LOPA and Human Reliability – Human Errors and Human IPLs (Updated)

William Bridges
Process Improvement Institute, Inc. (PII)
1321 Waterside Lane, Knoxville, TN 37922, USA
wbridges@piii.com

Dr. Tony Clark
Process Improvement Institute, Inc. (PII)

Prepared for Presentation at
American Institute of Chemical Engineers
2011 Spring Meeting
7th Global Congress on Process Safety
Chicago, Illinois
March 13-16, 2011

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.[1] 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 (based on a similar paper from 2010[2]) 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.** This paper also provides an overview of an alternative method for estimating the PFD of a Human IPL, based on plant and scenario specific factors (such as stress factors, complexity, and communication factors).

1. Human Error Fundamentals

Human errors are sometimes mistakenly called procedural errors. This is not true anymore than saying all equipment failures 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 are described in more detail elsewhere[3] but they include:

- Procedure accuracy and procedure clarity (the number one most sited root cause of accidents):

  - A procedure typically needs to be 95% or more accurate 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 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.
  - For human response IPLs, a checklist is not always practical; in fact a trouble-shooting guide (rather than a prescriptive, linear procedure) is usually required since we do not always know what condition a process will be in before a serious problem begins that requires humans to respond (as a human-based IPL)

- 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. For human IPLs, the action must be demonstrated to be "fast enough" as well. For the Site-Specific Data approach, validation can be done in actual performance (measuring errors to responses to actual alarms) or in drills/practice. Or the human performance can be validated by Expert Judgment, Generic data, or by Prediction. Appendix B and C provide examples of two methods for validating human IPLs (and estimating human IEFs).
  - Documentation – the actual performance of the humans or alternative validation approaches 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, 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 one of the most common causes 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, and as mentioned earlier, we cannot predict what state the rest of the process will be in when then alarm comes on.

In addition to these human factors, 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 considering human IEs and IPLs, the site must ensure that the human factors just listed are consistently controlled over the long-term and that they are controlled to an adequate degree during the mode of operation for the particularly LOPA scenario. 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 or 20 times during startup and so a LOPA scenario for startup would need to take these into consideration.

Revealed versus Unrevealed Errors for Human. As with equipment failures, human errors can lead to a revealed fault in the system (e.g., the flow does not start) or to an unrevealed fault, such as the block valve downstream of a control valve is left closed, but the failure is not revealed until the control valve is 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 for a revealed failure than for an unrevealed failure, which is only discovered upon testing or inspection.

2. General Relationship of Human Factors 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 (and of other IPLs indirectly). Table 1 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 human-based independent protection layers (IPLs).

Table 1. Considerations for Getting Low Human Error Rates

Issue	Requirement for claiming a low error rate for humans causing an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure on demand for a human as a protection layer against an accident (usually a response to alarm)	Comments
Time to perform action	NA	The total time to: detect deviation, diagnose the problem, decide on proper action, and take action must be less than the time required to reach the consequence of interest or MART. The term "maximum allowable response time" (MART) is used throughout this guideline.	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 to quickly respond.
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. Follow the current rules for procedure development and writing.[3]	Step should 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; follow the current rules for procedure development and writing. [3]	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 an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure on demand 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) can be 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. Per CCPS, *Guidelines for IPLs and IEs (2011)* [4], there are limitations on error rates that account for "practice" versus "error rate per task demand" that are incorporated within the Human IEs.	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 validated, to ensure response time limits are not exceeded. For the Site-Specific Data approach to validation, each person who is expected to implement the action "on demand" must perform the practice/drill per the assigned frequency of about once per year. Per the Predicted Data approach to validation (see the brief example at this end of this paper for an example), the PFD for the human IPL is estimated by calculations using factors based on the quality of the underlying management systems, procedures, and training; and based on the level of stress involved with the response task. Other methods of validation may also be used, as described in the upcoming textbook from CCPS, *Guidelines for IPLs and IEs.* [4] It is normally possible to group similar IPLs along the classification of similar types of actions and similar available response times. When grouping is possible, then validating one action will in effect be validating all similar actions. This paper provides an example of grouping, if validating using Site-Specific drills of response actions.	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 often enough (to be determined by the site).

Issue	Requirement for claiming a low error rate for humans causing an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure on demand for a human as a protection layer against an accident (usually a response to alarm)	Comments
Documentation of validation of human action	Documentation methods depend on the validation method chosen. Validation based on Expert Judgment or Predicted Data approaches would require the site to maintain sufficient documentation, including calculations and description of methods, for the individuals IEs or groupings of IEs. The Site-Specific Data approach to validation requires more effort since the site would need to document each practice (i.e., each startup and each shutdown) and document failures rates for 'critical steps'. For this same validation method, the site would keep the file of the validation associated with the task and also with each worker. Though more time consuming on documentation, this method may produce more side benefits that out weight this cost.	Documentation methods depend on the validation method chosen. Validation based on Expert Judgment or Predicted Data approaches would require the site to maintain sufficient documentation, including calculations and description of methods, for the individuals alarm response actions. For the Site-Specific Data approach to validation, the site would need to document each practice (i.e., each response) and document failures rates for 'critical response action.' The site would also 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 necessary (to be determined by the site) 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 (if drills are part of the validation method).

