Relating Risk and Confidence: A Structured Approach to Constructing Assurance Cases

Tim Kelly
University of York, UK

E-mail: tim.kelly@cs.york.ac.uk

Purpose of a Safety Assurance Case

Principal Objective:
- Safety case presents the argument that a system will be acceptably safe in a given context

Safety Cases must clearly define / explain:

- ‘System’
- ‘Acceptably’
- ‘Context’
Argument & Evidence

A safety case requires two elements:

- **Supporting Evidence**
  Results of observing, analysing, testing, simulating and estimating the properties of a system that provide the fundamental information from which safety can be inferred

- **High Level Argument**
  Explanation of how the available evidence can be reasonably interpreted as indicating acceptable safety – usually by demonstrating compliance with requirements, sufficient mitigation / avoidance of hazards etc

Argument without Evidence is unfounded
Evidence without Argument is unexplained

Presenting Clear Arguments

- Basic argument structure
  - **claim** – what we want to show
  - **argument** – why we believe the claim is met, based on
  - **evidence** – test results, analysis results, etc.

- In general, argument broken down hierarchically
  - claim, argument, sub-claims, sub-arguments, evidence

- Important to ‘carry’ the reader with you through the argument
  - Not ‘lose them in the chase’!
  - Force to make too big a leap

- Easier to show graphically?
The Goal Structuring Notation

Purpose of a Goal Structure

To show how goals are broken down into sub-goals, and eventually supported by evidence (solutions) whilst making clear the strategies adopted, the rationale for the approach (assumptions, justifications) and the context in which goals are stated.

A Simple Goal Structure
A Simple Goal Structure

Safety Requirements & Objectives

Safety Argument

Safety Evidence
Step 1 – Example

G1
Press is acceptably safe to safe to operate within Whatford Plant

As with conventional safety case report – we wish to clearly set out the objective and scope of the safety argument being presented.
Step 2 – Example

- Terms – Press, Operate and CCC Whatford Plant drawn out explicitly as contextual information
- Acceptably Safe left for expansion through the supporting argument

Step 3 – Example

- Two separate strategies – for reader’s benefit
Step 4 – Example

- Needs Justification?
- Any Assumptions?

Step 5 – Example

- Needs Justification?
- Any Assumptions?
**Step 6 – Example**

- **G3**: Hazard of ‘Operator Hands Caught in Press Drive Machinery’ sufficiently mitigated
- **G8**: Motor / Clutch / Drive Belts surrounded with safety cage
- **G9**: Press operation will (safely) halt if safety cage tampered with

**Reference to source of information that would substantiate claim**

More explanation required here

---

**Safety vs. Confidence**

Need to distinguish arguments about:

1. **A)** why a (real, product, people-get-hurt) hazard risk is acceptably managed, e.g.
   - Why is this hazard sufficiently unlikely to occur?
   - How is this hazard mitigated if it did occur?

2. **B)** why there is sufficient confidence in the arguments and evidence of risk management, e.g.
   - Is that testing exhaustive?
   - Is that COTS experience representative of my usage context?
Time to Join AA

AA - Arguers’ Anonymous

First step in recovery is to admit:

“My arguments and evidence are not perfect, they’re not proof, there are gaps and flaws in them”

Possible addition: “… but I think they’re OK”

Otherwise, there’s no issue of confidence to be argued

Two areas of Concern:

The “Logic” isn’t infallible, e.g.:

- Conclusion “The software will enforce Safety Property X”
- (Sole) Supporting Claim: “The software has been developed to SIL 4”

or

- Conclusion “The software is safe within a system context”
- (Sole) Supporting Claim: “The software satisfies its functional safety requirements”

The Evidence isn’t infallible, e.g.:

- Conclusion “The software will enforce Safety Property X”
- Evidence: (Non-exhaustive) testing
Mind the Gap

- Numerous factors may affect confidence:
  - Assumptions made & scope drawn
  - The “inductive gap”
  - Trustworthiness of evidence
  - Visibility of information
  - Etc.

