Proven Approaches to Ensuring Operators Can Respond to Critical Process Deviations in Time (Human Response IPL)

William G. Bridges, President
PROCESS IMPROVEMENT INSTITUTE, INC. (PII)
1321 Waterside Lane, Knoxville, TN 37922
Phone: (865) 675-3458
Fax: (865) 622-6800
e-mail: wbridges@piii.com

2017 © Copyright reserved by Process Improvement Institute, Inc. "PII"
Prepared for Presentation at
13th Global Congress on Process Safety
San Antonio, TX
March 27-29, 2017

UNPUBLISHED

AIChE shall not be responsible for statements or opinions contained
in papers or printed in its publications

Proven Approaches to Ensuring Operators Can Respond to Critical Process Deviations in Time (Human Response IPL)

William G. Bridges, President
PROCESS IMPROVEMENT INSTITUTE, INC. (PII)
e-mail: wbridges@piii.com

Keywords: Risk control, process safety management, PSM, LOPA, Layer of Protection Analysis, Independent Protection Layers, IPLs, human error prevention, Human Factors, human response

Abstract

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). Validating Human IPLs has been a show stopper for many companies considering the use of human response as an IPL. Human IPLs 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 first describes the fundamentals of clear alarms, practical actions, and having enough time to perform the action, all without being in harm's way at the end of the action. This paper builds upon earlier studies (based on similar papers from 2010 and 2011)[13, 14] of how to collect the data needed for directly measuring the probability of failure on demand (PFD) of the human response. The preferred 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. An example is provided of how a simple data collection and validation method can be set up within a company. This paper also provides an overview of alternative methods for estimating the PFD of a Human IPL, based on plant and scenario specific factors (such as stress factors, complexity, and communication factors); and the paper evaluates and compares alternative approaches to validating human IPLs, including expert judgment based on frequent practice of trouble-shooting by operators. All of these methods were available in Appendix B and C of the initial full draft of the CCPS book, *Guidelines for Initiating Events and Independent Protection Layers, April 2012 (unpublished), but unfortunately o*nly a small fraction of one of the Appendices was retain in the version that was published in January 2015. This paper provides that missing information.

Response to Critical Process Deviations – Fundamentals

Humans response to an alarm or troublesome reading or sample result can be a great safeguard layer against major accidents because the human has the capability to diagnose false alarms and in other ways prevent a spurious shutdown of a process. Many organizations are reluctant to use human response as a protection layer because "humans are human" and make mistakes. Of course, instrumented systems fail as well. Nothing is perfect. The key to using human response as a layer of protection is to follow the general guidelines for qualifying a response as a Human Independent Protection Layer (IPL).

There are 5 steps to using a human to respond to a critical process deviation as a Human Response IPL:

1. Determine which parameters limits (announced by alarms or other triggers) should have human response, and why human response is best
2. Ensure human response action meets the definition of an IPL
3. Develop a trouble-shooting guide (general steps for the operators to take) for each response
4. Perform initial training on each human response IPL
5. Validate that human response success rates are high enough

This paper explains what we have found to be best practices for each of these steps

1. Determine which parameters limits (announced by alarms or other triggers) should have human response, and why human response is best

This important first step is difficult for many organizations since they either have a prejudice against using a response by a human as an IPL or because they do not trust their hazard evaluation teams to make such judgments. The CCPS book on LOPA (2001)[1] and the follow-on book, *Guidelines for Initiating Events and Independent Protection Layers*, CCPS (2015)[2] both allow the use of Human IPLs and these books given criteria on their use. Neither book tells how to determine if and when a Human IPL should be used, but instead state that the PHA (HAZOP, What-If, etc.) is the setting for determining what layers of protection are appropriate and if that fails, then LOPA itself allows the use of a Human IPL as one IPL in a scenario, but it is up to the LOPA analyst to determine that the Human IPL is the best selection from alternatives for risk reduction.