Issue	Requirement for claiming a low error rate for humans causing an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure on demand for a human as a protection layer against an accident (usually a response to alarm)	Comments
Control of related physiological and psychological stressors (see SPAR-H, NRC, 2007)	• 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 guard against miscommunication by having 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 an accident (usually skipping a step or doing one wrong)	Requirement for using a low probability of failure on demand 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 difficult to use a human response safeguard, if a cause of the predicted accident sequence is a human error. A human reliability analysis (HRA) is typically done to verify the human protection layer probability in this situation.

If all of the factors were optimized, one would expect about 0.01 per task as an IEF and about 0.05 PFD (which would normally be rounded up to 0.1) 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.

3. Types of Human Errors of Importance in LOPA

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

Human errors that directly relate to human IEs and human IPLs

- 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

Human errors that indirectly relate to LOPA component-based IEs and IPLs

- Errors when following management systems that lead to higher failure rates of components

The examples below illustrate both types of human error that directly affect LOPA. (Human errors that indirectly effect component type of IEs and IPLs are discussed near the end of this paper.)

3.1 IE Caused by Human Error

In NUREG/CR-1278[5]– 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 have been extracted and included in slightly amended form below as Table 2 below.

Table 2. Errors of Omission versus Error Probability[5]

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

This table 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 (for a continuous process) is when most accidents occur (50% of major accidents in the chemical-related industry occurs during startup mode of operation) [6, 7, 8,] and since the causes are typically due to human error, it is appropriate to use that as an example. If all human factors are optimized, the probability of human error taken could be 0.003 per step, which we would round up for Basic LOPA to 0.01 errors per step. It may appear then that 0.01 is a conservative value for human error, but consider that there are few (if any) chemical plant sites that have optimized most of the human factors. And poor control of certain human factors can have a multiplying effect on the human error of 3 to 50 times, so the real error rate for use of a startup procedure after a maintenance turnaround could be as high as 0.2 (as was approximately the case during the BP Texas City Accident in March, 2005). ***It is dangerous to use the value of 0.01 errors per step per opportunity as representative of the error rates at a given site without site-specific data.*** 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 are ***not*** optimized. This concept is described at the end of the paper.

3.2 Human Error Related to Human-Based IPLs

Again, from The Human Reliability Handbook [5], estimates were provided for the probability of the human to respond correctly and in time following an alarm (annunciation). Figure 1 below (extracted from the handbook) is 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 that this study was done for nuclear plant control rooms that always have several operators available to potentially respond. Many chemical plants do not have this level of control room staffing.

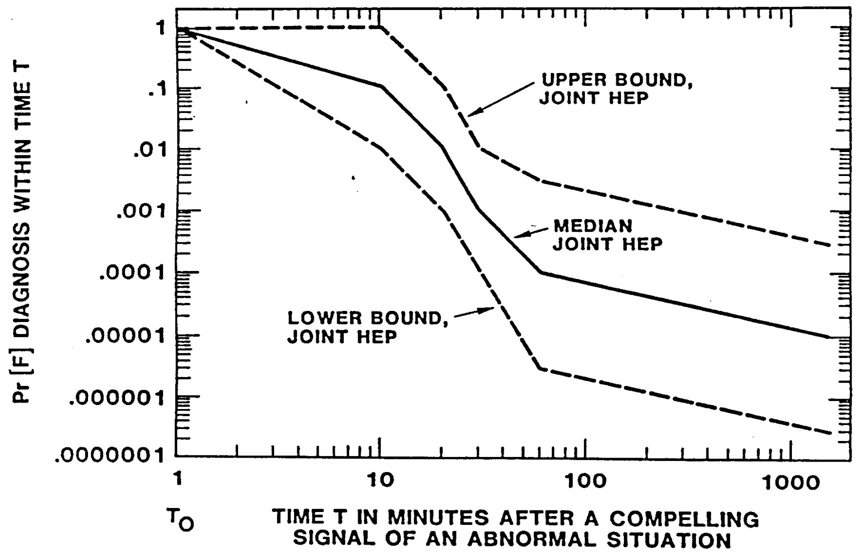


Figure 1. Errors Probability versus Time Available to Respond[5]
NOTE: HEP refers to Human Error Probability. Pr refers to the probability of making a diagnosis error.

Note from Figure 1 that the more time to detect and diagnose the problem, the lower the chance of error. But Figure 1 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, annunciate, diagnose, and complete the proper response must happen within the time available, is best termed "maximum allowable response time (MART)," which is used in this paper (and in the upcoming CCPS text, *Guidelines for IPLs and IEs (2011)*[4].

MART is the maximum duration for a given IPL to detect an out of limit condition, fully execute its safety function, and bring the process to a safe state. It is a property (constraint) of many IPLs which the LOPA team imposes based on knowledge of the scenario timeline and feeds directly into setpoint selection for ISF (and other active IPLs) and checking frequency for proactive human actions such as routine rounds. It is useful where time is a critical element of IPL design (i.e., applies to active but not passive IPLs). The MART must be known for a site to judge if a specific human response can qualify as an IPL (given other criteria of an IPL is met; see LOPA 2001[1] for guidelines on qualifying an IPL).

When an operator on rounds observes a failure that he/she is specifically looking for (e.g., small leak in a pump seal or flange) he/she must assure the leak is repaired. The LOPA team specifies this MART based on the anticipated progression of the failure. MART would be significant for an analytical test which detects that the process is leading to (but has not yet achieved) an unsafe state (e.g., wrong composition feeding a post reaction tank based on low conversion where there is still time before the tank overheats). For other analytical tests such as identity check before offloading a railcar the procedure step order is critical and time is not (thus MART doesn't apply).

