Seminar: What Makes Good Evidence for a Safety Case



The "What Makes Good Evidence for a Safety Case" seminar was held on 30th Nov 2023 at the Wellcome Collection in London. This seminar examined the role of evidence in supporting a safety case and explored the differences between compelling and weak evidence. It also considered what could be done to fill the gaps if evidence is not available or of poor quality or if evidence is continually evolving. The SCSC Newsletter Editor provides a summary of the day.

Mike Parsons opened the seminar and chaired proceedings, introducing each speaker and facilitating questions from the audience after each talk. The talks covered a wide range of domains including Aviation, Automotive and Healthcare. The event concluded with a lively open group discussion on a range of topics.

Evidence for a Satellite Datacomms to Civil Aircraft System



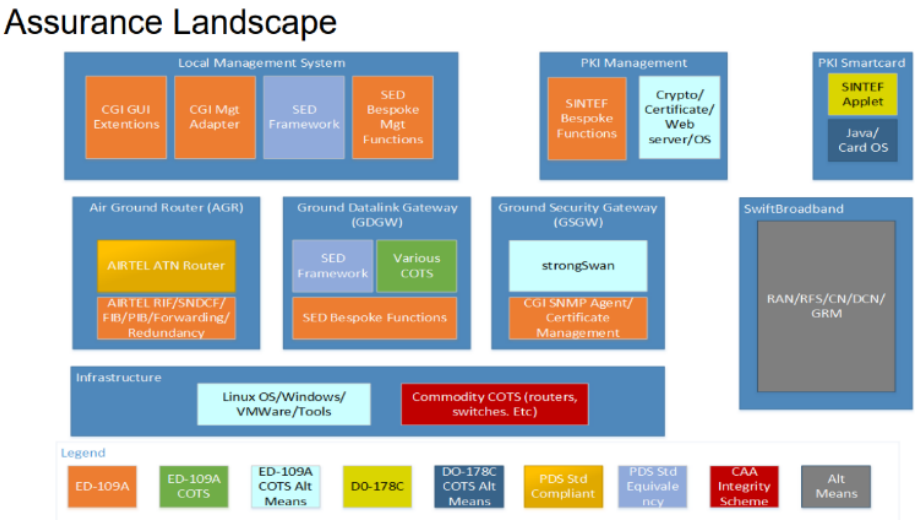
Paul Hampton from CGI gave the first talk of the day based around a system providing a data link for air traffic controller communications to commercial airline pilots. Paul discussed the risks and challenges of standards-based assurance when the system comprises subsystems of very different natures such as new, legacy, predeveloped software and software where alternative means of compliance arguments are required.

Paul first introduced the Iris system, which he has been working on for almost a decade and is now being deployed on live commercial flights. The system provides a SATCOM-based data link to supplement traditional VHF voice controller/pilot communications, which are seeing channels become increasingly congested. The use of data-based communication links also paves the way to more efficient route planning that will drive fuel saving costs and lead to reduced CO_2 emissions.

Paul provided the history of the project noting that when the project began, data link standards were not mature and did not cater well for SATCOM as a bearer for the data. The focus was therefore on the four assurance "pillars" of: safety, security, performance and interoperability and Paul showed how a structured approach to understanding the lifecycle of each of these assessment processes, and how they interacted, was key to developing an overall assurance argument for the entire system. Paul noted that the result was a very large and complex assurance argument and tools were required to help make the assurance case more accessible.

Paul then went on to explain how the regulation finally matured into requiring a 'Safety Support File' from the communication service provider that required evidence to demonstrate that: "the system would operate only as specified and would continue to do so in live service". Paul noted that producing such a case was akin to a traditional 'V' model and was well aligned with the earlier work that focused on explicitly breaking out each distinct phase of the development lifecycle from design through into service. Much of the evidence generated could then be allocated to different parts of the 'V' model to help support the overall assurance case.