If qualitative methods such as HAZOP or What-If brainstorming teams are used to determine when a Human IPL is best, then it is recommended that the team follows the protocol in Bridges & Dowell, 2016.[3] This approach provides for qualitatively determining when safeguards (including human response safeguards) meet the definition of an IPL. Once determined, the Human IPL should be documented as such in the PHA analysis as shown on in Table 1.

Table 1. Documentation of IPLs in a PHA (in this case HAZOP method) Analysis Table[3]

No.: 2 XXXX storage spheres xxx-T-XX A/B/C/D/E/F/G/H/I/J/K/L (1 of 12)					
#	**Dev.**	**Causes**	**Consequences**	**Safeguards**	**Recommendations**
2.1	High level	Too much flow to one sphere from XX Plant (through their pump; about 40 bar MDH)	High pressure (see 2.5)	High level SIF with level sensors voted 2oo2, to close inlet valve - **SIL 1** Overflow thru pressure equalization line to other spheres (through normally open [NO] valve) - **IPL**	
		Misdirected flow - Liquid from xxx Plant(s) to spheres (see 1.4)	Overpressure of sphere not credible from high level, for normal operating pressure of the column (which is 1.75 MPa), unless all spheres are liquid filled and then thermal expansion of the liquid could overpressure the spheres	High level SIF with level sensors voted 2oo2, to close inlet valve - **SIL 1** Overflow thru pressure equalization line to other spheres (through normally open [NO] valve) - **IPL** Spheres rated for 1.95MPa (19.5 Bar, approx) and the highest pressure possible from the column feeding the spheres is 1.75 MPa Level indication and high level alarm in DCS, used by operators to manually select which tank to fill – **Human IPL**	
			Overflow into the equalization line will interfere with withdrawal from the column, but this is an operational upset only		
			Excessive pressure on inlet of high pressure liquid pumps, leading to excess load on pumps and trip of pumps on high pumps, causing trips of xxx, xxx, etc. - significant operability issue		
2.2	Low level	Failing to switch from the sphere with low level in time (based on level indication)	Low/no flow - Liquid from spheres through high pressure product pumps to the vaporizer (see 4.2)	Level indication and low level alarm, inspected each year, per government regulation **(not IPL; part of the cause)** 9 other spheres with possibly enough level to switch to Feeding from two spheres at all times, so unlikely for BOTH spheres to have low level at the same time - **IPL** Two level indication from SIS level transmitter, with low level alarm, with more than 60 min available to switch tanks (SIF driven alarm and response) - **possible IPL, if action of the operator is quick enough**	Rec 4. Make sure the Human IPL of response to low level in all spheres and tanks is described in a trouble-shooting guide (like an SOP) and practiced once per year per unit operator. This will make this response a valid IPL.
			Low/no flow - Unqualified liquid from spheres back to Plant (see 6.2)		
2.3	High temp.	Large area of damaged insulation Loss of cooling. when the tank is isolated from column	High pressure - vapor from spheres through condenser and return to liquid pump out line (only used when plant is shutdown) (see 3.7)		

2. Ensure Human Response Action Meets the Definition of an IPL

The prior step should have already performed this check when determining if a Human Response IPL is the best choice of protection layer for the scenario. Regardless, before going to the further effort on ensuring proper human response, the organization should ensure all of the criteria for a human response IPL has been met. Per the *Guidelines for Initiating Events and Independent Protection Layers, 2015*[2] this includes:

A. The Human IPL (and any associated alarm) must be truly independent of the other protection layers. That is, there must be no failure that can deactivate two or more protection layers.

The IPL (also applies to IE) includes the ENTIRE sub-system, including any root valves, impulse lines and bypasses. The other IPLs cannot share any of these or other components (except for the mother board of the BPCS loops).

A device, system, or action is **not** independent of the initiating event and cannot be credited as an IPL for either approach if either of the following are true:

- Operator error is the initiating event and the candidate IPL assumes that the same operator must act to mitigate the situation. Human error is equivalent to the failure of a system and once a human has committed an error it is not reasonable to expect the same operator to act correctly later in the sequence of events. This approach is justified because the error may be due to fatigue, illness, incapacity (drugs or alcohol), distraction, work overload, inexperience, faulty operating instructions, lack of knowledge, etc., that are still present later when the action is required.

