MIT ESD

Massachusetts Institute of Technology
Engineering Systems Division

ESD Working Paper Series

A Systems Thinking Approach to Leading Indicators in the Petrochemical Industry

Nancy Leveson
Professor of Aeronautics and
Astronautics and Engineering Systems
Massachusetts Institute of Technology

ESD-WP-2013-01 January 2013 esd.mit.edu/wps

A Systems Thinking Approach to Leading Indicators in the Petrochemical Industry[1]

Nancy Leveson

There are always warning signs before a major accident, but these signs may only be noticeable or interpretable as a leading indicator in hindsight. Before an accident, such "weak signals" are often perceived only as noise. To ask people to "be mindful of weak signals" is asking them to do something that is impossible. There is always a lot of noise and always a lot of signals that do not presage an accident. The problem then becomes how to distinguish the important signals from all the noise. Defining effective leading indicators is a way to accomplish this goal by providing specific clues that people need to look for. Asking people to "look for anything that might be an important sign" is usually asking them to do the impossible.

Almost all of the past effort to identify leading indicators has involved finding a set of generally applicable metrics or signals that presage an accident. Examples of such identified leading indicators are quality and backlog of maintenance, inspection, and corrective action; minor incidents such as leaks or spills, equipment failure rates, and so on. There is commonly a belief—or perhaps, hope—that a small number of such "leading indicators" can identify an increase in risk of an accident. While some general indicators may be useful, large amounts of effort over decades has not provided much progress.[2] The lack of progress may be a sign that such general, industry-wide indicators do not exist or will not be particularly effective in identifying increasing risk. An alternative is to identify leading indicators that are specific to the system being monitored.

This paper proposes an approach to identifying and monitoring system-specific leading indicators and provides some guidance in designing a risk management structure to use such indicators effectively. The approach is based on the STAMP model of accident causation and tools that have been designed to build on that model. STAMP extends current accident causality to include more complex causes than simply component failures and chains of failure events. It incorporates basic principles of systems thinking and is based on systems theory rather than traditional reliability theory. The next section briefly describes STAMP and STPA, the latter being a new hazard analysis technique based on STAMP. Then the proposal for a new approach to generating and managing leading indicators is outlined.

1. Background

Almost all major accidents (at least viewed after the fact) have multiple precursors and cues that an accident is likely to happen. The reason is that most major accidents do not result simply from a unique set of proximal, physical events but from the drift of the organization to a state of heightened risk over time as safeguards and controls are relaxed due to conflicting goals and tradeoffs.[3] Almost all the factors involved in the Bhopal accident, for example, existed before the actual triggering event that led directly to the loss. The plant was losing money and Union Carbide had ordered that costs be reduced, without considering how these cuts might conflict with safety. Requirements in the operating manual, such as never filling the tanks more than half their volume, the use of safety equipment for potentially hazardous operations, and the operation of a refrigeration unit to keep the MIC at a safe temperature were not followed. In fact, when the

[1] This research was partially supported by a research grant from BP.

[2] Ibrahim Khawaji, Developing System-Based Leading Indicators for Proactive Risk Management in the Chemical Processing Industry, SM Thesis, MIT, 2012.

[3] Jens Rasmussen, Risk Management in a Dynamic Society: A Modeling Problem, Safety Science 27 (2/3): 183-213, 1997.

refrigeration unit was turned off (most likely to save money), the high temperature alarm threshold was raised, which eliminated the possibility of an early warning of rising temperatures. Valves leaks and gauges frequently were inaccurate or out of order. Maintenance procedures were severely cutback and critical jobs were left unfilled in shifts when someone called in sick. A review and audit two years before had noted that many of the safety devices, such as alarms, the flare tower and the gas scrubber, were inoperable or inadequate. Most of the specific practices leading directly to the accident, such as filter-cleaning operations without using slip blinds, leaking valves, bad pressure gauges, etc., were noted in the report and never fixed. Union Carbide did not follow up to ensure the deficiencies were corrected. Qualifications of personnel went down. Training and oversight was reduced. A similar accident had occurred the year before at the plant but under circumstances where the results were less severe (one person was killed), but nothing was done about fixing the hazardous operation of the plant. In this state, some events were bound to occur that would trigger an accident.

While the events and practices at Bhopal were strikingly bad, in hindsight nearly every major accident has similar migration toward the accident over time that potentially could have been detected and the accident prevented. The challenge in preventing accidents is to establish safeguards and metrics to prevent and detect migration toward a state of unacceptable risk before an accident occurs.

But this detection alone is not enough—there must be a management process in place to act when the leading indicators show that action is necessary. Note that at Bhopal there had been an audit report showing the conditions existed but they were not adequately addressed.

The process of tracking leading indicators of increasing risk embedded within an effective risk management structure can play an important role in preventing accidents, but a way to derive effective leading indicators is required. The signs are not always as clear as at Bhopal, and, of course, we cannot wait until hindsight shows us what we should have noted at the time.

Traditional accident causation models explain accidents in terms of a chain of directly related events that cause the accident. Such models, however, are limited in their ability to handle accidents in complex systems, organizational and managerial (social and cultural) factors in accidents, and the systemic causes of the events. STAMP (System-Theoretic Accident Model and Processes) is a new model of accident causation that extends the old models to include non-linear and indirect relations and thus can better handle the levels of complexity and technical innovation in today's systems.[4] A systems-theoretic model allows capturing the non-linear dynamics of interactions between system components and anticipating the risk-related consequences of change and adaptation over time.

In STAMP, accidents are conceived as resulting not simply from system component failures but more generally from interactions among system components (both physical and social) that violate system safety constraints. Examples of safety constraints are that a highly reactive chemical must be stored below a maximum temperature, pressure in a well must be within acceptable levels at all times, and the level of liquid in the ISOM tower must always remain below a maximum level and pressure. The constraints must be enforced in the operating process and contingency action must be taken if the constraints are somehow violated.

In STAMP, process safety is treated as a control problem, not as reliability problem: accidents occur when component failures, external disturbances, and/or potentially unsafe interactions among system components are not controlled (handled) adequately. The controls may be managerial, organizational, physical, operational, or manufacturing. Major accidents rarely have a single root cause but result from an adaptive feedback function that fails to maintain safety as performance changes over time to meet a complex and changing set of goals and values.[5] The

[4] Nancy Leveson, *Engineering a Safer World: Applying Systems Thinking to Safety*, MIT Press, 2012.