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, as long as there is time to fully complete the response action. The norm therefore is to use a PFD of 0.1 if the response can be performed within the MART and if the MART is equal to or greater than 20 minutes (so the complete detect-diagnose-response steps are accomplished in 20 minutes. Note that some organizations use a lower limit of 10 minutes or even 5 minutes for MART for Human IPLs if the detect-diagnose-response steps are all accomplished in a control room. 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 MARTs greater than 40 minutes but care must be taken to assure that controls used for annunciation or response are appropriately improved to SIL integrity. 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[4].

4. Human Errors as Related to IEs

Human errors that result in IEs of major process incidents are usually occur during startup; one definitive study (Rasmussen, 1989)[7] and a survey of the major accidents in the past 20 years[8] show that roughly 65% of **major** process accidents in chemical-related process plants occur during startup and on-line maintenance. Further, errors during normal shutdown or emergency shutdown together represent about 10% of the IEs caused by human errors. (Other studies show that most **minor releases**, such as packing leaks, occur during normal operation.) Human errors during normal operation result in less than 25% of the and include such issues as performing a step wrong during operation in manual mode, such as when a process controller is bypassed or by external impact (one major accident during normal operation occurred when a fork-truck hit and broke off a drain valve).

The two primary ways to find accident scenarios caused by human error (before the accident occurs) are:

- Prediction of the accident scenario in PHA of all modes of operation, especially startup and shut down procedures. Most PHA/HAZOPs world-wide are weak on finding accident scenarios during non-routine modes of operation. Overcoming this weakness is critical and related methods are explained in Chapter 9.1 of the *Guidelines for Hazard Evaluation Procedures, 3rd Edition* [9] and in several papers [10, 8].
- Getting a lot of near misses reported and investigated; usually having at least 15 times the number of near misses reported as accidents occurring is needed to have a 90% chance or higher of preventing the major accidents [11, 12]. The ratio of near misses report to accidents occurring is relatively low at most companies world-wide, so this is low hanging fruit for many companies.

For LOPA, the ways to estimate and validate human IEFs (and validate PFDs of IPLs) include expert opinion, using industry values, calculation from human reliability factors, or direct measurement at the site. This paper provides examples of how to validate an IPL by Predicted Data (a brief example at the end of the paper) from human reliability factors and by using direct measurement at the Site-Specific Data, respectively.

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 one effective way to reduce 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. They can also be roughly estimated by analysis of near misses reported (if the number of near misses reported is at least 15 times higher than the number of loss events or accident[11, 12].

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 Table 3 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.

5. Validation of Human IEs and IPLs by Site-Specific Data

A 0.1 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-based IPL.

As required for all IPLs, a human IPL must be validated. The Preferred approach to validation is direct measurement or testing of the human response (under controlled conditions or drills); but other methods of validation can include Expert Judgment, using data from other comparable settings (Generic Data method), and estimation of the PFD of human IPLs by mathematical modeling (Predicted Data method, see an example of this method later in this paper).

On the next few pages are the options for validating human IPLs with **site-specific data**. These include:

- 100% testing for each human responder and for each action to be taken
- Sample plan testing of random combinations of human responders and actions.

One key focus of this paper is discussion of 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 this paper to provide the reader with some examples of how a simple data collection and validation method can be set up within their companies.*

Table 3. 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**

As mentioned earlier, if a site has a very good system for reporting and investigating near misses, then this can be used to find site specific data for human errors (including both IEs and failure of human IPLs). Another way to collect site-specific data on error rates is to **measure them with tests or drills of the action**; and for human IPLs discussed later, the results may need to be adjusted to account for actual stress levels of a response. This is not commonplace in the chemical industry, but the US Nuclear Regulatory Agency (NRC) in 10 CFR 55.45 and 55.59[13] 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 this section of the paper.

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 think 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). This equals about 3% of the work-year for an operator. (The example below and also the example of the sample-plan approach discussed later shows how this investment of time to measure response effectiveness can be reduced by a factor of 10 or more.)

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

> ***Background:*** The operating area being evaluated has 20 operators (spread across 4, rotating shifts of work) 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 (longer than the MART, defined earlier).
>
> ***Estimation of Resources Requirements for 100% Testing Scheme:*** Below is an estimate of how much test effort is needed to ensure that training and retraining programs are sufficient to validate a 0.1 value for the PFD of all of these identified human response IPLs:

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 2600 (20 x 130 = 2600) discrete tests for one test of each combination of trigger and human responder (which makes up the 2600 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. On the surface, testing once per year per alarm may seem unlikely to give a 90% chance of proper response to every alarm, except for the fact 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 TVA tests, but also inferred from, nuclear power plant data[5], 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 26,000 minutes or about 430 staff-hours (22 hours per operator per period; most likely the period would be once per year). 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. Also note that testing only a sample of these responses, discussed later, would reduce the load considerably.)