- Having such uncertainty is normal and acceptable
- As long as it is identified, understood and managed

Current Practice

- Many safety cases don’t recognise the distinction between safety arguments and confidence arguments

  - All mixed together
    - Arguments of confidence alongside safety
    - E.g. “COTS component is acceptably safe” because “COTS component doesn’t exhibit failure mode Z” and “COTS component supplied by a trustworthy vendor”
  
- Transition from safety argument to confidence argument
  - Arguments of safety turn into arguments of confidence
  - E.g. “Software System will satisfy safety property X” because “Software developed to Development Assurance Level A”
Consequences of Mixing

- Arguments tend to become large and unwieldy
  - Too much information in one argument
  - Unnecessary material is sometimes included in arguments “just in case”
  - Voluminous, rambling, ad infinitum arguments
  - Arguments become difficult to review

- Weaknesses of argument are sometimes not evident
  - Easily overlooked
  - More difficult to spot incompleteness or poor structure in either

- Link between elements of the argument and risk is often lost
Haddon-Cave

- Difficulties are serious since they detract from the basic purpose of using safety cases
- Many of these problems with current practice in safety cases were highlighted by Haddon-Cave

- Bureaucratic length
- Failure to see the wood for the trees
- Disproportionality
- Compliance-only exercises
- Audits of process only
- Safety Cases were intended to be an aid to thinking about risk but have become an end in themselves

Clear Separation Required

- The safety argument documents the asserted arguments and evidence of risk reduction
  - RULES:
    - Everything cited in the safety argument should have a direct role as part of the causal chain to the hazard;
    - All claims in the safety argument must be claims about the system or parts, properties, or properties of parts thereof
    - Artefacts from system development (e.g. test reports and, by extension, their contents) may be referenced only as evidence or context
- The confidence argument documents the reasons for having confidence in the safety argument
  - RULES:
    - confidence argument claims must address (only) the structure of the safety argument (i.e. it’s not a free-for-all!)
Safety Arguments as Assertions

- (For non-deductive arguments) the recorded argument that ‘Hazard X is acceptably mitigated’ *because* ‘Safety Measure Z is sufficiently reliable’ is an ASSERTION
  - It’s ‘Say so’

- (For non-deductive arguments) the recorded argument that ‘Safety Measure Z is sufficiently reliable’ is evidenced by ‘Fault Tree Analysis Results’ is an ASSERTION

- The argument that declares ‘Hazard List’ is the relevant and appropriate context for the Risk Argument is an ASSERTION

- Safety Case Arguments are bags of ASSERTIONS
Assurance Claim Points

- These assertions could, and should, be debated
  - This is the role of the CONFIDENCE argument

- These ACPs correspond to three different types of assertion:
  - Asserted inference (ACP1)
  - Asserted context (ACP2)
  - Asserted solution (ACP3)

Confidence Argument

- Qualitative argument to demonstrate sufficient confidence in an assertion:
  - Residual uncertainties (assurance deficits) in the assertion have been identified
  - Residual uncertainties (assurance deficits) in the assertion are insufficient to cause concern

- Quantify?
  - If you can, do
  - However, confidence 'problems' with the safety argument will almost always relate to a qualitative omission of something
  - There is no science to the encoding of the impact of that omission in terms of a confidence value (where no relevant prior evidence exist it is merely an encoding of beliefs)
  - Encoding and quantification of beliefs doesn’t really help identify the real issue to be addressed (worse: it can obscure it)
Assurance Deficits

- Recognised assurance deficits = Something we don’t know (haven’t addressed in the case)
  - A known unknown
  - Potential source of counter evidence
- Increase assurance by addressing deficits

How much confidence is enough?

- Are the identified assurance deficits acceptable?
- Necessary to reason about the ‘consequences’ of deficit
  - … on the safety argument claims
- Which aspects of the claims (in the safety argument) are still assured, and which are not?
  - What are the worst implications of ‘not knowing’?
- Worst case: uncertainty, when resolved, undermines (is counter-evidence for) your case
  - When you check your blind spot, there’s a motorcycle...
  - Considering the potential counter-evidence can help determine impact
Mitigating ADs

Example mitigations:
- Change the design of the system
  - e.g. adding a hardware backup when it is impractical to demonstrate with adequate confidence that software has the properties necessary to ensure system safety
- Change the system operation
  - e.g. by limiting the conditions under which the system is used
- Change the safety argument
  - e.g. adding an independent source of evidence
- Generate additional evidence for the confidence argument,
  - e.g. gather additional evidence about the effectiveness of previous similar safety arguments and evidence

How much confidence is enough?
- Are we moved to act?
  - When have we done enough mitigation?
  - Need some stopping criteria
- There will always be some residual ADs
  - Diminishing returns
  - Inevitably we consider Costs vs. Benefits
  - The effort should reflect the risk
    - That's why understanding the effect of AD on the safety argument is so important
Confidence Argument Structure

What should the confidence argument contain?

ACP1
Sufficient confidence exists at an assumed inference at ACP

CC1
Credit support exists for the truth of an assumed inference at ACP

CC2
Assurance deficits at ACP have been identified

CC3
Residual assurance deficits in (assumed inference at ACP) are acceptable

What grounds are there for believing this assertion? What are the assurance deficits associated with this assertion? Why are the residual assurance deficits believed to be acceptable?