Examples where the Human IPL is not independent include

- Assuming that the same operator acts correctly after operator error initiated the event.
- Alarms that are annunciated on the BPCS are not independent of the BPCS; if the BPCS is counted as an IPL, then such alarms cannot be counted as an IPL (again, see the exception discussed later).

B. The Human IPL is specifically designed (capable) to prevent or mitigate the consequences of a potentially hazardous event.

- Is the Human IPL valid for the mode of operation for the scenario (startup, shutdown, normal, batch, etc.)?
- Is the Human strong enough to perform the required action, such as closing a manual isolation valve
- Is the Human fast enough (discussed in a little more detail later in this section)?
- What is the maintenance/reliability practices and plant/company history for any related equipment that the Human must use the complete the desired action? How much likelihood reduction credit will you take for the alarm working?
- How good are the procedures and related training (and drills)? Were the operators trained in specifics of how to respond to this alarm/indication? Are they test often enough?

C. The Human IPL must be *Maintained, Tested,* and *Validated* periodically; it must be proven that the Human IPL can be relied upon to do what it was intended to do.

The IPL must be periodically maintained and it must be proven or validated. The site must have data that supports the reliability factor. The frequency and test method must comply with best industry practices for such IPLs. Also, the site must maintain a database for each IPL that statistically supports the PFD stated. For a component or instrumentation IPL, this requires maintaining a statistical failure rate database that justifies the PFD listed for each IPL. For a human IPL, the site must maintain data from "drills" of the action of the worker that statistically demonstrates that the worker(s) can indeed implement the required action (of the IPL) with the time specified in the IPL, or else they must use another means of validation as discussed in Section 5 of this paper.

D. **The Human IPL maintenance and validation must be *Audited*.** Auditing is required to ensure the validation, procedures, training, and resulting data are adequate. This is an administrative check. This auditing cycle is set frequent enough (typically 1 year for the first audit and then 5 year frequency after that) to ensure that validation is being carried out as planned and is sufficient to justify the IPL and its PFD.

Specific Criteria on Speed of Response versus Process safety Time.

The criteria for setting the alarm level (that sets the time available for response) should be true before going to the effort of developing a procedure for response (before developing a trouble-shooting guide explained in Section 3):

- The response is typically still possible, but it is time dependent. The time available is called the process safety time (PST). The operator must complete the diagnosis, make the necessary change(s), and make sure they are out of harm's way by the end of the Maximum Allowable Response Time (MART). [5]
- We usually set an alarm or a pre-alarm to trigger this action. This is usually before the shutdown triggers (ESD occur automatically) or release points (PSV set points) are reached
- The Min and Max shown in a Trouble-shooting Guide are not the absolute safety limits for a system, but are instead some values that leave us some time to take action to prevent from reaching the absolute limits.
- There is still time to prevent or avoid the final consequence that could occur if we reach the ultimate limits of the process. Usually, we want the MART to be ½ or less of the PST, and we want MART > 10 minutes for trouble shooting in the field/plant and MART > 5 min for trouble-shooting only from the control room.[5]

See the PII database of IPL Datasheets for detailed criteria on qualifying a human IPL[4]. Similar datasheets are also available in *Guidelines for Initiating Events and Independent Protection Layers,* CCPS (2015)[2]

3. Develop a Trouble-Shooting Guide (TSG) for Each Response

TSGs are a special form of operating procedure. They are written for the actions we want the operators to take to recover from a process deviation, *before an emergency situation occurs.* They are called guides since rarely can we predict the process conditions at the time the action is required. Trouble-shooting guides (and necessary training and drills) are required for any action that is considered a Human IPL. The Action Limit is what we show as the Min or Max in a Trouble-Shooting Guide. The action limit triggers the demand to use the trouble-shooting guide.