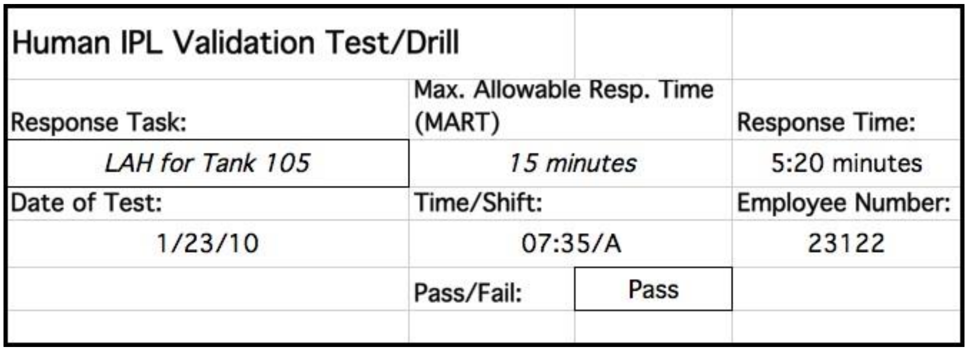
5.1 Actual Example of Using Tests/Drills (site-specific data collection) to Validate Human Response IPLs

Up until now, the actual effort to collect such data has not been well documented in the literature, though many chemical companies, refineries, and nuclear power plants do validate human response using this method. Recent research by three chemical companies (one in Malaysia and one in Texas, USA and one in Canada) documented the effort required to validate human response IPLs using Site-Specific Data. The following is an excerpt of the research results.

5.5.1 Validation Setup:

A simple test was 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. Three chemical plants belonging to large organizations performed the test.

The test involved having multiple operators in one unit of one plant/site (one for each company) 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, within the allotted MART. 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. Figure 2 is an example of such an index card.



Human IPL Validation Test/Drill			
Response Task:	Max. Allowable Resp. Time (MART)		Response Time:
LAH for Tank 105	*15 minutes*		5:20 minutes
Date of Test:	Time/Shift:		Employee Number:
1/23/10	07:35/A		23122
	Pass/Fail:	Pass	

Figure 2: Example of card used to administer validation of a single human IPL

Note that the card contains an estimate of the MART – as defined earlier in this appendix and elsewhere in this guide; 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.

5.1.2 Validating/Testing:

A human response IPL "failed" if the operator could not perform the required action to prevent the hypothetical outcome within the MART (defined earlier). The person administering the test timed the operator response and recorded the results. (Again note that these tests did not validate the probability of an operator failing to detect an alarm.) Each test took 10-15 minutes to administer and less than 1 minute to record the data. The validation was performed by multiple operators on multiple shifts for randomly selected alarms. The tests were administered by a shift supervisor, a shift engineer, or in some cases, a process safety coordinator. It is likely another operator could administer most proof tests (validations) 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. The total time for the test varied, but the sites that performed the test considered the time to administer the test to be minimal; the largest effort was simply for someone other than the operator to be there to “independently” measure the operator’s response time.

For the most part, the tests came with little warning and occurred on all shifts. (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 and later transferred to a summary spreadsheet

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). Table 4 provides a sample of the site-specific data for several similar Human IPLs from the three sites.

For the Malaysia and Canada sites, 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.

All of the IPLs passed (all operators performed each action correctly within the allotted time) during these tests/drills. The labor to perform the test took less than 15 minutes per test (including documentation time). After the initial resistance at each site to performing the test, the subsequent tests were not resisted and in fact the operations staff embraced the testing/drills since they saw many side benefits from the test/drills, including 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). Lead operators and supervisors administered the test; very little training or coaching was necessary to have the drills done properly.

All three sites believe it is possible to use a sampling of human error for similarly qualified humans doing similar response or proactive tasks. (This is because the responses for all IPLs were the same when the same operator acts on different alarms or when different operators act on the same alarms.) 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 how human take the necessary actions for similar situations. Obviously, sampling can greatly reduce the measurement and documentation load for validation of human error rates. If sampling is used, the sites suggested that:

- 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.

Table 4. Site-Specific Validation of Human Response IPLs[a]

IPL No.	Response Task	Number of Test Performed	Average Response Time (minutes)	Maximum Average Response Time (minutes)	Number Failures	PFD (average)	LOPA PFD
Company A (USA), Site 1							
IPL 1	ABP: High Temp in Generator	6	2.3	10	0	0	0.1
IPL 2	ABP: Loss of Acid Flow to Generator	12	2.2	10	0	0	0.1
Company B (Canada), Site 1							
IPL 1	Low Seal Gas Pressure to Turbo-Exchanger Bearings	5	5.7	15	0	0	0.1
IPL 2	High Lube Oil Temperature – XX Compressor	5	6.3	30	0	0	0.1
IPL 3	Low Level Emergency Alarm -- Steam Drum	5	4.9	15	0	0	0.1
IPL 4	High-High Lube Oil Temperature – XX Compressor	5	6.1	30	0	0	0.1
Company C (Malaysia), Site 1							
IPL 1	High Level on Ammonia Absorber Column	10	5.1	15	0	0	0.1
IPL 2	Low Temperature Alarm on Ammonia Compressor Discharge	10	4.9	15	0	0	0.1
IPL 3	Low Level in CO_2 Compressor Interstage KO drum	10	7.0	15	0	0	0.1
IPL 4	High Level on High Pressure Carbamate Heat Exchanger	10	6.1	15	0	0	0.1
IPL 5	High Pressure High Pressure Carbamate Condenser	10	5.0	15	0	0	0.1
IPL 6	High Pressure on Steam Controller to Rectifying Column Recirculation Heater	10	5.4	15	0	0	0.1

[a] Data provided by 3 companies (in the USA, Canada, and Malaysia)

- Perhaps do just 1 or 2 drills per shift per group per year for the simpler ones on some periodic basis; that gives a chance to test feasibility (make sure valves are not locked, make sure a valve wrench or other tools are available, etc.).

