LOPA and Human Reliability – Human Errors and Human IPLs

By: William Bridges
Process Improvement Institute, Inc. (PII)
1321 Waterside Lane
Knoxville, TN 37922 USA
+1-865-675-3458
wbridges@piii.com

Prepared for Presentation at
American Institute of Chemical Engineers
2010 Spring Meeting
6th Global Congress on Process Safety
San Antonio, Texas
March 22-24, 2010

Keywords: human factors, Layer of Protection Analysis, LOPA, process safety, PSM, human error, independent protection layer, IPL, SIS, SIL, safety instrumented system

Abstract

Layer of Protection Analysis (LOPA) is a simplified risk assessment method that provides an order of magnitude estimate of the risk of a potential accident scenario. Humans can be the cause on an accident scenario (the Initiating Event [IE]) or humans can serve or participate as an independent protection layer (IPL). In either case, estimating the likelihood of the human error and measuring the human error rate at a site are troublesome tasks within the LOPA framework, which is one of the reasons some companies do not give any credit for a human IPL.

Identifying and sustaining independent protection layers (IPLs) is the heart of LOPA. Each IPL must be:

- independent of the initiating event (IE) and the other IPLs
- capable (big enough, fast enough, strong enough, etc.)
- maintained or kept in practice/service
- validated/proven to provide the probability of failure (PFD) chosen

and all of the above must be documented and audited periodically to ensure compliance with these definitions.

There are many types of IPLs, and some are more trustworthy than others, hence the difference in the PFD of IPLs. As just mentioned, one possible type of IPL is a Human IPL. These include preventative steps that may stop a scenario from progressing once it is initiated, but more typically the human IPLs are responses to alerts or alarms or troubling readings and sample results.

This paper discusses the data needed for adequately counting the human in a LOPA (and other risk assessments), and includes discussion of the theory of human factors. One key focus of the paper is on practical means for collecting raw data in a plant setting for substantiating the error rates for the site, and especially for crediting a human IPL. The method for data collection covers the training requirements that should be met, proof drills for response to alarms, simulations and tests, and frequency of proofs, and of course the effect of human factors on human error rates. *Actual plant data and tests are included in the paper to provide the reader with some examples of how a simple data collection and validation method can be set up within their companies.*

Human Error Fundamentals

Human errors are sometimes mistakenly called procedural errors. This is not true anymore than saying all equipment errors are due to design errors. Over the past 5 decades of research and observation in the workplace on human error, we have come to know that human error probability depends on many factors. These factors (described in more detail in *Human Factors Missing from PSM*, LPS/GCPS 2010), include:

- Procedure accuracy and procedure clarity (the number one most sited root cause of accidents):
    - A procedure typically needs to be 95% or better accuracy to help reduce human error; humans tend to compensate for the remaining 5% of errors in a written procedure.
    - A procedure must clearly convey the information (there are about 25 rules for structuring procedures to accomplish this) and the procedure must be convenient to use.
    - Checklist features – These should be used and enforced either in the procedure or in a supplemental document

- Training, knowledge, and skills
    - Employees must be selected with the necessary skills before being hired or assigned to a department.
    - Initial Training – There must be effective, demonstration based training on each proactive task and each reactive (e.g., response to alarm) task.
    - Ongoing validation of human action is required and usually must be repeated (in either actual performance or in drills/practice) at least once per year *(as discussed later in this paper).* For human IPLs, the action must be demonstrated to be "fast enough" as well.
    - Documentation – the performance of the humans must be documented and retained to demonstrate the error rates chosen are valid.

- Fitness for Duty – Includes control of many sub-factors such as fatigue (a factor in a great many accidents), stress, illness and medications, and substance abuse.

- Workload management – Too little workload and the mind becomes bored and looks for distraction; too many tasks per hour can increase human error as well.

- Communication – Miscommunication (of an instruction or set of instructions or of the status of a process) is the second or third most common cause of human error in the workplace. There are proven management systems for controlling communication errors.

- Work environment – Factors to optimize include lighting, noise, temperature, humidity, ventilation, and distractions.

- Human System Interface – Factors to control include layout of equipment, displays, controls and their integration to displays, alarm nature and control of alarm overload, labeling, color-coding, fool-proofing measures, etc.

- Task complexity – Complexity of a task or job is proportional to the (1) number of choices available for making a wrong selection of similar items (such as number of similar switches, number of similar valves, number of similar size and shaped cans), (2) number of parallel tasks that may distract the worker from the task at hand (leading to either an initiating event or failure of a protection layer), (3) number of individuals involved in the task, and (4) judgment or calculation/interpolation, if required. For most chemical process environments, the complexity of the task is relatively low (one action per step), but for response actions (human IPLs) there are almost always other tasks underway when the out-of-bounds reading occurs or the alarm is activated.

In addition to the human factors listed, other considerations for use of a human as an IPL include (1) time available to perform the action and (2) physical capability to perform the action.

When using human error data for IEs and IPLs, the site must ensure that the factors above are consistently controlled over the long-term and that they are controlled to the same degree during the mode of operation that the LOPA covers. For instance, if the workers are fatigued following many extra hours of work in a two week period leading up to restart of a process, then the human error rates can increase by a factor of 10 times or 20 times during startup.

Revealed versus Unrevealed Errors for Human. As with equipment failures, human errors can lead to a revealed fault in the system (the flow does not start, for instance) or to an unrevealed fault (the block valve downstream of a control valve is left closed, but the failure is not revealed until the control is needed/used). If the error is revealed, then the error can be corrected or compensated for. If the restoration/correction time is sufficiently short, then the probability of being in the failed state is much less than for an unrevealed failure that is only discovered upon testing or inspection.

General Relationship of Human Factors Consideration to LOPA

Every risk assessment must consider the likelihood and effect of human factors. For LOPA, poor human factors can lead to higher human error rates that increase IE frequencies and that increase the PFD of a human IPL. The table on the next page summarizes the human factor issues that relate directly to LOPA. The table contrasts the impact of good and poor human factors on initiating event frequency (IEF) and on the PFD of independent protection layers (IPLs). If all of the factors were optimized, one would expect the IEF to be about 0.01 per task for IEs and about 0.05 for response to alarms or call for actions (human IPLs), but such an analysis must be made on a case-by-case basis, especially for the PFD of IPLs.