[5] Rasmussen, op.cit.

accident or loss itself results not simply from component failure or human error (which are symptoms rather than root causes) but from the inadequate control of safety-related constraints on the development, design, construction, and operation of the entire socio-technical system. We will use this causality model in developing leading indicators.

Safety-related constraints are enforced by a safety control structure that must be carefully designed and evaluated to ensure that the controls are adequate to maintain the constraints on behavior necessary to control risk. Figure 1 shows the safety control structure existing at the time of the Macondo well blowout. Each component has specific assigned responsibilities for maintaining the safety of the system, that is, enforcing the safety constraints and preventing losses. For example, the mud logger is responsible for creating a detailed record of a borehole by examining the contents of the circulating drilling medium, the cementer is responsible for properly sealing off a wellbore, and local management has responsibilities for overseeing that these and other activities are carried out properly and safely. The government oversight agency may be responsible for ensuring that safe practices are being followed and acceptable equipment being used. And so on.

Major accidents are rarely the result of unsafe behavior by only one of the components but usually the result of unsafe interactions among and behavior by all or most of the components in the control structure. When accidents occur and they are investigated thoroughly, as was true for the Macondo well blowout, it almost always turns out that more than one component (and often all) did not fulfill its control responsibilities. Also, as shown in Figure 1, more than one company may participate in the safety control structure, with the controllers of the components (whether part of their own company or another) having individual responsibilities for ensuring that the controlled processes or components are fulfilling their safety responsibilities.

Figure 2 shows a more general example of a safety control structure with a focus on producing a product. This structure might be more typical for an oil refinery where the government agency involved in safety oversight might be OSHA or the EPA. As with Figure 1, each component in the control structure has responsibility for controlling the behavior of some lower level components in the structure.

Between the levels of the safety control structure there are classic feedback control loops: the controllers provide control actions to maintain a "set point," in this case a set of safety constraints on the behavior of the controlled process. In turn they get feedback from the controlled process to assist in providing appropriate and effective control actions. Feedback may be direct from the physical process, such as sensors that provide information about the state of the well at that point in time, or may involve feedback from lower level controllers to higher level controllers to provide information about the current state of their activities and the perceived level of risk.

Note that the use of the term "control" does not imply only managerial and operator controls. Physical component behavior and interactions among components can be controlled through the use of physical controls such as interlocks or through various types of barriers and fault tolerance features.

In addition to physical and managerial controls, all behavior is influenced and at least partially "controlled" by the social and organizational context in which the behavior occurs. Control is provided not only by engineered systems and direct management intervention, but also indirectly by policies, procedures, shared values, and other aspects of the organizational culture, sometimes called the "safety culture."

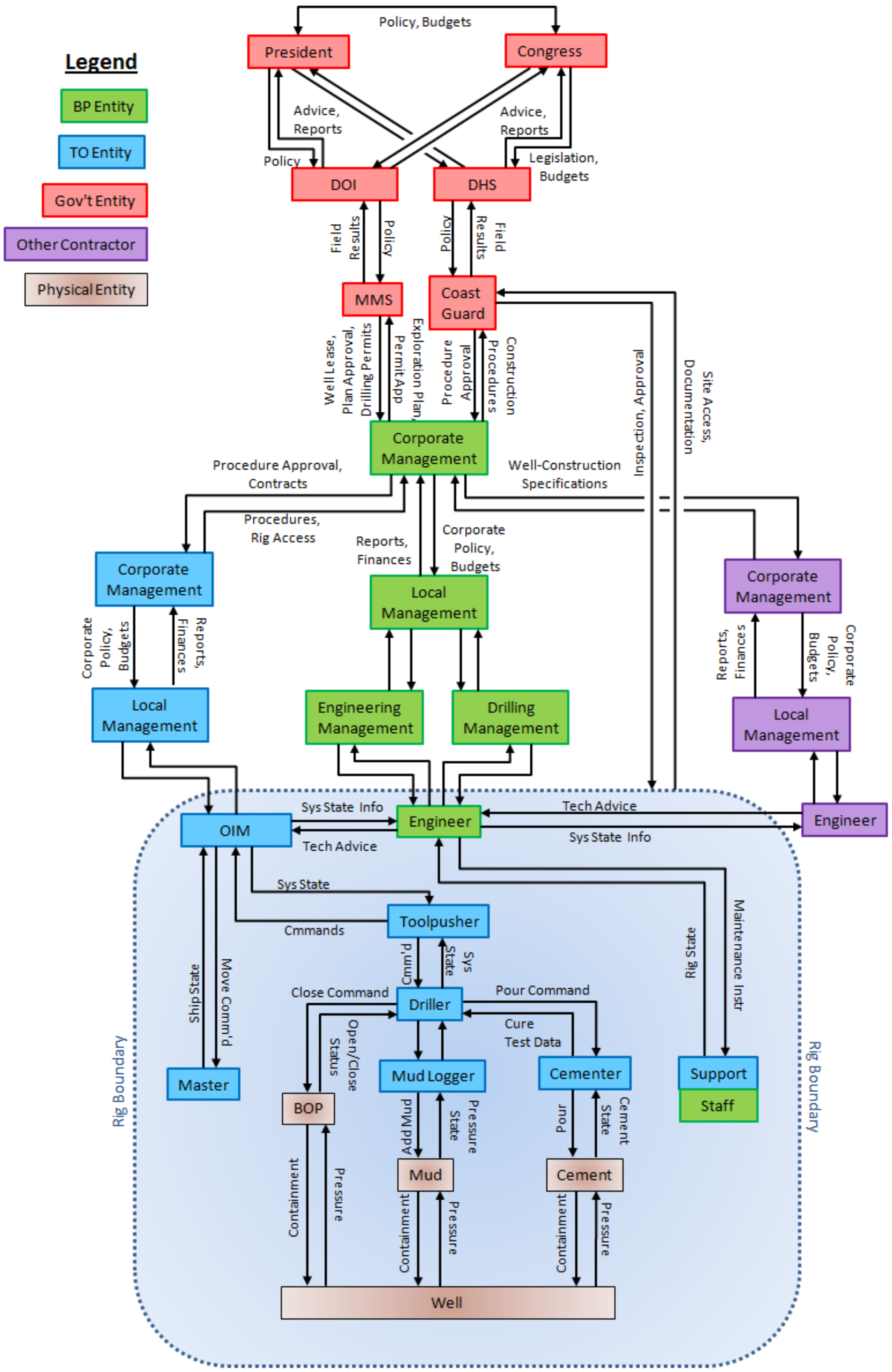


Figure 1: The General Safety Control Structure Existing at the Time of the Macondo Accident