Human performance sampling is discussed in more detail later in this paper.

5*.1.3 Adjustment of Results for Stress*

As mentioned, this data (as with all drills) is collected during a simulation of a call for action. In a real event, ***the stress to perform the task correctly would increase the average error rate***. NRC estimates[14] the stress for this type of pre-emergency response action (versus emergency response and evacuation) will likely not be "extreme" but it will be "high," in which case a conservative estimate is that error rates would double from the test/drill case. It is likely not possible to get a drill that accurately mimics the stress of a real alarm event, so there will likely always be a need to adjust data for increases in errors due to stress. It is likely more appropriate to double the observed error rates (observed PFDs) rather than doubling the observed response time. But in either case, the IPL data collected above must still "pass" when adjusted for stress for an IPL to be validated using Site-Specific Data.

5.1.4 Comments from Site Personnel on Collection of Site-Specific Data

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

- ***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 4th to 5th 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?***
    - This varied across the three sites, but were either plant production engineer, shift supervisor (mostly), PSM Engineer, or Shift Executive (Engineer) although these latter two were 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 Malaysian 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, and a specialty continuous chemical plant in Alberta Canada.

- ***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.

5.2 Approach to Using a Statistical Sample Plan for Validation of Human IPLs.

It is important to ensure that a 0.1 value is indeed valid before using a human IPL. Rather than testing 100% of the responses by 100% of the operators, it is normally valid to use a sampling plan. This is 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. (See Walpole, 2006, and other standard textbooks for typical approach on statistics and sampling.) 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% to 5% of your total population of data, but not smaller than 30 samples and not greater than 350 to 500 samples. 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 that hardly moves above a 350 sample size no matter how large the population, for a 10% error margin (and approaches a limit of 500 for a 5% error margin). The sample size can also be approximated using Equation 1 below:

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

Where:

- Z = Z value (e.g. 1.96 for 97.5% confidence level for single-sided, normal distribution; note that though human action is pass/fail and so is typically described using binomial distributions, for large populations/groups, a normal distribution approximates a binomial distribution)
- 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):

Equation 2: Validation SS (finite population) =

$$\frac{SS}{1 + \dfrac{SS - 1}{\text{population}}}$$

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 0.1 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." For Human IPLs, the group is a combination of similar operators (likely all "qualified, independent" operators are treated equally for this purpose) and similar responses to similar triggers. Creating this 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.

Next, the sample size and pass/fail targets must be estimated. This is based on the size of the groupings, the confidence level desired in the result, and the expected error margin (distribution) of the results.

- One method for determining sample size is to use Equation 1 and 2 discussed earlier. After the sample size if determined, then the pass/fail target (to prove if the hypothesis of a PFD of 0.1 is valid or invalid) can be estimated (refer to statistical textbooks for this derivation).

- Another recognized standard for determining the sample size and the pass/fail targets is to use ANSI Z1.4; this standard is now incorporated by reference within the US MIL standards for lot sampling and pass/fail determination. Examples calculations and the resulting savings achievable by use of sample plans are also provided in ANSI Z1.4.

Example Calculations: Using the same theoretical example as before with 20 operators and 130 alarms that require response, the total population of combinations of responses and operators is 2,600.

Case A: For one case, assume that all actions and all operators are equivalent and so the population makes up one group. In this case, at a confidence level of 97.5% and a confidence interval of 5%, the sample size adjusted for the finite population would be 131. So, only 131 combinations of operators and alarms/actions would have to be validated (tested) each period (and a typical period is each year). This is about 5% of the total population of human IPLs and so the validation would likely only take .05% of a staff year per operator, or 1 hour per year per operator (a very small investment in time for validation of IPLs). For this sample of 131 validation test, if 8 or more fail to accomplish the action in the required time, then All human IPLs have failed their validations. In addition, this is without counting for the stress of an actual response. If the error rate is doubled to account for stress, then the number of acceptable failures is cut in half, so for the sample size of 131, if 4 or more fail, then All human IPLs have failed their validations. If the validation failed, then the site would likely enlarge the sample size and re-test and/or find out where the problems are occurring and work to improve the response success. Regardless, the workload would be less with sampling and the company would obtain valuable insights into where to focus improvement efforts for Human IPLs.

Case B: For this case, assume that not all alarms/actions are equivalent, but still assume as in Case that all operators are equivalent. Further assume that the alarms and actions can be separated into 10 groups of 13 similar alarms/actions each. In this case, the population of any one group is 20 operators multiplied by 13 alarms or actions for a total of 260. In this case, at a confidence level of 97.5% and a confidence interval of 5%, the sample size adjusted for the finite population would be 91. So, 91 combinations of operators and alarms/actions from each of the 10 groups would have to be validated (tested) each period (and a typical period is each year). This is about 35% of the total population of human IPLs and so the validation would likely only take 0.35% of a staff year per operator, or 6 hour per year per operator (a small investment in time for validation of IPLs). For this sample of 91 validation test, if 6 or more fail to accomplish the action in the required time, then all human IPLs in this grouping of 13 alarms or actions have failed their validations. In addition, this is without counting for the stress of an actual response. If the error rate is doubled to account for stress, then the number of acceptable failures is cut in half, so for the sample size of 91, if 3 or more fail, then all human IPLs in this grouping have failed their validations. If the validation of a group of alarms failed, then the site would likely enlarge the sample size for that group and re-test and/or find out where the problems are occurring and find ways to improve the response success. Regardless, the workload would be less with sampling and the company would obtain valuable insights into where to focus improvement efforts for Human IPLs.

