Improving the Risk Matrix

Nancy Leveson
MIT

A Standard Version of the Risk Matrix

- Used throughout the life cycle
- Assumes Risk = f (severity, likelihood)

RISK ASSESSMENT MATRIX				
SEVERITY / PROBABILITY	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
Frequent (A)	High	High	Serious	Medium
Probable (B)	High	High	Serious	Medium
Occasional (C)	High	Serious	Medium	Low
Remote (D)	Serious	Medium	Medium	Low
Improbable (E)	Medium	Medium	Medium	Low
Eliminated (F)	Eliminated			

Severity

- Defined as a set of categories, such as

  Catastrophic: multiple deaths

  Critical: one death or multiple severe injuries

  Marginal: one severe injury or multiple minor injuries

  Negligible: one minor injury

- Relatively straightforward but
  - Worst case? Most likely? Credible? Predefined common events?
  - How define credible? (blurs with likelihood)
  - Design basis? (nuclear energy)

- ARP 4761 example:
  - Loss of deceleration capability
    - Not annunciated during taxi: Major (Crew unable to stop a/c resulting in slow speed contact with terminal, aircraft, or vehicles)
    - Annunciated during taxi: No safety effect (Crew steers a/c clear of any obstacles and calls for a tug or portable stairs)

"Improved" Disembarkation Method


Gate
114
114
114

Likelihood

- Example:

  Frequent: likely to occur frequently

  Probable: Will occur several times in the system's life

  Occasional: Likely to occur sometime in the system's life

  Remote: Unlikely to occur in system's life, but possible

  Improbable: Extremely unlikely to occur

  Impossible: Equal to a probability of zero

- More problematic than severity
  - Historic events may not apply
    - System, environment, or way used may change
    - Software "failure" is always 1
  - Sometimes associate with probability levels (can this be determined?)

How Accurate is the Risk Matrix?

- Almost no scientific evaluation
  - Two studies I know about, both had poor results (orders of magnitude different evaluations by experts)
- Empirical (from practical use)
- General technical limitations
  - Mathematical and theoretical
  - Heuristic Biases

Empirical Evaluations and Practical Limitations

Caveats

- Nothing available so only our own evaluations on real systems
- Not criticizing individual engineers or companies
  - They were following standard practices
  - Our goal was to figure out how to improve what is done today
  - Same flaws in hundreds of these I have seen in my career

Empirical Evaluations (2)

- Common problem: Assess risk of failures not hazards
  - Loss of external communication or breaking piston nuts vs. aircraft instability or violation of min separation from terrain
  - Reliability, not safety
  - What about non-failures?
  - Individual failures but not combinations of low-ranked failures (and usually assumptions that pilot will behave appropriately)
    - Infeasible to consider all combinations
    - Assumption of independence
    - Affects accuracy of results

Empirical Evaluations (3)

- Assumptions about correct pilot reaction to failures (then blame them for the accidents)
  - Pilot mental model is critical. Where is this in the risk assessment?
- Unrealistic assumptions about hardware and software
  - Redundancy as a mitigation:
    - Doesn't work for software or for design errors in hardware
    - Software ONLY has design errors
  - Virtually all software-related accidents stem from requirements errors, not implementation errors
    - Redundancy and rigor of software development will not help here

Empirical Evaluations (4)

- We found items categorized as

  Severity = Catastrophic

  Likelihood = Low

  that had been involved in multiple accidents for those systems

- Only improbable if ignore software requirements flaws, human behavior aspects, etc.
- STPA found non-failure scenarios leading to catastrophic events that were omitted from official risk assessment
- STPA identified realistic and relatively likely scenarios leading to all of specific failures dismissed as improbable in official risk assessment.
- Likelihood can differ significantly depending on external environment and operations in which a failure occurs.