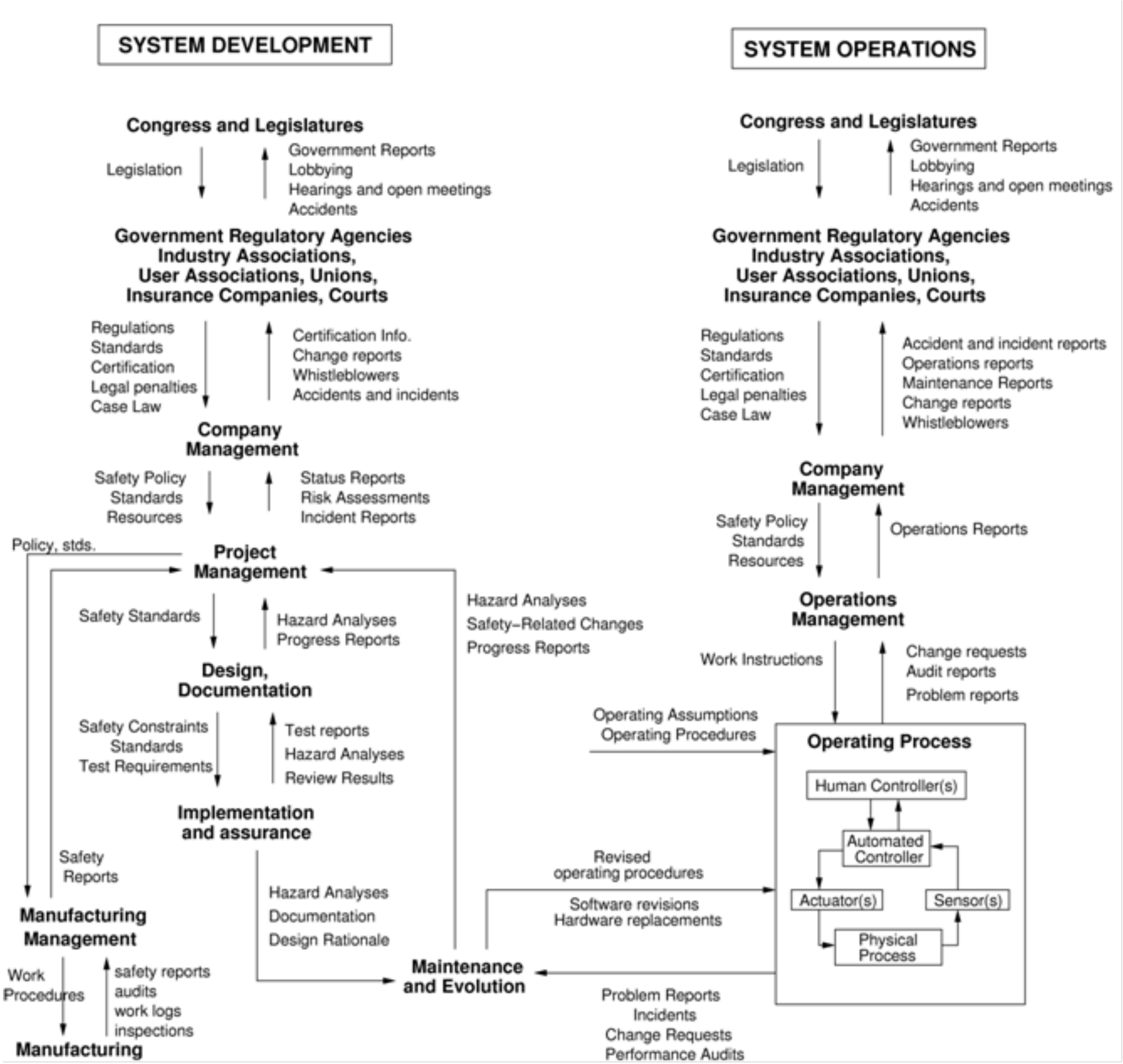


Figure 2: An Example of a Generic Safety Control Structure

In addition to the concepts of safety constraints and a safety control structure, one other important concept is needed in formulating safety as a control problem. In basic systems and control theory, in order to provide effective control, the controller must have an accurate model of the process it is controlling (Figure 2). For human controllers, this model is commonly called the mental model. For both automated and human controllers, the process model or mental model is used to determine what control actions are necessary to keep the system operating effectively.

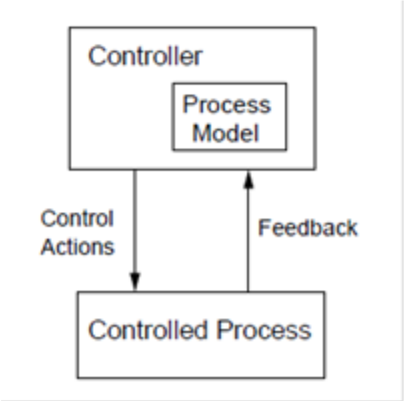


Figure 3: Every Controller Uses a Model of the State of the Controlled Process to Determine What Control Actions are Needed

The process model includes assumptions about how the controlled process operates and about the current state of the controlled process. Accidents in complex systems, particularly those related to software or human controllers, often result from inconsistencies between the model of the process used by the controller and the actual process state. The inconsistency contributes to the controller providing inadequate control. The local BP manager on Deepwater Horizon thought the cement had properly sealed the annulus and ordered the mud to be removed, the operators at Texas City thought the level of liquid in the isomerization unit was below the appropriate threshold, and the operator at Whiting in an overflow accident did not know that the CO_2 tank was filling. Usually, these models of the controlled system become incorrect due to missing or inadequate feedback and communication channels, Deepwater Horizon had limited sensors to provide information about the state of the well; Texas City had no sensors above the maximum fill level of the tank; and the operator at Whiting had closed the intake valve, the control board showed that it was closed, the flow meter showed no flow, and the high level alarm did not sound because it was inoperative.[6]

The effectiveness of the safety control structure in preventing accidents is greatly dependent on the accuracy of the information about the state of the controlled system each controller has, often in the form of feedback from the controlled process although other sources of such information can and often does exist. Performance metrics and leading indicators of changes in the safety control structure are a form of feedback and can provide a means of measuring the risk in the current state of the process and the safety control structure. They provide important signals about the potential for an accident.