Considerations for Getting Low Human Error Rates

Issue	Requirement for claiming a low error rate for humans causing (initiating event) an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure for a human as a protection layer against an accident (usually a response to alarm)	Comments
Time to perform action	NA	Time to detect deviation, diagnose the problem, decide on proper action, and take action is recommended to be less than 1/2 the time required to reach the consequence of interest.	The reliability of the indication (sample & analysis in lab, field reading, etc.) or annunciation (alarm of the deviation) may limit the value for the human action IPL. Operator must be reliably available.
Capable	NA	It is physically possible to perform the control action required to forestall the consequence of interest, and the capability has been demonstrated and is not the limiting factor in the human action (e.g., the operator is strong enough to do the required task)	If the final control requires closing a manual block valve, the action and strength required must be verified in the field (some valves take several operators more than 1 hour to close, even with a valve wrench or cheater). This usually requires preventive maintenance or equipment use as part of a 'critical action.'
Procedure	Step must be in procedure. Not completing the step as stated and in that sequence (most aspects of errors of omission and commission) is the IE. The step is a critical step and is indicated as critical in the procedure or with a warning or caution. The procedure must follow "best practices" for control of human error; best practices relate to content accuracy, page format, and step presentation. See rules for procedures presented elsewhere.	Step must be in a trouble-shooting guide or similar contingency procedure (including emergency shutdown procedure), that is in paper form or reliably available on demand electronically (including via linkage to trouble-shooting guides within a distributed control system [DCS] for an alarm point) that describes how to respond to a process deviation. The response procedure follows best practices as well; see rules for procedures development and writing.	Note that we assume there is a 100% chance of skipping the required step or doing it substantially wrong if the step is not in the written procedure or if the step written is wrong or in the wrong sequence.
Checklist	Checklist in place and its use is enforced.	NA	Without a checklist, the error rate goes up by a factor of 3 to 5 times.

Issue	Requirement for claiming a low error rate for humans causing (initiating event) an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure for a human as a protection layer against an accident (usually a response to alarm)	Comments
Initial Training	There is initial training focused on this step as a critical step	There are initial training and drills on the response to a deviation; and the deviation is annunciated.	Note that we assume there is a 100% chance of skipping the required step or doing it substantially wrong if the step is not emphasized in initial training.
Ongoing validation (or practice) of human action	Steps must be practiced, but for seldom-used procedures, this practice (normally in the form of a talk-through for shutdown or startup of a continuous unit) is not until just prior to use. Practice rate increases reliability per action (due to increase in skill) but Actual Use Rate also increases the number of opportunities for failure. These factors can offset each other completely. See limitations on error rates that account for "practice" versus "error rate per task demand."	Steps must be practiced routinely enough to assure the reliability of the action under the increased stress caused by the alarm/indication. The time it takes to complete the action must be measured during each practice/drill, to ensure response time limits are not exceeded. Each person who is expected to implement the action "on demand" must perform the practice/drill per the assigned frequency. The minimum amount of practice or drills and type of practice should be at least one per year. (Note: One refinery practiced one scenario for a human IPL, selected at random, every shift and found no issues on shift workload and actual had improvement in general worker performance.) It is possible to group similar IPLs along the classification of similar types of actions and similar available response times. If such grouping is possible, then practicing one action will in effect be practicing all similar actions. For determination of sample size for tests/drills based on number of human IPLs and number of human responders, refer to standard methods for sample size determination.	Note that we assume there is a 100% chance of skipping the required step or doing it substantially wrong if the step is not practiced as often as specified here.
Documentation of validation of human action	Document each practice (i.e., each startup and each shutdown) and document failures rates for 'critical steps'. Keep the file of the validation associated with the task and also with each worker.	Document each practice (i.e., each response) and document failures rates for 'critical response action.' Keep the file of the validation associated with the 'action' and also with each worker. If a sample plan method is used, the same data as above is still required, and further, the data should be checked each quarter to ensure against inadvertently using the same IPL and same operator too often.	Note that we assume there is a 100% chance of skipping the required step or doing it substantially wrong if the step is not practiced as often as specified here or if there is No documentation of the completed refresher training or just-in-time training for an IE, or of drills for an IPL.

Issue	Requirement for claiming a low error rate for humans causing (initiating event) an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure for a human as a protection layer against an accident (usually a response to alarm)	Comments
Control of related physiological and psychological stressors	• Fatigue – Company policy and enforcement to limit fatigue, including limit of hours worked per day and per week, and including restrictions on hours worked outside of work. Can affect error rates by factor up to 20X. • Other fitness of duty issues – Company policy and enforcement to limit effect of alcohol & drug abuse, illness, prescription drug effects, and personal life stress effects. Can affect error rates by factor up to 50X. • Workload – Workload needs to be managed to optimize worker performance. Enough stress (time/task load-based), but not overstress. (Workload is also related to the complexity of the work. Complexity increases mental workload, even if the number of tasks per hour is not affected.) Workload evaluation may be necessary. Can affect error rates by a factor up to 10X. • Communication –The second most common cause of human error, the site must have a policy and rules and follow-through for controlling verbal and visual communication. Rules include repeat-back on instructions received and use of common jargon. Poor control of communication rules can increase human error rates by 10X. • Work environment – Temperature, lighting, noise, distractions, ventilation, etc., has been optimized to improvement worker performance. Can affect error rates by factor up to 5X. • Human-System Interface – Includes control-display-information integration, labeling, error-proofing designs, color-coding, alarm management, etc. Poor human-system interface control can increase human error rates by 10X or higher (especially if “norms” are violated, where a negative effect of 20-50X is assumed). • Complexity – Complexity of a task or job is proportional to the (1) number of choices available for making a wrong selection of similar items (such as number of similar switches, number of similar valves, number of similar size and shaped cans), (2) number of parallel tasks that may distract the worker from the task at hand (leading to either an initiating event or failure of a protection layer, (3) number of individuals involved, and (4) judgment or calculation/interpolation, if required. For most chemical process environments the complexity of the task is relatively low (one action per step), but for response actions there is almost always other tasks underway when the out-of-bounds reading or the alarm is activated. Complexity is difficult to predict (since it is not known when a human action will be needed), but higher complexity can increase error rates by 2X to 10X.		All of the listed human factors (beyond the procedure and training and skills listed above) have a large, independent influence on human error rates. A large negative swing in any one of the factors can increase human error rate by a factor of 3 to 50 times.