If the unit has a good PHA/HAZOP, then it is best to extract information from the HAZOP (or What-If) analysis tables to start the development of each guide. (Table 2 shows examples of the conversion of HAZOP entries into TSG entries.) The guide is then finished with input from the process experts.[6]

Table 2 Examples of Creating a TSG from a HAZOP Table

HAZOP Table Entry	**Trouble-Shooting Guide Entry**
Cause: *Bypass valve is open or passing*	Make sure the bypass is tightly closed
Safeguard: *Isolation valves for the vessel*	Isolate the vessel, if necessary
Safeguard: *Relief valve*	Make sure the relief valve block valves/relief path are open

The key categories of information needed in a trouble-shooting guide are:[6]

- IMMEDIATE ACTION (by system or by operator)
- DECIDE IF ALARM is REAL
- FINDING and FIXING the CAUSE
- FIX or BYPASS PROBLEM

Figure 1 provides an example of a trouble shooting guide for one critical alarm/action. This guide is in portrait format and follows best human factors practices for formatting.

Optimal Presentation of Trouble-Shooting Information

Ideally trouble-shooting information should be imbedded in the basic process control system (DCS) so that the operators can access the information on demand, with the click of mouse or key. Using the DCS for display of the steps for response to alarms (trouble-shooting), to be displayed "on demand" is becoming more of the norm each year. So, for more than 15 years, many companies have been taking the TSG information, such as that shown in Figure 1, and imbedding it in the DCS; the operators can then access this reminder of the proper response with a click of the human system interface (such as a click of the mouse).

Figure 1: Example of Trouble-Shooting Guide that Follows Best Practices

Trouble-Shooting Guide

Alarm or Indicator:	**PAL 4446 – Low Pressure Alarm for Suction of Organic Feed Pump 40-PM-18.445**

Action Limit:	5 kPa		
Consequence:	Possible pump seal failure, releasing or spraying organic waste into the berm.		
Process Area:	FB&D Incinerator; Liquid Organic Liquid Feed	**Oper. Mode:**	Normal
Drawing #s:	D-400-PI-013		

IMMEDIATE ACTION (by system or by operator)

- DCS should shut down the organic feed pump (40-PM-18.445).
- From the DCS display, MAKE SURE the organic feed pump is shutdown.
- HAVE the field operator check for leaks near the organic feed pump.
- IF there is a large leak/release, THEN use the ESD switch to shutdown the unit and then follow/complete the shutdown and isolation procedure, OPS-ESD-117.
- IF there is a minor leak or no leak, THEN:
  - COMPLETE the rest of the trouble-shooting,
  - and DECIDE how to contain the leak for now,

DECIDE IF ALARM is REAL

- From the DCS, CHECK the pressure and feed tank level trends. IF the trends indicate the alarm if valid, THEN continue with finding the cause or fixing or bypassing the problem.

FINDING and FIXING the CAUSE

- CHECK valves upstream of the organic feed pump to see if any are closed too far, including checking ESD valves.
- CHECK, by feel with hand, if the heat tracing is on; IF Not, then TURN ON or open heat trace valves
- MAKE SURE nitrogen to the pump seal is at the normal operating pressure.
- CHECK if the line if plugged or frozen (skill)

ABC Chemical Company Prepared by: *Printed copy of this procedure is good for one job task duration.*	OPS-76-TSG-233	Incinerator Unit Revised 8/24/2015 Printed 2/21/16 Page 1 of 2

Figure 1: Example of Trouble-Shooting Guide that Follows Best Practices (continued)

Trouble-Shooting Guide

Alarm or Indicator:	**PAL 4446 – Low Pressure Alarm for Suction of Organic Feed Pump 40-PM-18.445**

- CHECK if the line if plugged or frozen (skill)
- CHECK if the level is actually low, use the Organic Feed Tank.
- IF the cause is low level in the feed tank, THEN resolve the problem if necessary based on cause that is found (skill)

FIX or BYPASS PROBLEM

- IF necessary, SHUT DOWN the Unit to allow fixing of the problem.
- FOLLOW the proper procedure to resolve problems (repair procedure, line clearing procedure, etc.)
- IF the decision is made to continuing operation without all equipment in normal condition, THEN:
    - FOLLOW Temporary Operating mode, if there is a temporary procedure already written for this possible problem/condition
    - FOLLOW MOC procedures to obtain approval for any non-standard temporary operating procedure or mode