STPA [2] is a new hazard analysis technique based on this theoretical STAMP accident causality model. It is basically a rigorous method for examining the control loops in the safety control structure to find potential flaws and the potential for (and causes of) inadequate control. Because the STAMP framework extends current accident models and thus includes accidents caused by component failure(s), STPA not only identifies the hazard scenarios identified by fault trees, HAZOP, and other traditional hazard analysis methods, but it also includes those factors not included or poorly handled by these traditional methods such as software requirements errors, component interaction accidents, complex human decision-making errors, inadequate coordination among multiple controllers, and flawed management and regulatory decision making. Note that

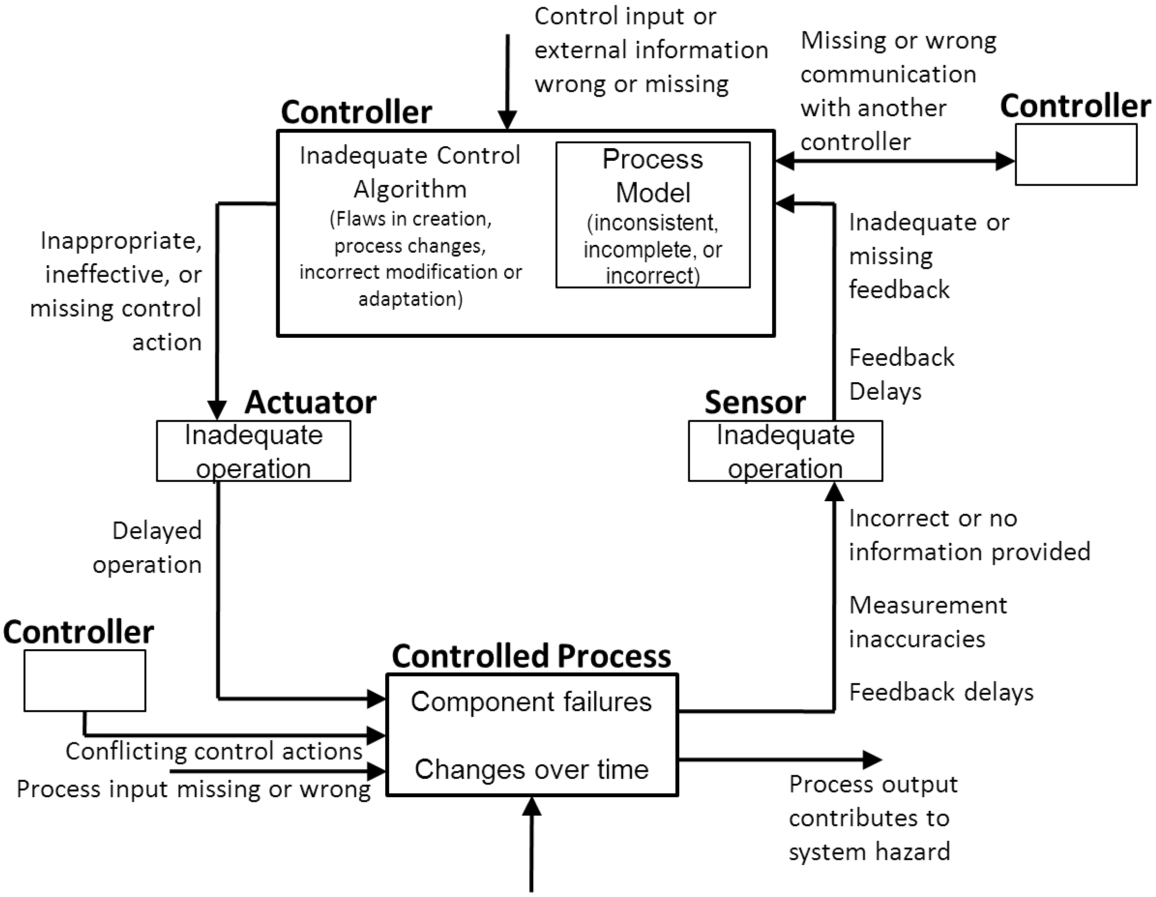
[6] Amazingly the operator still received primary blame in the accident report for not responding quickly enough although he got no information there was a problem and was distracted by another concurrent emergency alarm in the plant.

STPA, unlike the traditional hazard analysis techniques, works not only on the physical system but on the management structure and organizational design.

Figure 4 shows some of the types of general flaws considered in an STPA analysis. The analysis is performed on a specification of the system's safety control structure and is broken into two steps in order to more carefully structure it.

The first step in STPA assists in identifying the safety control requirements. There are four types of inadequate control that can lead to accidents:

1. An action required for safety is not provided or not followed, e.g., the operator does not close the intake valve when the tank is full.
2. An unsafe control action *is* provided, e.g., mud is removed before the well has been properly sealed.
3. A potentially safe control action is provided too early or too late, that is, at the wrong time or in the wrong sequence, e.g., the plant evacuation signal is delayed or, as in Bhopal, the operators do not investigate a reported leak until after the tea break.
4. A control action required for safety is stopped too soon or applied too long.



Control input or external information wrong or missing
Missing or wrong communication with another controller
Controller
Controller
Inadequate Control Algorithm
(Flaws in creation, process changes, incorrect modification or adaptation)
Process Model
(inconsistent, incomplete, or incorrect)
Inappropriate, ineffective, or missing control action
Inadequate or missing feedback
Feedback Delays
Actuator
Inadequate operation
Sensor
Inadequate operation
Delayed operation
Incorrect or no information provided
Measurement inaccuracies
Feedback delays
Controller
Controlled Process
Component failures
Changes over time
Conflicting control actions
Process input missing or wrong
Process output contributes to system hazard

Figure 4: Some General Factors involved in Unsafe Control

After the potentially unsafe control actions are identified, the second step in STPA determines how these unsafe control actions could occur, that is the scenarios that can lead to a hazardous system state or accident. Like HAZOP (but not fault trees), STPA works on a system model and guidance is provided by STPA on what to look for so that omissions of scenarios or causes are less likely to occur. A major difference from HAZOP is that in HAZOP the system model is a physical model of the plant, while STPA uses a functional control structure model.