Issue	Requirement for claiming a low error rate for humans causing (initiating event) an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure for a human as a protection layer against an accident (usually a response to alarm)	Comments
Qualitative hazard evaluation	There has been a hazard evaluation of the scenario for skipping the step and doing it wrong and a hazard evaluation team has evaluated the consequences and safeguards.	The human action in this case is a safeguard that is called into play once a sequence of events is initiated. The human action (following the trouble-shooting guide or the emergency shutdown procedure) is to make a change of state in the system to prevent propagation of the event to the stated consequence. The hazard evaluation has evaluated the human action (and any annunciation device unavailability) in judging that this "critical responsive action" is capable, reliable, and audited.	It is essential that a company finds accident scenarios that can arise during all modes of operation, including startup, shutdown, emergency shutdown, and online maintenance.
Use of multiple successful human actions within a scenario	Making multiple mistakes (including shortcutting multiple steps) is the same probability as making a single mistake, if the "safeguard" steps amount to ensuring the same state in the system. The human who makes the errors sees this as performing the task by method B instead of method A. Method B could have different steps or less steps that method A.	If one person in a workgroup is assumed to make the mistake that is the initiating event, then it is assumed that no one in the workgroup can be part of the protection layers. Further, supervisors are also part of the workgroup, due to the trust/relationships that build quickly. Seemingly unrelated workgroups, such as maintenance and operations, usually build enough trust to rule out counting as separate groups. For a group/individual to be counted as a safeguard against the errors by another human error, the detection and response group must not share staff, must have an incentive to find mistakes and report/respond, and near miss reporting and auditing both indicate that independent measurement and action is occurring. An exception that is normally allowed is to re-use a workgroup or individual if the second action/response is separated by distance (location) and/or separated by enough time and/or there is a reminder alarm for the same alarm (the alarm that was missed earlier) that recurs often enough in the allowed time to still avoid the consequence.	It is nearly impossible to use a human response safeguard, if a cause of the predicted accident sequence is a human error.

TYPES of HUMAN ERRORS of IMPORTANCE IN LOPA

Human error comes in many forms, but the primary ones of interest in LOPA are:

- Errors in following proactive procedures, such as startup of a unit, that results in an initiating event of a LOPA accident scenario.
- Errors in responding to a call for action; if performed correctly such actions would interrupt the scenario and prevent the consequence.

The examples below illustrate both types of human error as they relate to LOPA.

Example of IE caused by Human error

In NUREG/CR-1278 – The Human Reliability Handbook (Dr. Alan Swain was the primary author), error rates were tabulated from various sources, including some empirical data. For errors that could be considered initiating events, these were summarized in Table 15-3 of that document, which is extracted and included in slightly amended form below:

Omission of Item	Human Error Probability
When procedures with check-off provisions are correctly used:	
(1) Short list, less than or equal to 10 items	.001
(2) Long list, greater than 10 items	.003
When procedures without check-off provisions are used, or when available check-off provisions are incorrectly used:	
(3) Short list, less than or equal to 10 items	.003
(4) Long list, greater than 10 items	.01
(5) When written procedures are available and should be used but are not used:	.05

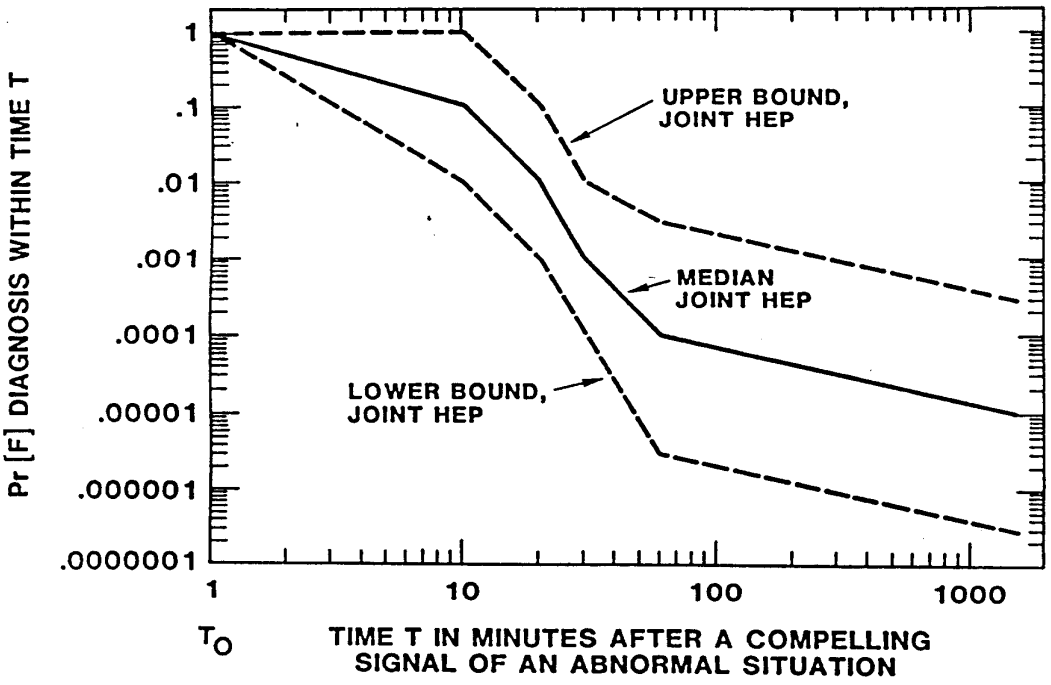
The table above shows the estimated probabilities of errors of omission per item of instruction when use of a written procedure for a task is required. The probabilities in the tables above assumes that human factors are optimized as well, which includes providing an accurate and easy to follow procedure, good control of fitness for duty factors (including controls against fatigue), good controls against miscommunication, training/practice (or actual performance) at least once per year, etc., as described earlier.

