc. □

Such control strategies, once established, can be successively refined throughout the hazard analysis process, eventually resulting in concrete hazard controls that are both precise and objective. A well-structured and well-maintained hazard log enables such refinement to be tracked and managed.

5 Conclusions

This paper has considered the issue of how hazard controls should be expressed. We have argued that the true purpose of controls is to establish a shared understanding of how safety is to be achieved. From this, it follows that precision and objectivity are essential in the expression of controls, for without these properties different readers will be apt to arrive at (dangerously) different interpretations. We have provided some general principles for the expression of controls that help to support this goal and discussed the role of successive refinement of controls throughout the hazard analysis process.

6 References

Bahr, N. J. (2015): *System safety and risk assessment: a practical approach*. 2nd ed. Boca Raton, FL, USA, CRC Press.

Becker, C., Arthur, D., & Brewer, J. (2018, August): *Functional safety assessment of a generic, conventional, hydraulic braking system with antilock brakes, traction control, and electronic stability control* (Report No. DOT HS 812 574). Washington, DC, National Highway Traffic Safety Administration.

Ericson, C. A. (2016): *Hazard analysis techniques for system safety*. 2nd ed. Hoboken, NJ, USA, Wiley.

Leveson, N. (2012): *Engineering a safer world: systems thinking applied to safety.* Cambridge, MA, USA, MIT Press.

Young, R. R. (2004): *The requirements engineering handbook.* Norwood, MA, USA, Artech House.

[2] The rate specified here is illustrative only.