Technical Limitations

- The use of the risk matrix itself has been shown to have mathematical and other limitations (see paper)
- Most important stem from Heuristic Biases (Kahnemann, Tversky, Slovic)
  - Psychologists who studied how people actually do risk evaluations
  - Humans, it turns out, are terrible at estimating risk

Heuristic Biases (Tversky, Slovic, and Kahneman)

- Confirmation bias (look for data that supports our beliefs)
- Construct simple causal scenarios
  - If none comes to mind, assume impossible
- Tend to identify simple, dramatic events rather than events that are chronic or cumulative
- Incomplete search for causes
  - Once one cause identified and not compelling, then stop search
- Defensive avoidance
  - Downgrade accuracy or don't take seriously
  - Avoid topic that is stressful or conflicts with other go

Heuristic Biases

Can avoid by: Providing those responsible with better information, obtained through a structured process to generate scenarios.

That goal be accomplished using more powerful hazard analysis techniques, such as STPA

Potential Alternatives to the Risk Matrix

1. Use hazards (not failures) and better information about potential causal scenarios

2. Change basic definition of risk and how it is assessed (not covered in this talk)

Use Hazard Rather than Failures

- Relationship between individual failures and losses is not obvious.
  - Assessing hazards is a more direct path to ultimate goal
  - Component reliability is not equivalent to system safety
  - Using hazards is traditional in system safety

Example: Why Should Use Hazards

- Helicopter Deice Function
- Final SAR included a failure of APU resulting from chaffing.
  - Important because APU used when loss of one generator occurs during blade deicing
  - But also another scenario identified by using STPA that could occur when APU has not failed

  UCA: *The flight crew does not switch the APU (Auxiliary Power Unit) generator power ON when either GEN1 or GEN2 are not supplying power to the helicopter and the blade de-ice system is required to prevent icing*.

  - Several causal scenarios and factors, but they are not in official SAR
  - Need to be factored into any risk assessment

Change Being Recommended

- Start from a prioritized list of stakeholder identified accidents or system losses.
- Identify high-level system hazards leading to these losses
- Assess severity and likelihood of *hazards*
- Only consider failures that can lead to hazards (identified by STPA) along with the non-failure scenarios (again, STPA can identify them)
- Consistent with MIL-STD-882 and most other safety standards

Likelihood as Strength of Potential Controls

- Severity now easy because can be traced directly to list of accidents or mishaps
- Heuristic biases lead to poor estimates of likelihood
- Following a rigorous STPA will result in
  - Reducing shortcuts and biases
  - More full consideration of potential causal scenarios
- Can be done early in development to identify where to place development effort
- Maybe focus on component behavior because have historical failure information

Example 1: Pilot's use of flight controls

- UCA: *The Flight Crew does not deflect pedals sufficiently to counter torque from the main rotor, resulting in the Flight Crew losing control of the aircraft and coming into contact with an obstacle in the environment or the terrain.*

One of causal scenarios:

- *Scenario 1: The Flight Crew is unaware that the pedals have not been deflected sufficiently to counter the torque from the main rotor.*
- The Flight Crew could have this flawed process model because:
  - *a) The flight instruments are malfunctioning and providing incorrect or insufficient feedback to the crew about the aircraft state during degraded visual conditions.*
  - *b) The flight instruments are operating as intended, but providing insufficient feedback to the crew to apply the proper pedal inputs to control heading of the aircraft to avoid obstacles during degraded visual conditions.*
  - *c) The Flight Crew has an incorrect mental model of how the FCS will execute their control inputs to control the aircraft and how the engine will respond to the environmental conditions.*
  - *d) The Flight Crew is confused about the current mode of the aircraft automation and is thus unaware of the actual control laws that are governing the aircraft at this time.*
  - *e) There is incorrect or insufficient control feedback.*

Example 1: Pilot's use of flight controls (Con't)

- Each causal factor used to generate requirements and design features to reduce their likelihood of occurring
- Likelihood can be based on *strength of potential controls*
  - Interface design (evaluated by human factors expert)
  - Redundancy and fault tolerant design
  - Training
  - System design (hardware, software, interactions)
  - Design of feedback
- Still need a way to link these to likelihood (will come back to that)

Example 2: Software

- What do now---rigor of development---makes no sense technically

  UCA: *One or more of the FCCs (flight control computers) command collective input to the hydraulic servos too long, resulting in an undesirable rotor RPM condition and potentially leading to the hazard of violating minimum separation from terrain or the hazard of losing control of the aircraft.*

- At least 5 causal scenarios why the FCCs might do this

Example (2): Software

Scenario 1: The FCCs are unaware that the desired state has been achieved and continue to supply collective input.

a) The FCCs are not receiving accurate position feedback from the main rotor servos.

b) The FCCs are not receiving input from the ICUs to stop supplying swashplate input.