Since startup is when most accidents occur due to human error (when human error is the initiating event), it is appropriate to use that as an example. Startup procedures normally have more than 10 steps, indeed more than 50 steps is more typical. So, if all human factors are optimized, the probability of human error taken should be 0.003 per step. If practice with the procedure is only done once per year or less often, then the IEF would

be 0.005, which we would round up for Basic LOPA to 0.01. It may seem conservative to round up, but consider that there are few (if any) chemical plant sites that have optimized most of the human factors. Poor control of certain human factors can have a multiplying effect on the human error of 3 to 50 times. For this reason alone, it is dangerous to use the data above as representative of the error rates at a given site. Some analyst believe that in the absence of site data, these values can be applied, but perhaps it is best to scale the values upward (use higher base error rates) if the site determines that the human factors described earlier in this paper and in other resources are Not optimized.

Example human error related to human-based IPLs

Again, from NUREG/CR-1278 – The Human Reliability Handbook, estimates were provided for the probability of human error to respond correctly and in time following an alarm (annunciation). The Figure below is extracted from the handbook for response by control room personnel to an abnormal event. This represents the time it takes to diagnose the problem given there are No other alarms and given that all human factors are optimized, including practicing the response once per year or more often. The error probabilities given are the joint probability of failing to notice the alarm and failure to diagnose the situation correctly.



Note from the figure above that the more time to detect and diagnose the problem, the lower the chance of error. (HEP refers to Human Error Probability. Pr refers to the probability of making a diagnosis error.)

This figure only tells part of the story. After diagnosis, the operator, perhaps with help of others, must then respond appropriately to prevent the scenario from propagating to the next

protection layer or to the ultimate consequence. The sum of the time to detect, diagnose, and complete the proper response must happen within the available "process safety time" (PST). The process safety time is the interval from when initiation of the event begins until the time when no further action can prevent the consequence. The PST must be known for a site to judge if a specific human response can qualify as an IPL. Some organizations require that the operator's overall response must be accomplished in much less (normally half) of the PST, for the human response to be considered an IPL (given other criteria of an IPL is met; see LOPA 2001 for guidelines on qualifying an IPL). It is generally assumed that if the problem is detected (such as by hearing an audible alarm) and if the diagnosis is correct, then the action is likely to be "more" accurate than the detection/diagnosis (though this conclusion does not account for the speed of the response action). The norm therefore is to use a PFD of 0.1 if the response can be performed within the PST (or within ½ the PST for some organizations) and the PST is equal to or greater than 20 minutes (so the complete detect-diagnose-response steps are accomplished in 10 to 20 minutes, depending on the organization's rules for LOPA). A value of better than this (lower error rate) is normally not allowed in Basic LOPA. From an advanced topic perspective, however, Human Reliability Analysis (HRA) will often use a PFD of 0.01 for response times greater than 40 minutes (PST greater than 80 minutes). Performance of HRA and documenting the basis for lower probability values requires a level of effort and a degree of education/experience far beyond the required learning to perform Basic LOPA.

HUMAN ERRORS as related to IEs

Human errors that result in IEs are usually because of operator errors during startup (roughly 65% of all accidents in chemical-related process plants occur during startup) or errors during normal shutdown or emergency shutdown (together these represent 10% of the IEs caused by human errors). The remaining errors usually occur during normal operation and include such issues as performing a step wrong during operation in manual mode (such as when a process controller is bypassed) or by turning the wrong switch or opening the wrong valve.

To find these errors, a very good system for reporting and investigating near misses is needed (but that topic has been covered in other research and papers, see Bridges 1997 and 2008). The US Nuclear Regulatory Agency (NRC) in 10 CFR 55.45 and 55.59 requires that all power plant operators be tested once per year on abnormal procedures. This testing is mostly related to humans involved as IPLs. 10 CFR 55.45 and 55.59 also allude to the possibility of using a test of a representative sample of human responses, but we address this option later in the paper.

To help minimize human errors that could be IEs, many companies will review a normal shutdown or normal startup just before its next use. This has proven effective in controlling human error in chemical-related industries. Nuclear power plants also perform such "just-in-time (JIT)" refresher training on normal shutdowns and normal startup procedures.

If a site records the errors that are made during such tasks, then the company can validate the average IEF for that task and so the site will have a better estimate of the error probability. Further, if the measured rate is too high, they can take proactive steps to lower the error rate.

Errors can be measured either in actual practice in the field, or via simulators, or roughly by analysis of near misses reported (if the number of near misses reported is at least 20 times higher than the number of loss events or accident; Bridges 1997 and 2008; CCPS 2003).

Below is an example of data from one site that was collected to estimate the IEF for mistakes made during normal startup:

> ***EXAMPLE - Initiating event frequency data collection and use:*** In one company, the operating group was able to provide the data on the following page for errors made during startup of various units. Based on this data, the average probability of human error was 0.0071 and the number of errors per year that could lead to a known accident sequence was 0.17 (but note that if all errors are counted, the average error rate would be 0.29/yr). For basic LOPA, this value is rounded up to an IEF of 1/yr, for initiating events arising from these and similar procedures at the site. ***See the Table on the next page for the raw data.***

Chemical-related process plants should collect such data to understand and validate the initiating event frequency caused by human error. This data would then be used to validate the values for IEF used in LOPA and other risk assessments.