Accidents usually involve factors in both the plant (product) and in the process (SEMS) and both need to be designed properly to prevent accidents. An effective risk management program starts with hazard analysis and design for safety, that is, with identifying and then eliminating or controlling the hazards. The information developed in the design process should be used to design the operational safety management system. Figure 5, from my book *Engineering a Safer World*, provides an overview of the entire safety process.

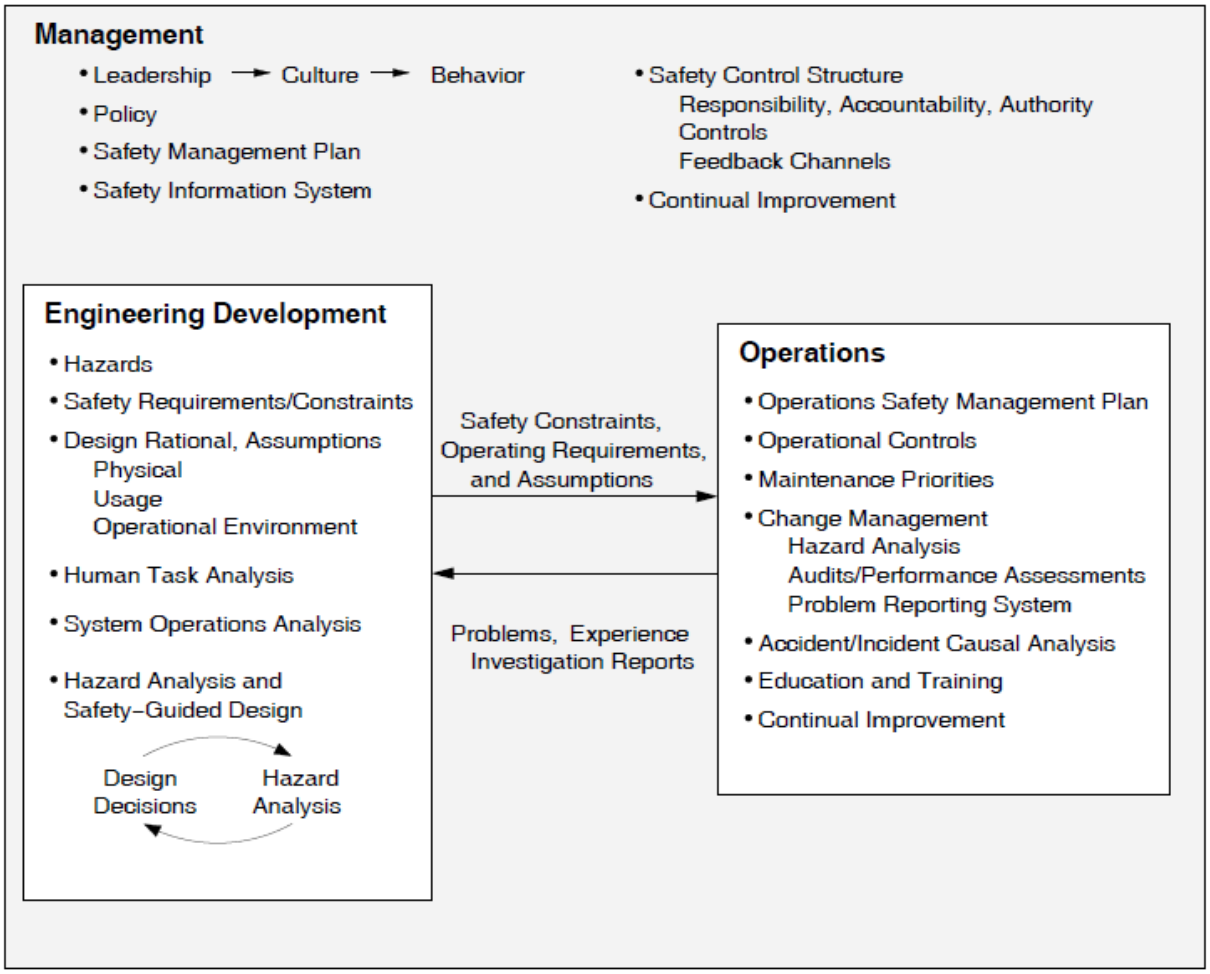

Management
• Leadership → Culture → Behavior
• Policy
• Safety Management Plan
• Safety Information System
• Safety Control Structure
Responsibility, Accountability, Authority
Controls
Feedback Channels
• Continual Improvement
Engineering Development
• Hazards
• Safety Requirements/Constraints
• Design Rational, Assumptions
Physical
Usage
Operational Environment
• Human Task Analysis
• System Operations Analysis
• Hazard Analysis and Safety–Guided Design
Design Decisions
Hazard Analysis
Safety Constraints, Operating Requirements, and Assumptions
Problems, Experience Investigation Reports
Operations
• Operations Safety Management Plan
• Operational Controls
• Maintenance Priorities
• Change Management
Hazard Analysis
Audits/Performance Assessments
Problem Reporting System
• Accident/Incident Causal Analysis
• Education and Training
• Continual Improvement

Figure 5: A Systems Engineering Approach to Managing Safety and Risk

Note that engineering development needs to pass, among other things, the operating and safety assumptions underlying the development process to operations so they can be used to design an Operations Safety Management Plan and to provide day-to-day safety control. Creating and monitoring leading indicators is an important aspect of that control and uses the safety constraints and operational assumptions provided.

As noted at the beginning of this paper, an alternative to identifying general leading indicators for risk in all systems is to identify more specific leading indicators for particular system designs and SEMS.[7] Using STAMP terminology, these indicators will be related to the system safety constraints, i.e., constraints on the behavior of the system to ensure accidents do not occur for that particular system or organization.

Leading indicators will be similar for different organizations only to the extent that the hazards, safety constraints, system design and safety control structure are similar. There are potentially many designs for an effective safety control structure and leading indicators need to be identified for the particular one being used. Leading indicators may only be applicable to a particular physical refinery or well design, but more general leading indicators at higher levels in the safety control structure, such as the organization's SEMS design, may apply organization wide. For example, everyone using a particular off-shore oil drilling rig may be able to use similar leading indicators related to the physical platform although appropriate indications may differ with respect to the management and operations of the organizations using the platform. For refineries owned by one company, such as BP, the leading indicators may be different for each physical refinery but may be the same at higher levels of the control structure (which will be common to all the refineries). Basically, the leading indicators should reflect the specific physical or organizational structure they are monitoring.