Paul then went on to describe the many different assurance approaches adopted on the project. Although nominally, the system was to be developed to aviation standards, only the new software developments could be assured in this traditional way, as much of the system was legacy, COTS or based on previously developed software. Paul noted that nine different methods were used in the end to provide an equivalency to the level of assurance required by the standards.



The "assurance landscape" showing all the system components and which assurance approach (each with a different colour) was adopted for that component.

Paul concluded his talk by highlighting his thoughts on what makes good evidence:

- "Goodness" is not a property of evidence itself … but more on the narrative that accompanies that evidence
- Favour evidence of "least surprise" – evidence following a standard's objectives or known techniques will be more accessible to regulator/reviewer
- Adopt a structured approach to organising and presenting the evidence
  - Be clear on methodologies adopted and their interactions
  - Build evidence incrementally on smaller, manageable lower-level arguments
- "Less" can often be "more" – avoid temptation to focus on quantity, it can just overwhelm
- Be open about assurance deficits, but then argue why "that's ok" or provide plans to address

Addressing the Challenges of Evidence in Healthcare IT



In the next talk, Sean White from NHS England set the scene by saying healthcare was different from traditional safety as it considers the safety of poorly people rather than focussing on the existing safe state of, say, an aircraft.

He then discussed the Medical Device Regulation (MDR), which applies to products such as smart watches. The MDR uses a classification scheme that involves increasing rigour including independent

review as risk increases. He noted however, that MDR is for products and doesn't cover traditional clinical services Health IT.

Health IT systems are now covered by their own Data Coordination Board (DCB) standards that are now mandated: DCB0129 for Health IT manufacturers and DCB0160 for the deploying health organisation. These are based on ISO/IEC 14971 but include requirement for a Clinical Safety Officer (CSO) and a Safety Case Report.

Sean covered some of the challenges and which forms of evidence might help address those challenges.

Subjective Assessment of Risk

There is variance in risk schemes experience and corporate attitude to risk leading to subjective assessment of risk.

We would therefore seek to have more around explanation of the risk scheme used and rationale for the severity and scale of impact.

Qualitative determination of Risk

There is no quantitative assessment of risk for software-based products and processes are not prescribed.

Evidence should therefore be more around the justification: why the software development process will deliver the required integrity and engagement of the CSO in the process.

Reliance of "clinician in the loop"

There is a perception that the clinician can always intervene before things go wrong but we need to be confident that this is indeed the case.

Evidence should focus on engagement of the clinicians, their training and competency and defining pathway as 'work as done' to ensure there are credible opportunities for intervention.

Perceived Safety Significant

Consequences of failures are in general less severe and major "crash" events are rare. This leads to a culture of playing down risk.

More is required on the justification to ensure the clinical context is explained and there is close control on incidents so they can be properly managed.

Complex Architectures

Systems are becoming increasingly complex with cloud, mobile and variance in use cases. Development methodologies like agile are also adding to complexity.

Evidence should therefore focus on understanding system design, justifying 3rd party products and defining the development processes.

Safety Management Competence

Sean noted that there are issues with awareness and commitment to the standards and they are often retrospectively considered.

The focus should therefore be in training; ensuring top management understand their responsibilities and there is sufficient resourcing of, for example, an influential CSO.

Sean described the work being undertaken to de-risk the deficiencies:

- England is the first country to mandate a risk management process
- There is now specific guidance in a healthcare context for managing data safety risk
- There is a national portfolio of training addressing the importance of evidence
- There is a focus on proportional risk management; with limited resource the focus is on greatest risk
- The establishment of a CSO is important in providing domain knowledge, influence and direction, responsibility and accountability.
- Health organisations are encouraged to build DCB into contracts to ensure safety is built into the service

But he ended saying that inevitably it is a "leap of faith".

Sean concluded with the following observations:

- Problematic safety evidence is a symptom of other issues
- There is a need to establish confidence rather than absolute proof
- There is a need for more emphasis on pre-emptive risk management
- There is a need for more openness across organisational boundaries
- AI will compound existing challenges and introduce new ones

Safety Evidence: The Good, The Bad, The Ugly

Helen Wright from Leonardo said that not enough thought is, in general, given to the production of evidence in a safety case and there are issues with confirmation bias as highlighted by Nancy Leveson and the Nimrod Report.