Site data on Human Errors related to Initiating Event Frequency (IEF)

Task for which a Potential Initiating Event exists	*Number of steps in the procedure*	*Number of Uses of the procedure*	*Number of Human Errors Noted*	*Number of Years of Data*	*Number of Errors That Initiated a Potential Accident Sequence*	*Human Error Probability per Step*	*Error Rate (probability of error per year)*
1. Startup of Process A	46	4	2	6	1	0.0109	0.17
2. Startup of Process B	71	5	3	6	1	0.0085	0.17
3. Startup of Process C	55	4	2	6	2	0.0091	0.33
4. Startup of Process D	56	4	0	6	0	0.0000	0.00
					Average:	0.0071	0.17
						IEF for LOPA for site:	**1/yr**

HUMAN ERRORS as related to HUMAN RESPONSE IPLs

A 10^{-1} PFD value for a human response indicates that the correct response occurs at least 9 out of 10 times (or no more than 1 wrong response for every 10 attempts). Most organizations will have identified many human responses involving a number of personnel, as part of their LOPA studies. Some organizations believe that if they have a procedure and a training program in place, that they can claim the PFD value of 0.1 for a Human IPL. This is no truer for a human IPL than it is for an active component IPL. A Human IPL must be tested and validated the same as a component IPL. On the next few pages are the options for validating human IPLs; these (1) include 100% testing for each human responder and for each action to be taken and (2) a sample plan testing of random combinations of human responders and actions.

Using a 100% Individual Test Plan Approach for Validation of Human IPLs.

One method to prove the operators will reliably respond for each Human IPL trigger is to have each operator demonstrate they can individually respond to each alarm (or other trigger). This response can be demonstrated by walkthroughs in the field or using simulators of the process unit. The nuclear power industry uses this approach for validating response by control room operators, but many in the chemical industry perceive this will take "too much time." As an example of the effort required, one nuclear power plant allocates about 200 hours per year per operator for refresher training activities, including 60 hours per year per operator for demonstration of skill in responding to critical alarms (from Tennessee Valley Authority [TVA] internal requirements to meet the performance-based requirements of NRC regulations 10 CFR 55.45 and 55.59). But is this time "too much"? We'll explore that answer in the rest of the paper.

Consider the following as an example of the application of this method of testing for 100% of the responses by 100% of the operators at an actual chemical plant:

> ***Background:*** The operating area being evaluated has 25 operators and 130 identified (from a hazards analysis and LOPA) human response IPLs.
>
> ***Validation Test***: The "tests" are documented on a set of index cards that call out various alarm conditions; these are the events (triggers) for a scenario identified in the hazards analysis (and LOPA) to be protected by the given human response IPL.
>
> ***Demonstrated Result:*** A correct answer (success of the IPL) is the desired response to the alarm scenario. A wrong answer (failure of the IPL) is any other response than the one desired or a response that takes too long.
>
> ***Estimation of Resources Requirements for 100% Testing Scheme:*** How much test effort is needed to ensure that training and retraining programs are sufficient to validate a 10^{-1} value for the PFD of all of these identified human response IPLs? An example estimate is as follows:

1. **Determine the number of test to be performed.** This would be the number of human response IPLs multiplied by the number of people who are expected to respond (perform as the human IPL) at some point in the future during their own shift. This example would yield 2750 (25 x 130 = 2750) discrete tests for one test of each combination of trigger and human responder (which makes up the 2750 IPLs).

2. **Determine the test frequency**. It is difficult to get consensus on this value. One documented example is from the nuclear industry. The US NRC regulation for control room operators (10 CFR 55.45 and 55.59) requires annual demonstration of proficiency in response to critical alarms and signals. This in fact would likely Not be enough testing to give a 90% chance of proper response to every alarm, except that response to one alarm is normally similar to response to other alarms, so the operator is in essence getting more **practice on similar alarms** as they perform each demonstration. Operators under this regimen of recertification have shown a 90% chance or better of proper response (detection, diagnosis, and action) within 10 minutes of an annunciation of an event of interest (from internal power plant records and also inferred from, Swain, 1983 and NUREG 1278 data, since the basic control room regimen of testing each operator with each action each year was essentially the same in the 1970s as they are now). For this example, a frequency of 1 validation per year is chosen.

3. **Determine the time required to perform each test.** Assuming 10 minutes of allowed response time for success (per alarm/trigger), then the test time for the organization would be 27,500 minutes or about 460 staff-hours (22 hours per operator per period). With a frequency (test period) of once per year per alarm, this equates to about 1% of the normal staff-hours for a worker in the USA. (Note: An equal amount of time would also be required to record/document the training record for the test [once the tracking system for such demonstrations is set up], but this is likely not a load the operator will have to bear.) An investment of 1% of each operator's time seems relatively inexpensive for the relative risk-reduction afforded by a Human IPL. (Note that testing only a sample of these responses, discussed later, would reduce the load considerably.)

EXPERIMENTAL DATA related to Human Response IPL.

Many companies have suggested that collecting data on human responses in order to validate that a human IPL is justified will be very difficult. However, the actual effort to collect such data is low and the benefits are great, as demonstrated by several chemical plants and refineries and nearly all nuclear power plants that have tried this.

To help demonstrate this, PII designed a simple test for use in measuring the response to an alarm condition. The test was not meant to measure the probability of detection of the alarm, but rather was meant to measure the time and success in determining and accomplishing the proper

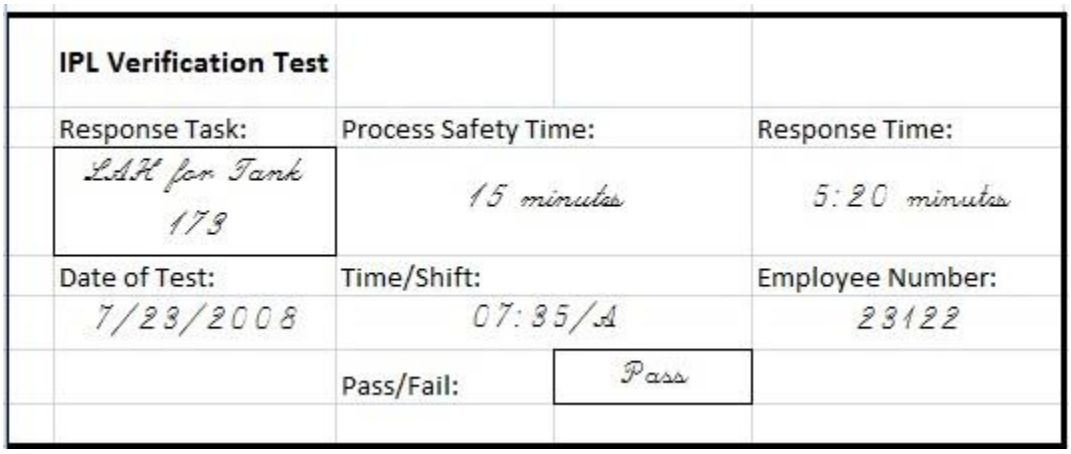
response to critical alarms as part of human IPLs. Two chemical plants belonging to large organizations (one in Malaysia and one in the USA) performed the test.