Scenario 2: The FCCs do not send the appropriate response to the aircraft for particular control inputs. This could happen if:

a) The control logic does not follow intuitive guidelines that have been implemented in earlier aircraft, perhaps because requirements to do so were not included in the software requirements specification.

b) The hardware on which the FCCs are implemented has failed or is operating in a degraded state.

Scenario 3: The FCCs do not provide feedback to the pilots to stop commanding collective increase when needed because the FADEC (engine controller) is supplying incorrect cues to the FCCs regarding engine conditions.

Scenario 4: The FCCs do not provide feedback to the pilots to stop commanding collective increase when needed because the FCCs are receiving inaccurate NR (rotor rpm) sensor information from the main rotor.

Scenario 5: The FCCs provide incorrect tactile cueing to the ICUs (inceptor control units) to properly place the collective to prevent low rotor RPM conditions.

Example 2: Software (con't)

- Scenarios used to identify appropriate FCC requirements and design constraints.
- For example, for Scenario 1:
  - *1. The FCCs must perform median testing to determine if feedback received from the main rotor servos is inaccurate.*
  - *2. The PR SVO FAULT caution must be presented to the Flight Crew if the FCCs lose communication with a main rotor servo.*
  - *3. The EICAS must alert the Flight Crew if the FCCs do not get input from the ICU every x seconds.*
- Translate these into "likelihood" (final piece of puzzle)

Translating Strength of Controls into Likelihood

Qualitative Ranking such as

1. The causal factor can be eliminated through design and high assurance.
2. The occurrence of the causal factor can be reduced or controlled through system design
3. The causal factor can be detected and mitigated if it does occur through system design or through operational procedures
4. The only potential controls involve training and procedures.

Maybe too simplistic?

- Could include how thoroughly the causal factor has been handled within each category
- Combinations of possible controls?

Translating Strength of Controls into Likelihood (2)

- May be able to come up with more sophisticated procedures for specific types of systems.
- Examples in paper on this topic at:

  http://sunnyday.mit.edu/Risk-Matrix.pdf

  Architectural trade study for space exploration

  Air Traffic Control enhancements

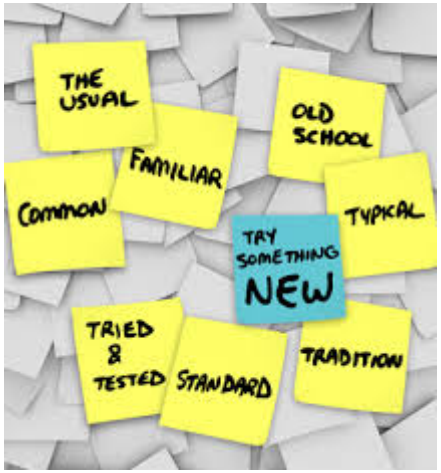
Additional Considerations

- Risk also affected by factors during manufacturing and operations:
  - Manufacturing controls
  - Designed maintainability and maintenance errors
  - Training programs
  - Changes over time in usage environment
  - Consistency and rigor of management and oversight
  - Assumptions during development about operational environment: how well communicated to users and how rigorously are enforced during operations
  - etc.

Additional Considerations (2)

- Including these factors will improve risk assessment
- Should also track factors and improve risk assessment over time
  - Risk assessment process need not stop at deployment
  - Risk-based decisions needed throughout life cycled
  - Castilho: Active STPA
    - Identify leading indicators of increasing risk during operations

Conclusions



- Can provide improved risk matrix processes
- Start from hazards, not failures, to get more realistic assessments of risk
- STPA and better causal analysis can greatly improve likelihood estimates
- Suggestions were provided and other people should be able to create even better processes
- But limited by the use of the Risk Matrix and current definition of risk
  - Alternative is to improve definition of risk and its evaluation
  - Suggestions for this goal will follow (soon)