END

Image and layout above copyrighted by PII, 2008-2017

4. Perform Initial Training on Each Human Response IPL

Once the TSG is developed then initial training of operators can start. Initial training is straight forward and includes a progression of activities such as:

- Reading the TSG
- Classroom or simulator training on the response action
- Discussions to increase understanding on the importance of the action
- Discussions to ensure operators know the options and judgments needed with the TSG approach
- Tools needed (if any, such as a valve wrench to get more leverage on a valve)
- Optional Written Exam on the task

As with all other activities in life, the real "training" occurs in practice in the control room or unit. With a Human Response IPL, the hands-on practice is doubly important, since if the triggering alarm sounds, the operator will many times Not have time to refer to a procedure; the response must be practiced enough to make it second nature to the operator. The amount of practice should be enough to ensure the probability of failure to accomplish the task in time is less than 0.1.

Repeat practice necessary to maintain adequate response capabilities (see Section 5 for details on Human IPL Validation)

As discussed in Section 5.3, this may require one drill per operator per alarm per year. But after the first year or two, the organization may decide to use a different method of Human IPL validation, such as Expert Judgment.

5. Validate that Human Response Success Rate is High Enough

The values for failure rates and probabilities of failure on demand (PFDs) used in LOPA should be conducted to ensure that the selected values are appropriate. Validation of the values used in risk analysis can follow any of the four methods used for initially establishing failure rates. The validation method used for any particular value can differ from the original method used to determine a failure rate or probability of failure on demand.

As mentioned earlier, much of Section 5 was originally included in the *Guidelines for Initiating Events and Independent Protection Layers and Initiating Events, 2015*[2], but was subsequently cut before publication of the final book. This paper provides that useful information once again.

Validation approaches

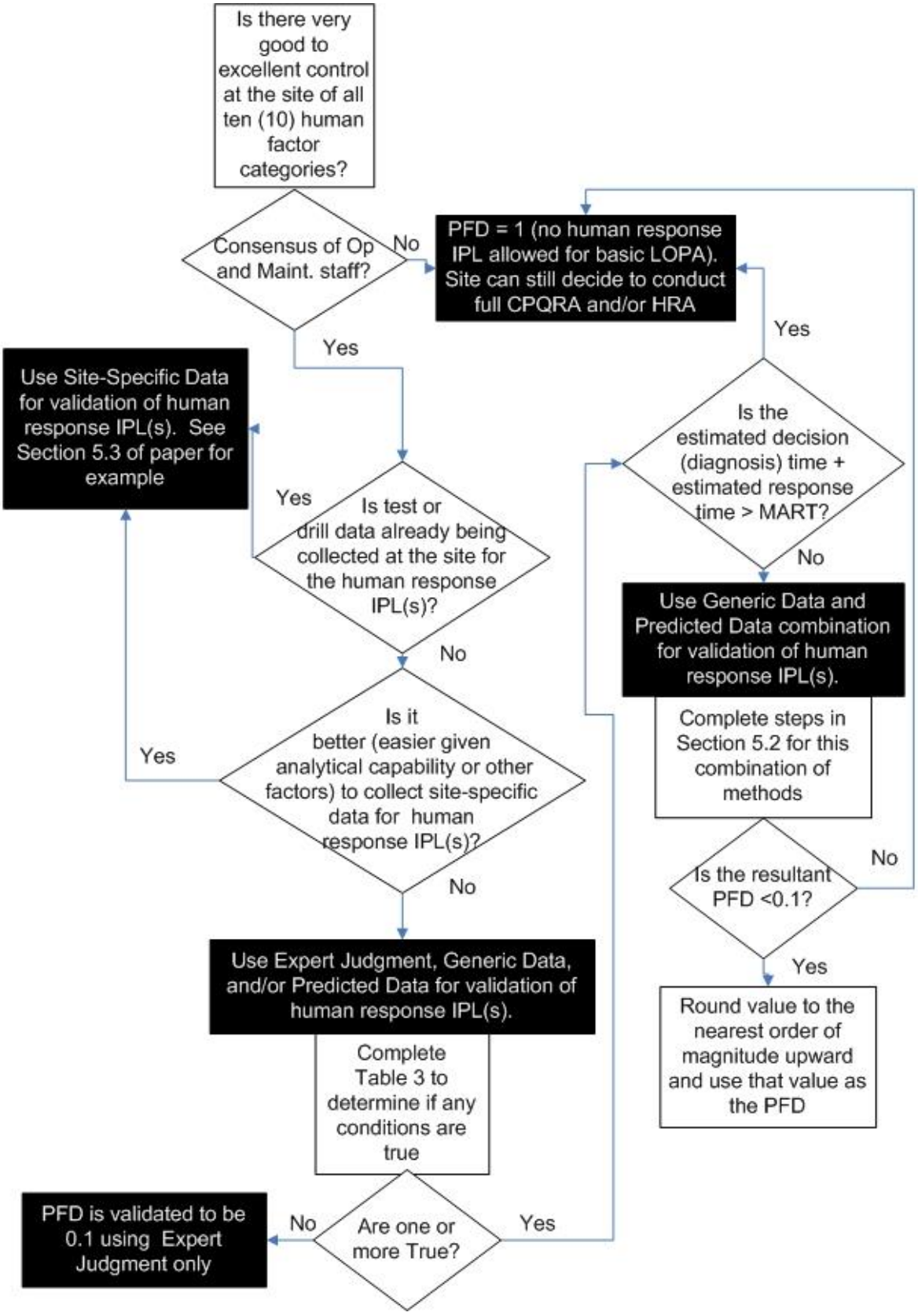
The four methods of *validation* are presented in increasing order of robustness. As values to be used in a risk analysis increase in the reliability claimed, consideration should be given to using more robust validation methods. In particular, site-specific data can be used to justify extending the reliability claimed for initiating events or independent protection layers beyond the values derived from generic data.