There are three requirements for using leading indicators to reduce accidents: (1) identifying appropriate and effective leading indicators, (2) creating a safety indicator monitoring program to use them, and (3) embedding this monitoring system within a risk management program. The next three sections propose a process for each of these in turn.

2. Identifying Leading Indicators

To start, we need to consider why accidents occur. Leading indicators will necessarily be related to our understanding of accident causality and the *assumptions* underlying our decision making.

The idea of assumptions being the basis for identifying leading indicators is not original but has been proposed for more general risk management programs. RAND developed the methodology of assumption-based planning (ABP) primarily to assist U.S. Army clients with mid- and long-term defense planning and to reduce uncertainty and manage risk.[8] Some terminology and ideas from ABP are used in the leading indicator process being proposed in this paper.

Why Do Accidents Occur?

Despite much effort to avoid them, accidents still occur. One notable exception is the SUBSAFE program, but that only targets a small part of the safety problem. Theoretically, if we design a safe system, that is, eliminate or adequately control or mitigate all the hazards and nothing changes, then we should not have accidents. The problem is that neither of these conditions is usually true in practice: no engineering process is perfect, and every system and its environment is subject to change over time.

The causes for accidents may arise in the development and implementation of the system or in operations or may reflect management and cultural deficiencies. The following list shows the way accident causes can arise in each of these phases:

[7] The term SEMS (Safety and Environmental Management System) is commonly used in the process industries and is the same as the SMS (Safety Management System) terminology used in most other industries. Depending on one's particular definition of SEMS or SMS, the safety control structure as defined in STAMP contains all or much of it. An existing SEMS can usually be mapped into a safety control structure although we often find that important controls are missing from the SEMS.

[8] James Dewar, *ibid*.

1. Development and implementation
   - Inadequate hazard analysis (assumptions about the system hazards or the process used to identify them do not hold)
     - HA is not performed or is not completed
     - Some hazards are not identified or are not handled because they are assumed to be "sufficiently unlikely"
     - HA is incomplete (important causes are omitted).
   - Inadequate identification and design of control and mitigation measures for the hazards (possibly due to inappropriate assumptions about operations)
   - Inadequate construction of control and mitigation measures
2. Operations
   - Controls that designers assumed would exist during operations are not adequately implemented.
   - Changes over time violate the assumptions underlying the design and controls
     - New hazards arise with changing conditions, were not anticipated during design and development or were dismissed as unlikely to occur
     - Physical controls and mitigation measures degrade over time in ways not accounted for in the analysis and design process
     - Components (including humans) behave differently over time (violate assumptions made during design and analysis)
     - The system environment changes over time (violates assumptions made during design and analysis)
3. Management
   - The design of the safety control structure (SEMS) is flawed
   - The safety control structure does not operate the way it was designed (assumed) to operate, While there may be many reasons for this, one general cause is that the safety culture, i.e., the goals and values of the organization with respect to safety, degrades over time. In addition, the behavior of those in the safety control structure may be influenced by competitive, financial or other pressures.

Although some of these causes involve engineering errors, for example, inadequate design of physical control mechanisms such as the BOP at Macondo, a large number can be framed in terms of assumptions that do not hold, either originally or as changes occur over time. Examples include assumptions that may be involved in identifying the system hazards or in excluding potential hazards that are deemed unlikely. There are also always critical assumptions about how the system components will behave and about the environment in which the system operates.

The evaluation of "likelihood" when assessing risk is a key assumption that, if incorrect, can lead to accidents that might have been prevented. Too often, there is no scientific basis for making such assumptions about likelihood and occasionally politics intervenes. After accidents, it is common to find that the hazard involved had been identified but not controlled because it was deemed too unlikely to occur. If the likelihood evaluation is based on scientific knowledge, such as the laws of thermodynamics, then likelihood seems safe to use. If it is not, then the use of likelihood is so subject to bias (see section below about bias in decision making about risk) that it should not be used.

Instead of trying to predict the likelihood that an event will occur or an assumption will fail, the similar but different concept of *vulnerability* can be used.[9] Vulnerability in the world of assumption-

[9] James Dewar, *Assumption-Based Planning*, Cambridge University Press, 2002.

based planning involves assessing whether an assumption could plausibly fail during the lifetime of the system, not the specific probability of that happening. Trying to assess the exact probability of a valve failing or operators making a mistake or changing their behavior is impossible in a world of imperfect manufacturing and changing human behavior. If an assumption is vulnerable and the cost of eliminating or controlling it is reasonable, then it makes no sense not to protect against it. Decisions about what form that protection takes, in the form of a shaping or hedging action (see below), or perhaps only controls during operations, will involve assessments of severity and cost.

The goals and values of the organization, i.e., the safety culture, is an important assumption that when wrong can be a major factor in accidents and must be reflected in the set of leading indicators. For example, a safety policy is a basic requirement for every company or organization to communicate the desired safety culture. There must be a way to measure how well that policy is being followed and if that behavior changes over time. Assumptions about management behavior and decision making are also commonly found to be violated after accidents occur and must be monitored.

Sometimes the safety-related assumptions underlying the system or operational design process hold originally but become untrue due to changes over time. The world is constantly changing, especially human behavior and, as noted, major accidents are usually preceded by the migration of the system to a state of unrecognized high risk. Using the terminology developed so far, that migration basically involves moving to states where the assumptions used during the design and development of the system are violated. So even if a great job in terms of hazard analysis and design for safety have been done during development and manufacturing, we still almost inevitably will have accidents. This potential for migration toward the violation of the assumptions underlying the safety of the system needs to be reflected in the set of leading indicators. For example, operators may start to take shortcuts or turn off safety devices in order to operate more efficiently.