Test setup: The test involved having multiple operators in one unit of one plant/site perform responses to critical process alarms. These alarms were related to human IPLs. The actual response and time of response was measured, but essentially the tests were setup as a "pass or fail" – in other words, the test were to determine if the operators were able to respond as desired/expected.

Each test took 10-15 minutes to administer and less than 1 minute to record the data. The data was administered to multiple operators on multiple shifts. The tests were administered by a shift supervisor, a shift engineer, or in some cases, a process safety coordinator; though in the future, it is likely fine for a lead operator to administer the test. The overall effort to administer the test was deemed minimal, though there was resistance to such performance/validation test at first (as would be expected).

A human response IPL "failed" if the operator could not perform the required action to prevent the hypothetical outcome within the process safety time (definer earlier).

To run each test, the plants printed a data card (the size of an index card) and handed it to an operator chosen at random. Below is an example of such an index card.



IPL Verification Test		
Response Task:	Process Safety Time:	Response Time:
LAH for Tank 173	15 minutes	5:20 minutes
Date of Test:	Time/Shift:	Employee Number:
7/23/2008	07:35/A	23122
	Pass/Fail: Pass	

Note that the card contains an estimate of the PST – as defined earlier in this paper and elsewhere, this is the time an operator has to perform the task once the alarm is received until it is too late to take any further action. The time it took to print and hand out the index card was minimal. The person administering the test then timed the operator response and recorded the results. Any time less than the PST was a successful validation. (Again note that these tests did not validate the probability of an operator failing to detect an alarm.) The total time for the test varied, but the two sites that performed the test considered the administration time to be minimal; the largest time effort is simply for someone other than the operator to be there to "independently" measure the operator's response time (i.e., time to administer the test). Perhaps in the future, another operator could likely administer most proof tests and then the site

management could audit some percentage of the tests to help ensure against bias. If the operators test each other, then the time to administer a test is likely not significant enough to measure since they have to be "there" regardless for their other duties.

Aside: Currently the Safe Automation guidelines by CCPS and the upcoming guideline for IPLs and IEs (pending, 2010) both stipulate that the action must be completed within ½ of the PST. But, the ½ PST rule was arbitrarily chosen and has no statistical basis that the author is aware of.

For the most part, the tests came with little warning and occurred on all shifts. Several critical alarms were tested using various operators, all randomly selected. (It is anticipated that unless a sample plan is used, each operator will perform roughly one response related to each human IPL each year.) The time to respond was recorded on the index card.

Based on such raw data, the site was able to (1) evaluate the degree to which they were controlling human factors for the units, (2) identify which human responses qualify as IPLs, and (3) validate that the response is accurate enough and quick enough to qualify for the PFD used as the credit for the IPL (which for basic LOPA, the PFD is limited to a value of 0.1).

The table on the next page provides a sample of the site-specific data for several similar Human IPLs from the two sites (one in Malaysia and one in Texas, USA). For the Malaysia site, the data was from the same operating area consisting of multiple operators across 4 rotating shifts of 8 hours per shift and for the Texas site, the shift was 12 hours.

The sites believed that it is probably reasonable to use a sampling of human error for similarly qualified humans doing similar response or proactive tasks. A sample plan of perhaps only 5% to 10% of the number of human-task pairs may be necessary to have a valid statistic for human error for a "type of action." Sampling is valid for human actions because of how the mind processes information and actions for similar situations. Obviously, sampling can greatly reduce the measurement and documentation load for validation of human error rates. Human performance sampling is discussed in more detail next in this paper.

Human Error Data from Two Sites (USA and Malaysia) related to Independent Protection Layers (IPLs)

IPL No.	Response Task	Number of Test Performed	Average Response Time (minutes)	Process Safety Time (minutes)	Number Failures	PFD (average)	LOPA PFD
Company A (USA) Site 1							
IPL 1	ABP: High Temp in Generator	6	2.25	10	0	0	0.1
IPL 2	ABP: Loss of Acid Flow to Generator	12	2.166666667	10	0	0	0.1
Company B (Malaysia), Site 1							
IPL 1	LICA-HL-05007	10	5.054	15	0	0	0.1
IPL 2	TI 09010 low alarm on 09-K001 discharge	10	4.943	15	0	0	0.1
IPL 3	LAL-21004	10	6.962	15	0	0	0.1
IPL 4	LAH 22002	10	6.058	15	0	0	0.1
IPL 5	PIAH 21006	10	5.036	15	0	0	0.1
IPL 6	PIC-29014	10	5.352	15	0	0	0.1

Below are responses from the sites to questions related to the measurements:

- ***How much labor did each performance test take (on average)?***
    - The majority of the labor was with the actual operator going out to the field to 'take action' on the 'alarm'. Prep time was 1 hour for the initial recording of 10 data points for each of 2 to 6 IPLs, the prep time for each test was minimal as the cards were prepared beforehand and the recording takes less than 10 seconds.
    - The test time was about 15 minutes per operator per test.
    - Recording time was negligible.

- ***How much work do you think this would be longer-term (after getting in the habit and working any bugs out)?***
    - Not much difference in time as the testing process was picked up after the 4^{th} to 5^{th} data point in the unit, so it did not take long to get to "minimal" effort. One benefit of the test was the re-enforcement of "what to do" with all parties involved (the person doing the test, the person recording the test, and the other operators who noticed the test in progress).
    - The hard part is getting that up-front commitment; after doing a few tests, doing more was no issue.