She said we ideally want the case to be comprehensive, compelling and valid but they tend to be disjointed, difficult to navigate and end up trying to find large volume of evidence, a process which can end up pulling on threads that unravel the case. Helen thought that we need to raise the bar on what good safety evidence looks like.

Helen gave an example of safety activities undertaken on a civil aircraft programme that followed SAE ARP 4754/4761 processes in support of aircraft certification; this sits outside of a typical safety case regime. Combined safety/design generated extensive evidence although this was a lengthy process. She gave an example where safety analysis prompted a redesign of the fuel tank boundaries. She noted that the Aircraft Safety Synthesis was the closest thing to a safety case but without the argument. The same approach was implemented on the next programme designed for military operations, however in this example Helen highlighted some of the issues encountered when subsequently trying to adopt the safety case approach (as required by the Military Aviation Authority (MAA)) for release to

service with the UK MoD. The MAA requirement for a safety case and to be able to declare system risk reduced As Low As Reasonably Practicable (ALARP) and tolerable was substantially different to the evidence set required to support certification; this presented a number of challenges:

- The need to prepare a safety argument using available Certification evidence produced through a different process where appropriate and identify the potential evidence shortfalls in order to address them
- The additional need to produce a Hazard Log to support risk-based decisions at an appropriate level and employment of additional tools (such as Bow Tie Analysis) to aid Stakeholder engagement and facilitate communication to meet different regulatory requirements

This example was used to highlight the need to consider the regulatory environment and challenge the assumption that evidence used for one application might not necessarily be suitable in another.

Helen's final example was a safety-critical software project where assumptions about reuse of existing code were made during the bid phase such that only limited updates would be required to the safety analysis, the Safety Case Report (SCR) and the Safety Management Plan. This turned out not to be the case and substantial effort was required to unpick the traceability of requirements both from a safety analysis and a compliance with standards perspective as a result of seemingly insignificant changes in the expanded customer scope/requirements. The issues were largely resolved by identifying safety evidence requirements and detailed safety planning identifying the dependencies into the safety deliverables and feeding back into the system engineering lifecycle. The key learning point was that Safety Evidence Requirements instantiated into the project worked well.

Helen then concluded by discussing how to make things better and suggested:

- Establish confidence in the evidence from the start and identify any potential shortfalls
- Detailed safety planning as early as possible that is integrated with the system lifecycle and identifies dependencies (inputs to the safety analyses) that are generated by the wider team
- Challenge ambiguity in review, especially assumptions in the early stages of a project that influence the safety approach
- Need to get others on board by improving awareness of the evidence that will be required to support the safety argument – could use safety evidence requirements to make it explicit

Helen also suggested some other possible ideas to improve:

- Include confidence criteria in Goal Structuring Notation (GSN) and counter safety-arguments
- Eliminative argumentation and adopt a more mathematical approach
- It may be better to focus more on why it is not safe
- Improving clarity in communication of hazards with Stakeholders to ensure they have suitable and appropriate information on which to make their decisions

Autonomous Vehicles and Safety Case Evidence



Rebecca Wilford and Norina Ratiu from Oxa discussed their work in developing an L4 autonomous vehicle. Norina discussed the safety lifecycle across a 'V' model from project definition through to Safety in-use monitoring.

Autonomous Vehicles (AV) take on some of driving tasks such as strategic (mission specification), manoeuvre (behaviour stay-in lane etc) and control (throttle brake steering). The safety scope of AVs therefore needs to cover behaviour competencies, malfunctions, human misuse and cyber threats.

Norina and Rebecca discussed their Positive risk balance (PRB) approach. MISRA 2019 have developed guidelines for Automotive Safety Arguments. This splits the argument into 4 claims: Means, Rationale, Satisfaction and Organisational Environment.