- *Expert Judgment* is often used as a validation method for all of the other methods of generation of values for failure rates and probabilities of failure on demand. The critical evaluation of values derived from other methods by one or more experts is a means of validating the values used in a risk analysis.
- *Generic Data.* Validation by the use of generic data is through monitoring of any updates to the data used in an analysis along with other sources of generic data to ensure that improved values developed by industry sources do not invalidate the values originally selected. Validation by the use of generic data can be used for original values derived from expert judgment, generic sources, and predicted values. Normally, values for failure rates derived from analysis of site data are more accurate than any generic source, and generic source data would not be used to change the value derived from analysis of site data.
- *Predicted Reliability Values* can be used for validation in situations where the original values selected were based on expert judgment or generic data and predicted values become available. As an example, a level transmitter might be assumed to fail once in 10 years based on expert judgment or generic data. If a predicted value of failure is later reported by the manufacturer for the particular level transmitter of one failure in six years based on calculation from component failure rates, the lower reliability may invalidate the LOPA based on the less conservative value.
- *Site-specific data*. The most robust means of validation of the values used in LOPA is through the collection and analysis of failure rate data in the area being analyzed. Site-specific data can be used to validate failure rates developed from expert judgment, generic data, and predicted values. The process of validation may support the values developed from the original method, show that the actual reliability of the systems are not as good as that predicted from the other methods, or reveal that the systems are actually more reliable than anticipated. When the site-specific data reveals that systems are not as reliable as originally developed, steps can be taken to improve the reliability or correct any potential deficiency in risk management by taking additional mitigating actions. In situations where the site-specific data reveals that the reliability of the systems in a particular application are better than originally derived, the better values can be used in the design of future applications.

Choosing the Approach for Validation of Human IPLs

Figure 2 describes the decision making necessary to choose Expert Judgment, Prediction/Estimation, or Site-Specific Data approaches for validation. (Note that CPQRA in Figure 2 refers to *Guidelines for Chemical Process Quantitative Risk Analysis* [CCPS 2000].)[7]

Figure 2 Example Decision Path for Validating Human Response IPL; *copyright PII, 2017*



5.1 Determine if Validation by "Expert Judgment Only" Is Allowed

The questions in Table 3 are used in the initial screening process. If ALL are False, *then a PFD of 0.1 is valid without further calculation.* If ANY of the statements in Table 3 are true, then the validation calculation approach (section 5.2) or the Site-Specific Data (Section 5.3) must be applied to verify that the required reliability is achieved.