- ***Do you feel such data is necessary for each human response IPL? (In other words, do you think a 5-10% sample of related type of responses would suffice?)***
    - The response for all the IPLs were the same when the same operator acts on different alarms or when different operators act on the same alarms. If this is the case, we probably can use a sample to test a "general" response to a type of alarm.
    - The site should first screen which response can be grouped together into general types of response IPLs. Then a lesser percentage will need individual validation drills.
    - Perhaps do just 1 or 2 drills per shift per group per year for the simpler ones on some periodic basis; that gives you a chance to test feasibility (make sure valves not locked, make sure valve wrench or other tools are available, etc).

- ***What improvements should be made to the tests/validation scheme?***
    - If we make the simulation a little bit sophisticated, we may be able to induce stress and surprise (maybe create some disorientation or carry out the tests in not-so-smooth situations – such as when Gene Hackman commenced a missile drill at the end of a minor crisis in the movie *Crimson Tide*) so that we can get closer to the actual critical situations. This would obviously need to be under very controlled conditions.
    - Perhaps with simulators or other methods, we could also include "detection and diagnosis" of the alarm in the same test.

- ***Who administered the test?***
    - Company 1. Plant production engineer
    - Company 2. As agreed upfront, about 80% was by the Shift Supervisor; 20% managed to be administered by the PSM Engineer or Shift Executive (Engineer) although more in an observer capacity.

- ***Who do you think should admin the tests in future?***
    - Same or a Lead Operator, which is basically an expert from the operator ranks.
    - Shift Supervisor level. In the Malaysia company, this is the highest level for non-degree holders in operations (or it can also be a newly graduated engineer).

- ***What unit was this in?***
    - This was a small batch plant in Texas
    - Ammonia Plant and Urea Plant in Malaysia

- ***Overall feeling of the exercise?***
    - The staff said it was easy to do and didn't have much difficulty in setting it up.
    - Initially, probably due to not fully seeing the whole picture, it was seen as a burden to do. Convincing was required to explain the potential benefits. But the plant staff rose to the occasion and the actual implementation was not so bad judging by what was able to be achieved even with the unplanned shutdowns that happen to occur in the month when the validation tests were run.

Using a Statistical Sample Plan Approach in the Future for Validation of Human IPLs.

It is important to ensure that a 10^{-1} value is indeed valid before using a human IPL. Testing every involved person on every identified response is not that burdensome, as shown in the test cases presented earlier that was run at two chemical plant sites. But, this burden could also be wasteful if testing the response to one alarm by one individual is essentially the same as testing the response of similarly trained and drilled individuals to the same or similar alarms.

Rather than testing 100% of the responses by 100% of the operators, it may be valid to use a sampling plan. This may be especially true for groups of responses that are similar in action and response time. US NRC alluded to a "sampling" of human response in 10 CFR 55.59, indicating that this may be acceptable for validating human response to triggers (i.e., for validating Human IPLs).

Statistical techniques developed decades ago are used to establish a basis of acceptance of all kinds of products, raw materials, components, etc. These methods can also be used to validate Human IPLs.

The sample plan approach must group similar type of actions and similar response time requirements. For a sample plan approach, choice of responder and trigger must be chosen at random. The lot size is the product of the number of responders multiplied by the number of similar response actions in the group.

As a rough rule of thumb, the sample should be about 10% of your total population of data, but not smaller than 30 and not greater than 350 to 500. The sample size and number of failures before the PFD is invalid is related to the confidence level and margin of error that is acceptable to the organization. A confidence level of 95% and an error margin of 5% indicate that the result

of your testing (validation of a Human IPL) will be within +/-5% of the true PFD 95% of the time the validation testing is performed.

The correct sample size is a function of those three elements – your universe (how many people multiplied by the number of actions; each pairing makes up a Human IPL), the desired error margin, and the preferred confidence level. For IPL validation purposes, it is likely reasonable to use a 5% error margin at 95% confidence. Below are typical sample sizes (the first at a 10% error margin, the second at 5%):

50 in the population, sample 33 or 44
100 in the population, sample 49 or 80
200 in the population, sample 65 or 132
500 in the population, sample 81 or 217
1000 in the population, sample 88 or 278

The trend above approaches a limit for a 10% error margin that hardly moves above 350 in the sample size no matter how large the population (and approaches a limit of 500 for a 5% error margin). The sample size can also be approximated using the equation below:

Sample Size, ss:

$$\text{ss (for infinite population)} = \frac{Z^2 * (p) * (1-p)}{c^2}$$

Where:

- Z = Z value (e.g., 1.96 for 97.5% confidence level)
- p = percentage picking a choice, expressed as decimal (0.9 used for Human IPL sample size determination)
- c = confidence interval, expressed as decimal (e.g., .05= ±5%)

This calculation then must be corrected to account for a finite population (validation SS):

$$\text{Validation SS} = \frac{Ss}{1 + \frac{ss - 1}{\text{population}}}$$

Meeting acceptance criteria for a human response means that the procedures, training, retraining, communication control, and all other human factors are achieving the desired result of no more than 1 wrong response in 10 demands. Rejection means that the PFD of 10^{-1} is not valid (or that the control of human error needs improvement for the PFD to remain valid). When applying this sampling plan to a validation of administrative IPLs, the "acceptance" criteria (the desired response) is the response that prevents the consequence that is being considered in the LOPA, e.g., *the operator response to the alarm prevents the overflow*.

Sampling plan schemes rely on developing a valid definition of a group to be the sample. For Human IPLs, the group is a combination of similar operators (likely all operators are treated equally for this purpose) and similar responses to similar triggers. The latter grouping takes careful consideration by a multiple-disciplinary team (with heavy emphasis on the operators on the team composition), to ensure the grouping of IPLs makes sense (i.e., use expert judgment). Although it is likely that all operators can be lumped into the same statistical group (if they are selected at random for validation drills of IPLs), the triggers and responses will need to be grouped to ensure that validation of one trigger/response is essentially the same as validating other triggers/response within the same group.

The upcoming guideline for IPLs and IEs (CCPS, 2010, pending) provides guidance on choice of sample size and on grouping of alarms and workers who respond to them. Included in the discussion is a method derived from ANSI Z1.4 (now incorporated by reference with US MIL standards for lot sampling). Examples calculations and the resulting savings achievable by use of sample plans are also provided in the guideline.