They then gave examples of supporting claim themes using GSN for each of these four areas. They then discussed evidence quality and gave examples of the different types they encounter:

- **Best**: completeness in documentation and configuration control with traceability and evidence of review
- **Average**: summaries, front pages only, partially controlled, in note form
- **Bad**: Verbally communicated only, second hand sources or implied

Keeping an Operational Safety Case Alive and Well



Andy Whalley from BAE Systems discussed air system safety cases in the Military domain.

He has responsibility for managing the safety cases of several aircraft and likened maintaining a safety case to raising young children where they need nurturing and eat a lot, that is, they require a lot of evidence. He also said it is also not something that is done and completed but needs to be maintained throughout operations.

He noted that end operational cases need to be thought about at the conceptual stage even though not much evidence will be there to start. There can also be two operational cases at one time, one for the airframer and one for the airline (including maintenance etc).

The Military Aviation Authority (MAA), the regulator in the military domain, likened safety to a bath – evidence heats up bath water to operational standard as evidence is added.

Andy covered 3 aspects of safety case design: risk, rules-conformance and confidence and discussed what evidence is important through each of the product lifecycle phase.

Design phase



Risk

- Understand and ensure the context is right; in particular looking for CONOPS, to explain the Who/What/Where/When
- Ensure design has safety targets, eg. quantitative loss targets for the given context

Rules/Conformance

- Be clear on which regulatory set are being worked to (there are about 300 in military aviation and be able to argue which ones are applicable and which have been disregarded)

Confidence

- Need to ensure there are competent contracted organisations
- Staff in the organisation will need to be SQEP and there should be accountable stakeholders
- Check there are clear plans, principles and processes in place

Build/Test Phase



The evidence 'bathtub' will now get slightly warmer as the project moves into the Build/Test Phase and there will now be test plans.

Risk

- System safety assessment can now be conducted to assess the design and determine the residual risk
- Trials risk assessments can also be conducted
- Assumptions/dependencies are important and often a forgotten aspect. Need to look for evidence that assumptions/dependencies have been accepted across stakeholders.

Rules/Conformance

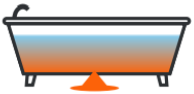
- Understand and handle non-compliances
- Compare the Design versus Build and versus Maintenance Standards to ensure they are consistent

Confidence

- Determine if targets have been met through post-test reports and risk-to-life review (assess residual risk)
- Undertake 1st 2nd 3rd party assurance, employ Independent Safety Assessors (ISA) and conduct regular safety/risk to life review

Andy provided further details on how to make the risk to life review through the building of a dynamic operational safety case based on a pyramid structure diagram. This allowed changes in the environment to be quickly assessed for impact on the safety of life assessment.

Entry-into-service and beyond



Once the system enters service the "bathtub" is now full of warm water, but it needs to be maintained to ensure the water does not get cold or leaks away.

Risk

- Ensure changes in configuration/environment, safety enhancing technologies/techniques, new technologies are managed and reflected in the safety case
- Ensure issues (the 'bath leaks') are managed (eg. operator or maintenance errors, accidents etc.) and review wider fleet occurrences

Rules/Conformance

- Monitor and assess the impact of regulatory change and conduct usage validation reviews

Confidence

- As with the previous phase: undertake 1st 2nd 3rd party assurance, ISA and regular safety/risk to life reviews.

Andy then looked at some of the future challenges:

- If AI has been used to generate evidence, how do we trust it? Similarly, what if it is used in the design, production and even in the safety case generation?
- Digital twins have also become more prominent in recent times. Could these be used for synthetic training of products? Virtual hazard analysis (interrogating a digital model) is already being done.