Table 5 provides various Sample Size and the Maximum Number of Failures (pass/fail target) using standard statistical methods (such as Equation 1 and 2); this table includes the data for Case A and B above.

6. Estimation of PFD Using Generic Human Factor Data and Calculation Approach

An alternative to collecting your own data to validate the PFD for a human IPL is to calculate the PFD by combining expert judgment and generic and predicted data (see CCPS text, Guidelines for Independent Protection Layers and Initiating Events, Appendix B[4] and the paper by Stack [15]). The calculated value is then used to validate the PFD value to be used in LOPA.

The PFD is calculated by combining a baseline unreliability value with Performance Shaping Factors (PSFs), also called human factors. The **baseline unreliability** value depends on the type of activity and generally is **set at 0.01 for a response activity** when used for basic LOPA. *(Note: the baseline error rate can be as low as 0.001 for an initiating event activity, where time pressure is not a factor, such as following a procedure to startup a unit.)* The calculation steps use the following PSFs that are specific to the response task:

- Number of steps in the procedure
- Number of activities or alarms per year
- External stress factors
- Communication factor
- Feedback factor

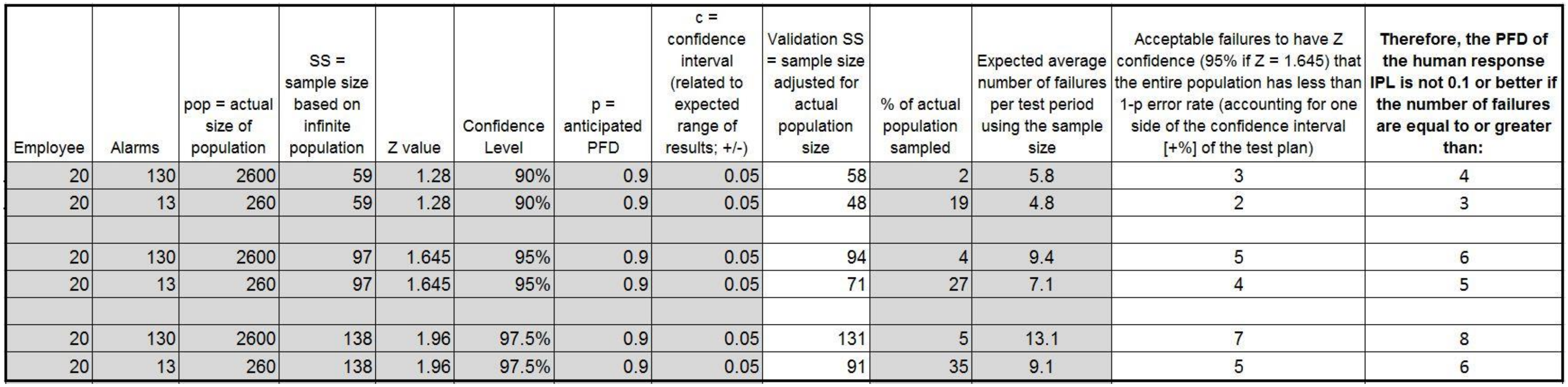
In addition, this value must be adjusted for a general score on global human factors issues at the site, such a control of fitness for duty (most notably control or not of fatigue), work environment, and human system interface. This global human factor issues adjustment is described at the end of this section.

6.1 Example of Calculation Approach, before Adjustment for Global Human Factors Issues

The example below illustrates this approach.

Number of steps in the procedure: If the number of steps in a procedure is less than 10, then the PSF multiplier = Number of steps/10, otherwise the factor is =1.0.

Table 5: Sample Size and Acceptance Criteria Estimation for Human IPL Validation (single-sided Z test; for normal distribution)



Employee	Alarms	pop = actual size of population	SS = sample size based on infinite population	Z value	Confidence Level	p = anticipated PFD	c = confidence interval (related to expected range of results; +/-)	Validation SS = sample size adjusted for actual population size	% of actual population sampled	Expected average number of failures per test period using the sample size	Acceptable failures to have Z confidence (95% if Z = 1.645) that the entire population has less than 1-p error rate (accounting for one side of the confidence interval [+%] of the test plan)	**Therefore, the PFD of the human response IPL is not 0.1 or better if the number of failures are equal to or greater than:**
20	130	2600	59	1.28	90%	0.9	0.05	58	2	5.8	3	4
20	13	260	59	1.28	90%	0.9	0.05	48	19	4.8	2	3
20	130	2600	97	1.645	95%	0.9	0.05	94	4	9.4	5	6
20	13	260	97	1.645	95%	0.9	0.05	71	27	7.1	4	5
20	130	2600	138	1.96	97.5%	0.9	0.05	131	5	13.1	7	8
20	13	260	138	1.96	97.5%	0.9	0.05	91	35	9.1	5	6

Number of activities or alarms per year: An activity frequency factor used as the PSF multiplier for the number of activities or alarms per year is obtained from Equation 3:

Equation 3: Activity Frequency factor = $(12/\text{frequency per year})^{0.25}$

External stress factors: The External Stressor Factors are provided for stress, conflict with normal expectations, and distractions with values ranging from 2 when an operator has significant distractions up to 15 in which some tasks in the procedure need to be executed in close proximity to the portion of the unit that is in a more dangerous mode for this accident scenario.