The table on the next pages provides an example of the relative savings possible by using a sampling of human response IPLs instead of testing each operator for each IPL.

Sample Size and Acceptance Criteria Estimation for Human IPL Validation

Employee	Alarms	pop = actual size of population	ss = sample size based on infinite population	Z value	Confidence Level	p = anticipated PFD	c = confidence interval	Validation SS = sample size adjusted for actual population	% of actual population sampled	Expected average number of failures per test period using the sample size	Maximum allowed failures to have XX% confidence (95% if Z= 1.96) that the entire population has less than 1-p error rate (at the chosen confidence interval)	From Hypothesis Test, for 95% confidence level and approximate PFD of 0.1 for entire population
	Double-tailed Z test						"+/-"					
25	130	3250	138	1.96	95%	0.9	0.05	133	4	13.3	12	8
25	13	325	138	1.96	95%	0.9	0.05	97	30	9.7	9	6
25	130	3250	195	2.327	97.5%	0.9	0.05	184	6	18.4	17	11
25	13	325	195	2.327	97.5%	0.9	0.05	122	38	12.2	11	7
	Double-tailed Z test						"+/-"					
25	130	3250	54	1.96	95%	0.9	0.08	53	2	5.3	5	2
25	13	325	54	1.96	95%	0.9	0.08	46	14	4.6	4	2
25	130	3250	76	2.327	97.5%	0.9	0.08	74	2	7.4	7	3
25	13	325	76	2.327	97.5%	0.9	0.08	62	19	6.2	6	2

HUMAN ERRORS as related to Impact on non-human IEs and IPLs (such as SIS and Relief Systems)

In addition to having a direct impact as an IE or as failure of a human IPL, human error can of course impact all other IEs and IPLs as well. This is because ALL safeguards are ultimately controlled and maintained by humans. Typically, a base human error rate of 1 error in 50 steps is reasonable, or for excellent control of human factors, an error rate of 1 error in 100 steps may be achieved. Below are some examples of how human factors control the reliability of all IPLs and IEs:

Example related to IE – wrong materials of construction received: If the humans fail to control or detect errors during selection and use of materials of construction, then perhaps the wrong grade of steel or other material will be used in a highly corrosive environment and then the IEF will be many times higher than expected. Industry data indicates that the materials delivered are different that what we specify about 3 to 7% of the time. Positive material identification (PMI) detect and correct such failures, theoretically, but plant data suggests that using 100% PMI at receiving will lower the composite error rate to 1% and that another 100% PMI of "as-welded in the field" lowers the combined error rate to about 0.3%. This still leaves a large number of components that may be of the wrong materials.

Example related to IPL – high integrity SIS: Safety Instrumented Systems (SIS), can be designed to provide risk reductions of 1, 2, or 3 orders of magnitude. However, even the best designed SIS can be negated if (1) the root valves for the sensors are left closed or (2) a bypass valve around an emergency isolation valve is left open (other errors are also possible of course). The current versions of SIS standards do not explicitly require validation of the control of human factors and the resulting human errors that can leave the valves in the wrong position. (In addition, the humans are always capable of revaluing trip points, etc., in the safety-rated PLC, given enough time to learn these shortcuts and given enough spurious trips to make them want to use such a shortcut.) Although an SIL 4 SIS is allowed by IEC 61508, with a PFD of $<10^{-4}$ to $\geq 10^{-5}$, looking at it pragmatically, we do not believe it is possible to control the human errors associated with in situ testing low enough to maintain a SIL 4 rating. Similar, due to the complexity of the analysis of human errors associated with in situ testing, basic LOPA only allows use of SIL 1. SIL 2 and SIL 3 rating is difficult to validate statistically, due to the consideration of the human factor, as a result, SIL of 2 and 3 require human reliability event tree or similar methods to quantify, and so is considered beyond basic LOPA.

Example related to IPL – pressure relief systems: Many relief system IPLs can have a PFD of 0.01 or better, assuming the rest of the process does not interfere or restrict the flow. However, if you have a block valve upstream or downstream of the a relief device, then leaving it closed will negate the entire value of the relief system. A human reliability analysis of existing plant practices is needed to establish the PFD is better than 0.1, and excellent sustained control of human factors is necessary to maintain a PFD of .01 or better for a relief system; because of the relatively high probability that a

human will leave a block valve closed upstream or downstream of a relief valve. Common industry data for human error rates indicates that the error of leaving such a block valve closed will likely be the dominating failure mode for the relief system, so including block valve upstream or downstream of a relief device (though allowed by ASME with administrative controls) will need to be evaluated by a site before assigning a PFD for the combination system (comprised of the relief system and block valves). The conservative PFD of 0.1 is therefore used as the default value for relief devices that have blocks valves upstream or downstream of relief devices, unless a human reliability for the plant proves otherwise.

Conclusion

Control of human factors is key to achieving high reliability of processes. Poor human factors lead to higher initiating event rates and less or no value for IPLs. Proper control of human factors can be achieved by following industry best practices (not minimal regulations and codes/standards).

Failing to control human factors will result in the time invested in LOPA to be wasted, and more importantly will lead to more accidents. Simply installing or claiming IPLs will not change the critical dependency all plants face on controlling risk – risk control always depends on human factor control.

References

1. Tew, R. and Bridges, W., *Human Factors Missing from PSM*, LPS/GCPS, 2010.

2. US Nuclear Regulatory Commission, *Training Requirements for Nuclear Power Plant Operators*, 10 CFR 55.45 and 55.59

3. NUREG/CR-1278 – *The Human Reliability Handbook*; guidelines from the US NRC on Human Reliability Analysis.

4. *Layer of Protection Analysis (LOPA) Guideline*, AIChE/CCPS, 2001.

5. *Guidelines for Independent Protection Layers and Initiating Events*, AIChE/CCPS, 2010 [pending].

6. *Guidelines for Investigating Chemical Process Incidents, 2nd Edition,* CCPS/AICHE, 2000.

7. Bridges, W., *Gains in Getting Near Misses Reported,* 8th Conference, ASSE-MEC, Bahrain, 2008.