TABLE 3 Determine *(Using Expert Judgment)* if Validation by Predicted Data Is Required

#	Criteria	True	False
1	Based on consensus of expert opinion, the operator has less than 15 minutes to successfully detect, diagnose, and perform the required action		
2	Operator response is required without explicit criteria and response instructions		
3	Critical or emergency responses involve multiple people		
4	Response actions provide no feedback that they are effective		
		If any are true, go through Predicted Data method for validation of the PFD	If all are false, use PFD = 0.1 (stop: no further calculation needed; Expert Judgment is sufficient for validation)

5.1.1 Determine if the Expert Judgment Estimate of the Baseline Human IPL Time Is Less than MART

The time available is critical to the reliability of any response activity. A shortage of time leads to hurrying and increased stress. In addition, under these conditions, any errors that occur may not be correctable. Ultimately, the action must be accomplished within the MART.

For a human response IPL, MART is the time from when the sensed parameter reaches the set point (and then perhaps a few moments later the alarm sounds, if alarmed) to the point of no return where the corrective action can no longer prevent the incident. This value is determined from process dynamics independent of any hoped-for human response. It includes any time delay in alarm activation and any time for automated actions (initiated by the operator) to occur.

This example of validation by predicted data requires three different time elements versus MART (maximum time available to stop the event):

A. *Detection time.* Time from when the parameter of interest exceeds the "safety" limit until the deviation is noticed by the human.

- Detection and annunciation could be via a sensor and alarm, followed by sensory perception of the annunciation by the operator.
- Detection could be by the operator taking a sample to the lab and then subsequent analysis and reporting of results by the lab technician. The detection time in this case includes time between sampling (at least one cycle) plus the time to take the sample plus the time to wait for analysis and perform analysis, plus the time to report the results to the appropriate operating staff.
- Detection could be the operator noticing a problem during routine operator rounds, in which case the time since the previous rounds is the major portion of the time consideration. So, for rounds every four hours, the detection time is greater than or equal to four hours; but note that it is best to rotate operators every round to enhance vigilance.

Use Expert Judgment for this estimate.

B. <u>*Decision time*</u> (time to decide what action to take; also called *diagnosis* time in HRA). The decision time was identified as a source of variability when people assessed the reliability of these activities. Consequently, the decision time is fixed within this validation method based upon the activity type. For purposes of alarms that a site would allow for LOPA, the decision time is normally less than one minute. But some HRA data developed for diagnosis time in control rooms (Swain, 1983) suggests that there is 90% chance the diagnosis will be correct if the worker in a nuclear power plant control room has at least 10 minutes to diagnose, and a 99% chance of correct diagnosis if they have 40 minutes. Because of these traditional values, the decision time is typically set at 10 minutes. However, for actions that require no or very little diagnosis or in simple process units, this value can reasonably be set to five minutes. Use Table 4.

C. <u>*Response time*</u> (time to complete all the alarm response activities). This is the time required to complete the tasks that will prevent the undesired event, as detailed in the alarm response procedure (e.g., after the procedure has been chosen as the correct course of action). Use Expert Judgment for this estimate. (For comparison, this is the time that is measured directly by testing/drills in validation using Site-specific data; see Section 5.3 for details.)

Estimate the task response time: Using solicitation of expert opinion (including at least two senior operators from the unit and one senior process engineer or equivalent), develop an Expert Judgment estimate of the time to complete the response activities, given that the diagnosis is performed correctly.

IF: Detection Time (including any delays in a related instrument system) + *Decision time + Task response time > MART*

THEN: The human response IPL is not valid.