It is not expected that any significant counter evidence exists in the assurance deficit ‘gap’

The residual assurance deficits are acceptable because...

The effect of the assurance deficit on the safety argument does not warrant further mitigation

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 33

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 34
Confidence Argument Structure

- Similar arguments for asserted solutions and asserted context too
- But two aspects of confidence to consider
  - Trustworthiness
    - Concerns the integrity of the evidence (or context)
    - Is the evidence what it purports to be?
    - Relates to confidence in the evidence descriptive assertion
  - Appropriateness
    - Concerns whether evidence (or context) is appropriate for its role in the argument
    - Relates to confidence in the evidence results assertion
**Overall Confidence Argument**

- Can assemble individual fragments of confidence argument to form an overall confidence argument.
- Number of important concerns for overall confidence argument:
  - Sufficiency may be more complex simple composition
     - Shortfalls in one part of the argument may be compensated by other parts
  - May be *common* underlying assurance deficits
     - Common modes of failure
  - May not be practical to argue confidence of *every* assertion
     - Selection and prioritisation of argument assertions required
Example: Insulin Pump Safety Argument 1

Insulin Pump Safety Argument 2
Insulin Pump Confidence Arguments

Pump Design (ACP.A1) Confidence Argument

- Important because intent defines ...
  - scope of concern
  - The ‘view’ of pump to be adopted within argument
- Assurance Deficits for Appropriateness (Right Thing?)
  - Is the pump design an adequate reflection of pump as built?
  - Is the pump design an adequate reflection of pump over the lifetime of each unit?
  - Does the pump design link to user operating instructions?
- Assurance Deficits for Trustworthiness (Thing Right?)
  - Is the pump design document complete?
  - Is the pump design document free of ambiguity?
  - Is the pump design document internally consistent?
Pump Design (ACP.A1) Confidence Argument

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 43

Pump Design (ACP.A1) Confidence Argument

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 44
Pump Design (ACP.A1) Confidence Argument

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 45

Pump Design (ACP.A1) Confidence Argument

Relating Risk and Confidence: A Structured Approach to Assurance Cases - 46
Selected confidence argument considerations

- **ACP.A5** – There is sufficient confidence that the list of delivery modes is complete and correct
- **ACP.S2** - There is sufficient confidence that considering the risk of excess insulin during each delivery mode will demonstrate that the risk of hypoglycaemia is adequately mitigated
- **ACP.A6** - There is sufficient confidence that the definition of commanded infusions is appropriate (Some modern insulin infusion pumps use a Bluetooth network connection to communicate)
**Assurance Arguments**

- Only talked here about **safety** cases
- Concepts apply immediately to **any** property of interest
  - Assured security cases
  - Assured reliability cases
- Content of assured security or reliability case would differ from assured safety case
  - Overall structures and approaches would be identical

**A Third Perspective - Compliance**

- For many safety critical industries, there are existing regulatory objectives, legislation etc:
- Natural to expect the safety case to address these

[Diagram showing the relationship between Assurance Argument, Safety Argument, Compliance Argument, and Confidence Argument.]
Compliance Arguments

- Compliance arguments can concern / be related to specific features of risk mitigation argument, e.g.:
  - Risk Target
  - Expected Risk Mitigation Features
- Compliance Arguments can also concern matters of confidence, e.g.
  - Suggesting the use of certain techniques or processes according to risk category, integrity level, or assurance level
- Examples:
  - IEC61508 Part 2 makes a clear distinction between between measure to avoid introducing systematic error (confidence) and measures to control any residual systematic errors (safety)
  - DO-178B talks in terms of assurance levels (confidence) in the justification of the satisfaction of software requirements (safety)

Twin Peaks

- Arguing compliance is not the same as arguing safety (or even sufficient confidence of safety)
- However, there can be plenty of commonality
Summary

- Existing safety arguments can often be ‘flabby’
  - Everything including the kitchen sink thrown in

- Often poorly argued
  - "Why is this relevant to that?"
  - Use of a structured approach (e.g. Claims-Argument-Evidence or GSN) is no guarantee in itself

- Need to acknowledge the weakness of safety arguments
  - They’re not proof

- Discipline of separating safety from confidence important
  - There are simple rules for what is permissible in each argument
  - Issues of confidence are otherwise often poorly handled
  - Provides opportunity to simplify the safety risk arguments

- Compliance is a necessary third perspective
  - Again, can help recognise that ‘top claim’ is distinct from a ‘pure’ safety claim