Communication factor: The effectiveness of communications has been shown to be a critical factor where a key operator is responding to an alarm. Values of 1 are assigned in cases where an alarm is received and acted upon by a single person up to a maximum 5 in situations where an alarm is received and acted upon by multiple people in different locations and the communication between the people is verbal by phone or radio.

Feedback factor: Feedback allows the operator to take additional action if the intended results do not occur, if there is sufficient time. If there is no feedback, then a PSF figure of 1 is used, with the lowest value being 0.01 if the feedback is immediate and compelling.

> **Example:** Consider the case of response to a high temperature situation in a monomer tank, which could lead to a runaway reaction and possible tank failure. The selection of values and the resulting calculation of the human response IPL for this scenario are presented in Table 5.
>
> The calculated PFD for the human IPL in this case (interim result) is:
>
> 0.01 x 1 x 1.57 x 6 x 5 x 0.1 = **0.047**
>
> Therefore a PFD of 0.10 would be valid in LOPA.
>
> However, if the tank farm fire water monitor were activated from the control room, by the control room operator, then the alarm would be received and acted upon by a single person (so the Communication PSF Value becomes 1) and is checked by a video monitor in the control room (so the Feedback PSF Value becomes 0.05), then the calculated PFD for the human IPL in this case (still an interim result) would now be:
>
> 0.01 x 1 x 1.57 x 6 x 1 x 0.05 = **0.005**
>
> Therefore a value of 0.01 should be valid for use of this human IPL in a LOPA scenario.

Table 5: Estimation of PFD of Human IPL Using Calculations

Step #	Validation Step Description	Value	PSF (multiplier)	Probability	Perception, Diagnosis, and Response Activities
1	Activities Involved (type of action required) - this sets the **baseline probability**			0.01	Board operator acknowledges alarm & contacts field operator to put fire monitor spray onto the monomer tank to control high temperature detected/alarmed in the tank
2	Number of steps	10	1		10 steps in the actual procedure, including immediate actions, diagnosis, and feedback checks
3	Number of activities or alarms per year	2	1.57		Estimate of 2, since operator is trained or tested annually on what to do and there is also about one event per year
4	External Stressor Factors		6		Field operator does not have to go to monomer tank (hazardous zone) to verify the temperature; there is remote display from multiple indicators in the control room; independent of alarm. Field operator does not have to go near the monomer tank. Control operator still under threat of taking wrong actions.
5	Communication Factor		5		Board operator and field operator have two way push to talk cell phones to work through the trouble-shooting guide
6	Feedback Factor		0.1		Board operator would observe tank temperature after water spray application by field operator, and field operator still required to observe water spray from safe distance, so feedback is good
			Interim Result	0.047	<< Calculated PFD for this Human IPL
			FINAL RESULT	**0.10**	**<< PFD to use for this Human IPL (must be equal to or less than 0.1 to allow this Human IPL)**

6.1 Example Adjustment for Global Human Factors Issues for the Site and Mode of Operation

The value above did not pre-adjust the baseline human error rate for the mode of operation and for the specific site. Global Human Factors Issues that are different for each (1) mode of operation and (2) site include:

- Fitness for Duty
- Work environment
- Human System Interface

Many approaches have been proposed for qualitative scoring (excellent, good, fair, poor) and for quantitative estimation of global human factors issues, such as SPAR-H[14]. These approaches share a common strategy of combining the multiplying effect that poor human factors can have on baseline error rates. For an IPL, the baseline rate is normally set at 0.01 because of the success rate reduction imposed by a limited time to take detect, decide, take action, and verify the response. As stated in Table 1, each human factor deficiency can have a multiplying effect on the number of human errors:

Fitness for Duty Control

- Fatigue – Can affect error rates by a factor of up to 20 times.
- Other fitness of duty issues – Can affect error rates by a factor of up to 20 to 50 times

Work Environment Control

- Can affect error rates by a factor of up to 5 times

Human-System Interface Weakness

- Can increase human error rates by 10 times or higher (especially if "norms" are violated, where a negative effect of 20 to 50 times is assumed, but we are assuming that a LOPA scenario evaluation or PHA/HAZOP analysis prior to LOPA would have already eliminated any violation of human-system interface norms.)

It has been noticed that even in accidents involving a combination of human system weaknesses, the error rates for humans appears to reach a plateau of 20% maximum (or 1 mistake in 5 steps). Therefore, 0.2 is the upper bound for a PFD for a human response IPL and 0.01 is the lower bound. One approach is to use the estimates of the "local" factors for fitness for duty control, work environment control, and human-system interface weaknesses to adjust the interim human error rate calculated in section 6.1, but limit to the bounds just stated. Table 6 lists typical multipliers we use for Global Human Factors Issues.

SPAR-H typically has larger multiplying effects, but we adjusted these to match the 0.01 to 0.2 range for PFD for a human response IPL.