Andy concluded by noting that:

- Evidence matures/evolves through-life
- Understanding operational context is key for operational safety case
- Evidence can grow stale, which can then lead to an erroneous safety case

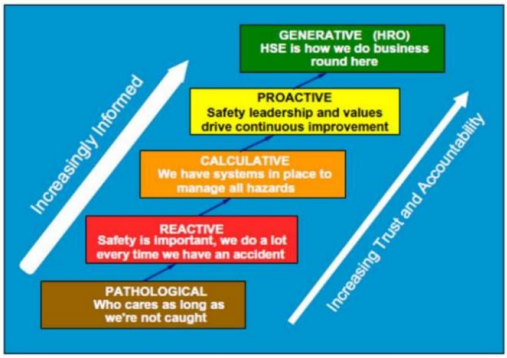
Confessions of an Auditor



Sarah Tovey from the MOD, described her roles in the independent technical services team and aimed to illustrate trends that were not domain specific.

The Journey

Sarah referred to the Parker-Hudson model, which is a model for safety culture in an organisation running from Pathological through to Generative.



Sarah thought the organisation was probably more reactive initially than "generative" and noted there were many issues such as not being consistent across domains, there were known problems and change programme upon change were disruptive.

3-5 years ago, the Safety and Environmental Management System (SEMS) documentation was being updated but not evolving quite as fast as they would have liked. 5 years ago, an in-house assurance team was developed that provided better oversight and robust evidence of trends, however there were still issues and new procurement models also complicated matters.

Where they are now

Sarah said they have now introduced consistent assurance across domains and internal delivery specialists established but there are still issues, known problems continue – and there is still change upon change.

Sarah described the Defence Safety Management System "wheel" from JSP185 that provides an SMS structure that is consistent across domains and then discussed trends and themes across some of these safety management areas.

Trends and themes

- Shelfware, lack of linkage of critical project milestones, quality variations, repetition of policy, resources, continuous improvement opportunities missed.
- Legislative compliance: deficiencies with demonstrating compliance and risk of deficiencies. Training systems and Compliance with government supplied equipment poorly addressed.
- Environmental management missing disposal plans.
- Different tools: risk to life v equipment contribution to risk of life. Issues with putting an air platform on a ship – are both parties evidence there?
- Underpinning evidence – when is enough enough? Some domains better than others.
- Personal competence resources and training: lack of SQEP, ongoing cycle of routine maintenance overlooked. Lack of succession planning. "niche" specialisms.
- Assumptions management; unknown risk as assumption that is unproven or not mitigated.
- Performance management information and reporting: what to report, trend management and learning from experience.
- Assurance: lack of resource and oversight.

Where they want to be

Sarah said that they want to be on the edge of calculative/proactive but noted this will take time but will require:

- Effective in-house and cross-domain assurance
- Embedded processes: internal delivery, change programme, psychological (people empowered to speak up)
- Known problems addressed

Takeaways and challenges

Sarah concluded with some final takeaways:

- It is an ongoing journey; cultural changes take time to embed
- Sharing of knowledge and consistency between projects could be improved; often see different teams making the same mistakes, which could have been prevented if there was better sharing

Group Discussion

The day ended with a group discussion to discuss what could be done to improve best practice. The following ideas were discussed:

- Expiry dates for evidence
- Standardising the weighting of evidence
- Evidence can be a byproduct but cost benefit may be difficult to evaluate
- Can evidence be quantified in isolation without the narrative; perhaps needs to be about relevance?
- High building safety new to systems safety – currently more compliance based – but big changes have occurred with every building over 80m needing a safety case and other important changes around how we engage with the residents.
- Impact of AI how are these going to be handled in the future. All depends on how it is used. May end up just being reviewers of AI outputs rather than authors.
- Should we do a hazard analysis ourselves of where AI can be used safely?
- May need a separate safety checker of AI outputs. Should it be focused on just showing where there are holes in a safety case?
- Safety culture is more important than we think. Culture can destroy the evidence base if not appropriate.
- Who is the readership of the safety case? Agreed that it is best if crafted as a narrative and story to bring the reader along.
- Is good safety evidence something you can trace back and understand its traceability to requirements?
- Should we be able to learn from rare events too?
- How does an author know that evidence is good? Should it be a group effort?
- Who "owns" the safety case? Authors can't be expected to know everything is correct in all aspects. Needs to be a collective – seen as a safety team activity

Mike wrapped up the session and concluded by saying the SCSC will be launching an AI Working Group.