We can now propose some relevant definitions:

> **Leading indicator**: A warning sign that can be used to monitor safety-related assumptions, events or thresholds that, if detected, signifies that a safety-related assumption is broken or dangerously weak and that action is required to prevent an accident. Indicators or warning signals that the validity or vulnerability of an assumption is changing.
>
> S**haping actions**: Actions intended to maintain assumptions, to prevent hazards and to control migration to states of higher risk. These are essentially actions taken during the design of the physical system or the SEMS to prevent hazards and to prevent the violation of the assumptions underlying the analysis and design. In control theory terms, these provide feed-forward control and are built into the system and SEMS design either originally or later incorporated in response to an accident or serious incident. An example might be an interlock to ensure that two events occur in a particular sequence or the use of a desiccant to prevent moisture that could lead to corrosion in a tank or pipe. For human behavior, shaping actions may be to make the operation of a safety control action easy and difficult to omit. A final example of a shaping action is the design of operational procedures to be followed under various types of conditions and following hypothesized events, such as creating an evacuation plan.
>
> **Hedging actions**: Actions that prepare for the possibility that an assumption will fail, including checking the assumptions during operations. Hedging actions come from thinking through a possible scenario (hazard analysis) in which the assumption collapses and asking what might be done now to prepare for that scenario. Another way of saying this is that the hazard analysis generates scenarios from broken assumptions (worst case analysis) to identify hedging actions that might be taken. In control theory and STAMP, hedging actions

involve feedback control using set points that are safety constraints during operation of the system. Examples include performance audits to determine whether the system and the safety controls are operating as designed and operators are following designed procedures.[10]

Signposts: Points in the unfolding future where changes in the current safety control structure design may be necessary or advisable. In essence, they involve planning for monitoring and responding to particular identified changes in the assumptions underlying the safety control structure. For example, new construction or planned future changes in the system or likely changes in the environment may trigger a planned response.

Assumption checking: The process of checking whether the assumptions underlying the safety design are still valid. The difference with signposts is that signposts are identified during the design and development process and specific responses designed. In assumption checking, risk managers and controllers monitor the system (perhaps for signposts or perhaps just for changes and failures of assumptions that had not been predicted) during the operation of the plant and ask whether the assumptions are still valid.

The Role of Psychological Biases in Hazard Analysis and a Leading Indicator Program

Psychologists have written extensively about the biases inherent in assessing risk.[11] These biases may have an impact on which indicators we design and how we react to them. For example, *confirmation bias* is the name given to the tendency of people to pay more attention to information that supports their views than to evidence that conflicts with them. So people tend to be overconfident in the accuracy of their forecasts, tending to deny uncertainty and vulnerability.

Another common bias is called the *availability heuristic* and suggests that people tend to base likelihood judgments of an event on the ease with which instances or occurrences of that or similar events can be brought to mind. While this heuristic may often be a reasonable one to use, it can also lead to systematic bias. For example, psychologists have found that judgments of the risk of various hazards or events will tend to be correlated with how often they are mentioned in the news media.

A third bias occurs when people think about future events whose likelihood cannot be based on past historical rates. They will often construct their own simple causal scenarios of how the event could occur, using the difficulty of producing reasons for an event's occurrence as an indicator of the event's likelihood. If no plausible cause or scenario comes to mind, an assumption may be made that the event is impossible or highly unlikely.

People also have difficulty predicting *cumulative causes*. They tend to identify simple, dramatic events rather than causes that are chronic or cumulative. Dramatic changes are given a relatively high probability or likelihood whereas a change resulting from a slow shift in social attitudes is more difficult to imagine and thus is given a lower probability. At the same time, the conjunction fallacy says that an outcome paired with a likely cause is often judged to be more probable than the outcome alone even though this conclusion violates the laws of probability.

A further bias is caused by an *incomplete search for possible causes*. Searches are often stopped once one possible cause or explanation for an event has been identified. If that first possible cause is not very compelling, stopping the search can mean that other, more plausible and compelling causes, are not identified and likelihood is underestimated.

[10] Finding out that the operators are not following procedures does not imply that the correct action is to retrain them or to enforce the procedures. Instead, the response should first be to determine why the operators are not following the procedures, perhaps because conditions have changed but the procedures have not.

[11] Some examples are Kahneman, Tverskly, Fischoff, and Sloman.

A final common psychological bias is called *defensive avoidance*. This type of bias may be reflected in the rejection or downgrading of the accuracy of leading indicators or people's inability to take them seriously or to accept that risk may be increasing. Defensive avoidance is based on the common psychological tendency to rationalize and avoid consideration of a topic that is stressful or conflicts with other pressing goals.

In addition to these psychological biases, organizational culture and politics can cause likelihood and vulnerability to be under- or over-estimated.

These common biases and others are one of the reasons we need a structured method for identifying assumptions and possible causes of hazards. Following a structure process diminishes the power of our biases and encourages us to do a more thorough search. Biases may also have an impact on decisions about which leading indicators to use and in recognizing the changes that do occur and accepting that the leading indicator is in fact accurately predicting increased risk (defensive avoidance).

In addition to using a structured process, biases can be controlled by concentrating on plausibility rather than likelihood. That is, thinking about whether an assumption *could* fail to hold in a given way, not whether it is *likely* to do so and concentrating on *causal mechanisms* rather than likelihoods. Anything that could happen within the expected lifetime of the system should be accorded serious attention as a vulnerability. Finally, engaging in worst-case thinking can assist in deterring people from concentrating on the more likely but usually less severe consequences of events or ignoring cases completely due to confirmation bias.

Characteristics of a Good Leading Indicator Identification Process

In some organizations, the desire to predict the future leads to collecting a large amount of information based on the hope that something will be obtained that is useful. The NASA Space Shuttle program was collecting 600 metrics a month, for example, right before the loss of the Columbia, none of which turned out to be useful in predicting the loss or identifying the clear migration of the program to states of increasing risk.

A structured process, based on STAMP and STPA, may provide a more effective set of leading indicators. There are several goals for this process and for the resulting set of leading indicators:

- Complete: all critical assumptions leading to an accident are identified. Of course, no process is perfect, but that does not negate the goal of aiming for perfection. The author and her students have created specifications for complex systems that rigorously specified the underlying safety-related assumptions in the design. It is not infeasible. In addition, part of the design of the leading indicators (risk management) program should include a process for continual improvement and updating over time. Because completeness may mean that a very large set of leading indicators is identified, a process for determining what should be checked, how, and when will again be a critical part of the leading indicators program.
- Consistent: inconsistencies in the assumptions underlying the leading indicators need to be identified and handled. Inconsistency may indicate a flawed safety design process.
- Effective: the indicators should appropriately address the underlying assumptions, uncertainties, and vulnerabilities and accurately evaluate risk.
- Traceable: Each leading indicator and the action attached to it (see Section 4) should be identified as a response to one or more assumptions.
- Parsimonious: there should be no extraneous assumptions, checks, or actions that are not necessary to prevent accidents.
- Unbiased: The leading indicator process should minimize (combat) standard biases in risk assessment and management.