Table 6: Adjustment Factors for Estimation of PFD of a Human Response IPL Global Human Factors Issues (based on SPAR-H[14])

Human Factor Category	Human Factor Issue/Level	Multiplier for Cognitive & Diagnosis Errors
Fitness for Duty	Unfit (high fatigue level, illness, strong medication, not physically capable of job today) Degraded Fitness Nominal	10 - 20 5 1
Work Environment	Extreme (high noise, heat, vibration; too cold) Good	5 1
Human-Machine Interface (includes tools)	Missing/Misleading (violates what is normally expected) Poor Nominal	10 - 20 5 1

Adjustment Approach:

1. Rank the Human Factors Issue/Level using the qualitative terms in the middle column of Table 6.
2. Select the matching multiplier in the right-hand column.
3. Multiply the Interim PFD (as calculated in Table 5) by the multipliers, but do not exceed a PFD of 0.2
4. If using Basic LOPA, do not use a human response IPL that has a PFD greater than 0.1.

One obvious benefit of this adjustment approach is that it raises attention to the gains possible on improving human factors to drop these multipliers closer to 1.

7. 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 baseline human error rate of 1 error in 50 steps is reasonable, or for excellent control of human factors, a baseline error rate of 1 error in 100 steps may be achieved (as described earlier). This baseline error rate then can increase if the control of human factors slip, as described earlier in this paper. Increasing human error rates in inspection, calibration, maintenance, quality control of part, commissioning (or re-commissioning activities) will impact the failure rate of engineered IPLs or cause more frequent initiating events.

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 process environment that is highly sensitive to materials of construction, such as highly corrosive service or cryogenic service.   So, the IEF will be many times higher than expected.

Industry data in the 1990s indicated that the materials delivered were different than what we specified about 3 to 7% of the time. Positive material identification (PMI) was employed to detect and correct such failures, but plant data suggests that using 100% PMI at receiving lowered the composite error rate to 1% and that another 100% PMI of “as-welded in the field” lowered the combined error rate to about 0.3%. This still left a large number of components that may be of the wrong materials.   In the past 10 years, the use of PMI throughout the entire supply chain has dramatically improved and third-party inspectors are now expert at PMI checks as well. This about a 10 fold reduction in the number of parts received that are the wrong materials of construction; the error rate is now below what is measureable by plant staff doing 100% or 200% PMI checks.

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; for instance the humans may be 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. 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 or otherwise increase the PFD of the SIS. The current versions of the standards also do not explicitly require validation of the likelihood of other systemic errors, such as plugging of sensor ports by contaminants through the system, which would be a common cause failure mechanism for all ports. Although a SIL 4 is allowed by IEC 61508, with a PFD of $<10^{-4}$ to $\geq 10^{-5}$, it is very unlikely to control human errors low enough to install or maintain a SIL 4. Similarly, due to the complexity of the analysis of human errors associated with in situ testing of SIS, basic LOPA only allows use of SIL 1, unless the systemic errors of leaving SIS bypassed is accounted for rigorously in the SIL verification step.

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 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 that 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. A preferred design is to install dual relief valves with a single valve arrangement (3-way valve) that helps ensure against blocking flow to the on-stream relief valve.

8. 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.

9. References

1. "Layer of Protection Analysis (LOPA) Guideline," CCPS/AIChE, 2001.
2. Bridges, W., "LOPA and Human Reliability," *6th Global Congress on Process Safety,* AIChE, 2010.
3. Tew, R. and Bridges, W., "Human Factors Missing from PSM," *6th Global Congress on Process Safety,* AIChE, 2010.
4. "Guidelines for Independent Protection Layers and Initiating Events," AIChE/CCPS, 2011 [pending].
5. Swain, A. D., Guttmann, H. E., "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, Final Report," NUREG/CR-1278, 1983, US Nuclear Regulatory Commission.
6. Rasmussen, J., and Rouse, W., "Human Detection and Diagnosis of System Failures," New York, Plenum Press, 1981.
7. Rasmussen, J., "Chemical Process Hazards Identification," *Reliability Engineering and Safety Systems 24*, Elsevier Science Publishers Ltd., UK, 1989.
8. Bridges, W. and Clark, T., "How to Efficiently Perform the Hazard Evaluation (PHA) Required for Non-Routine Modes of Operation (Startup, Shutdown, Online Maintenance)," *7th Global Congress on Process Safety,* AIChE, March 2011.
9. "Guidelines for Hazard Evaluation Procedures, 3rd Edition", CCPS/AIChE, 2008.
10. Bridges, W., Kirkman, J. and Lorenzo, D., "Strategies for Integrating Human Reliability Analysis into Process Hazard Evaluations," *International Conference*

on Hazard Identification and Risk Analysis, Human Factors and Human Reliability in Process Safety, CCPS/AIChE, January 1992.

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

12. “Guidelines for Investigating Chemical Process Incidents, 2nd Edition,” CCPS/AICHE, 2003.

13. “Training Requirements for Nuclear Power Plant Operators,” 10 CFR 55.45 and 55.59, US Nuclear Regulatory Commission,

14. Gertman, D.; Blackman, H.; Marble, J.; Byers, J. and Smith, C., “The SPAR-H Human Reliability Analysis Method,” NUREG/CR-6883, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, Washington, DC, August 2005.

15. Stack, R. and Delanoy, P., “Evaluating Human Response to An Alarm for LOPA or Safety Studies,” *6th Global Congress on Process Safety*, AIChE, 2010.