A Structured Approach to Identifying Leading Indicators

What needs to be identified can be derived from the reasons why accidents occur discussed earlier. For example, we want to identify assumptions made during design and development about the hazards and the hazard analysis, the effectiveness of the underlying mitigation and control measures, the behavior of the components during operations, and the behavior of the environment (system context) and then check those assumptions during operations to detect unexpected or unpredicted hazards and condition, changes in the assumed behavior of the system components and environment, etc. We also need to determine whether the control structure is working as designed and the state and effectiveness of the designed controls. For example, the investigation of many (most?) accidents, including the Bhopal, Macondo, and Texas City examples used in this paper, usually uncovers the fact that the requirements in the operations manuals are not being followed.

The STAMP/STPA process for safety-guided design and hazard analysis provides the framework for a structured leading indicator identification process. System hazards are first identified and used to derive the safety constraints and system safety requirements. Hazards are categorized, if necessary, with respect to potential worst-case severity and vulnerability. The functional safety control structure is designed with safety responsibilities identified for each component, where these control responsibilities are traceable to the system safety constraints. Once the safety control structure is created, STPA is used to identify unsafe control actions and their causes. An attempt to eliminate the causes is first attempted and, if elimination is not possible, to mitigate and control them.

During this process, the information necessary to identify the assumptions being made during design and development can be recorded and used to plan operations, to design the data and feedback that must be collected (see Section 3), and to design the overall leading indicator management program (Section 4). The relative importance of the leading indicators (the consequences of not detecting something) and potential action plans upon their failure is also determinable by the ranking of the hazards to which they are traceable.

The type of process being proposed here involves planning for the worst case instead of the predicted case. One drawback of considering vulnerability rather than likelihood is simply cost. One of the reasons for using likelihood early—even though the information to determine such likelihood is usually not available for new systems—is to avoid a lot of hazard analysis expense. There are two ways to tackle this problem. In practice, we have found that STPA is much cheaper and requires many fewer resources than the traditional hazard analysis techniques. So it may be feasible to examine more hazards in depth. In addition, if full analysis is not possible, then we can identify where some risk was allowed because we did not want to or could not eliminate or adequately control the hazard. This information can be used to create leading indicators to identify when those decisions were flawed. Finally, even without detailed analysis of all the potential causes, it is often possible to provide protection (a shaping or contingency action) against a hazard only knowing the hazard itself and not all its causes. The protection may not be as efficient as that which can be created with more causal information, but if the hazard truly is unlikely, the drawbacks of an inefficient or more costly response if it occurs may not be very important.

Dewar also cautions against implicitly assuming that if a worst case can be handled, then everything else will be handled too, that is, the assumption that everything else in a "lesser-included" case.[12]

3. Operationalizing a Leading Safety Indicator Monitoring Program

Ioannis Dokas has created a process called EWaSAP (Early Warning Sign Analysis using STPA) as an addition to STPA. EWaSAP adds steps to (a) define the data indicating the violation of safety

[12] Dewar, *op.cit*.

constraints and design assumptions within the system or process and (b) specify the capabilities and characteristics of the sensors, in order to be able to perceive these data.

Dokas operationalizes an early warning sign and defines it as the value of an observation provided by a sensor, which according to the process models and accident scenarios identified by STPA indicates the presence of causal factors for a potential loss or the violation of safety-related constraints and assumptions. He adds an additional type of control action, an *awareness action*, to the general control loop shown in Figure 4. An awareness control action allows a controller to provide a signal to other controllers inside or outside the system boundary whenever data indicating the presence of vulnerabilities have been perceived and comprehended. He also adds a set of additional guidewords to STPA that specifically relate to the transmission of early warning signals. Perceived signs and warning signals must be designed so they do not contribute to system hazards.

EWaSAP was used experimentally on a drinking water treatment plant to identify leading indicators and compare them to the ad hoc leading indicators already used by the operators and managers of the plant. The new process resulted in identifying 43 warning signs of which 37 were new and 6 already were used in everyday operations. Fourteen warning signs were deemed by management and staff to be of sufficient importance that they were incorporated into the existing safety management system at the plant.

The Dokas procedure focuses on physical leading indicators. It is also necessary to generate leading indicators to detect weaknesses and changes in the safety control structure and the safety culture. The same or similar approach to operationalizing a monitoring program may apply to the safety control structure.

Assumptions about the prevailing safety culture must also be monitored. Leading indicators can be generated from the corporate safety philosophy and safety policy and used as a basis for designing surveys to detect degradation of the process safety culture within the organization. While there are a lot of supposed culture surveys that have been suggested and used, investigating the Texas City accident required writing a new one because everything the committee could find focused only on occupational safety and did not apply to process safety. Hopefully things have improved since that time. The survey created by the committee was very general. It would be much better to identify leading indicators related to the specified safety policy of the company and use them to create a more tailored survey.

4. Managing a Leading Indicator Safety Program

Even if the right information is clear and available about elevated risk in the plant or organization, many companies are still slow to process these signals and respond. This paper so far has discussed how to identify leading indicators and operationalize them as shaping and hedging actions and warning signals. But having leading indicators will not help if they are not used or do not result in appropriate action being taken. As described in the section on psychological biases about risk, too often defensive avoidance is practiced and clear leading indicators are ignored until there is an accident.

To encourage effective action, leading indicators have to be integrated into the risk management program. Not only must they be communicated to appropriate decision makers, but detailed action plans for critical scenarios should be developed and triggers specified for implementing those action plans.

Some necessary steps in developing a leading indicator safety program include, obviously, identifying and documenting the safety-assumptions and designing leading indicators to identify weakening effectiveness of the controls over the safety constraints and ways to measure or identify the leading indicators. Responsibilities need to be assigned for checking them and for following through if problems are found.

Detailed action plans should be created *before* the assumptions are found to be invalid in order to lessen denial and avoidance behaviors and overcome organizational and cultural blinders. Responsibility for monitoring and action may need to be assigned to an independent organization and not to the project managers and those under conflicting stresses.

Periodically the list of leading indicators needs to be revisited and, if necessary, updated. A continuous improvement process should be created that both reevaluates the current indicators over time in the light of experience and diagnoses any identified lack of effectiveness. For example, if an accident or serious incident occurs that was not presaged by the leading indicators, an analysis should be made of why the leading indicators did not identify the problems in time to prevent them or, if they did, why effective action was not taken.

5. Next Steps

This paper has outlined a process for identifying and using leading indicators. A next reasonable step would be to evaluate its feasibility on a realistic system and then to determine, given that it is feasible, if it is useful.