TABLE 4 The Decision-Time Factor Assigned to the Different Activity Types

Activity Type	Decision Time (minutes)
Unambiguous cue in a continuously staffed control room or similar staffing near an alarm annunciation location, with simple process and little or no diagnosis (with a decision tool, such as a troubleshooting guide).	5
Unambiguous cue in a continuously staffed control room or similar staffing near an alarm annunciation location, with complex process unit that requires diagnosis to deduce the failure cause and the proper action to take (with a decision tool, such as a troubleshooting guide).	10
Requires diagnosis of a novel fault situation (cannot be used for IPL in LOPA).	Beyond LOPA

If the total human IPL time is too great, then the site may:

- Decide to use other methods to validate the human response IPL
- Decide to redesign the human response IPL so that it can be done in less than the MART.
- Decide to redesign the system to eliminate or reduce the risk.
- Decide to install or upgrade other types of IPLs (such as IPS, which are faster to respond) in lieu of the human response IPL not being available (because it is currently invalid).

IF: Detection Time + Decision time + Task response time < MART

THEN: Proceed to the steps in the next section

5.2 Approach for Validation of PFD of HUMAN IPL Using Combination of GENERIC Data and PREDICTED Data

The approach shown below is based largely on the methods described in SPAR-H (NUREG /CR 6883)[8], which is mainly a simplification of the much more complicated and detailed HRA methodology developed for the analysis of critical tasks (Swain, 1983)[9]. The approach starts with a baseline human error rate (0.0008 per year) is the lowest human error rate and then corrects that rank based on the multipliers related to good or bad human factors. See Table 5 for PII's method of how to estimate the PFD of a human response based on adjustment for human factors at the site and adjustment for the number of practices per operator per year (total practice is the combination of drills and actual responses to that same alarm by the same operator).

TABLE 5 Estimation of the PFD for Human Response IPL based on Basic Human Factors

		Absolute Baseline at practice multiple times per day	0.0008
Human Factor Category	**Human Factor Issue/Level**	**Multiplier for Cognitive & Diagnosis Errors**	**USED**
Available Time (includes staffing Issues) – ***for responses only***	Inadequate time	P(failure)=100%	1
	Barely adequate time (≈2/3 x nominal)	10	
	Nominal time (1x what is expected)	1	
	Extra time (at least 2x nominal and >20 min)	0.5	
	Expansive time (> 4 x nominal and > 20 min)	0.1	
Stress/Stressors (includes staffing issues)	Extreme stress (threat stress; unloading ship with crane non-stop for more than 2 hours, imminent hazard nearby)	5	1
	High stress (time pressures such as during a maintenance outage; issues at home, etc.)	2	
	Nominal	1	
Complexity & Task Design	Highly complex task. Or very low complexity/boring task that requires 100% attention for more than 45 min.	5	1
	Moderately complex (requires more than one staff)	2	
	Nominal	1	
	Obvious diagnosis	0.2	
Experience/Training* (see the practice rate adjustment in at end of table)	Low experience relative to complexity of task; or poor/no training	10	1
	Nominal	1	
	High	0.5	
Procedures	Not available in the field as a reference, but should be	20	1
	Incomplete; missing this task or these steps; or untrustworthy (< 85% accuracy)	8	
	Available and >90% accurate, but does not follow format rules ***(normal value for process industry)***	3	
	Good, 95% accurate, follows >90% of format rules	1	
	Diagnostic/symptom oriented	1	
Human-Machine Interface (includes tools)	Missing/Misleading (violates populational stereotype; including round valve handle is facing away from worker)	20	1
	Poor or hard to find the right device; in the head calc	10	
	Some unclear labels or displays	2	
	Good	1	
Fitness for Duty	Unfit (extreme fatigue level at >80 hrs/wk or >17 hr/day, no day off in 6-day period; or illness, legally intoxicated, etc.)	20	
	Highly degraded fitness (high fatigue such as >15 hr/day or >72 hr/wk, no day off after 4 shifts of 12 hours or more, illness, injury, legally barely intoxicated, etc.)	10	

	Moderately Degraded Fitness (≥12 hr day or ≥ 60 hours/wk; 1 day off [break] or more per week)	5	2
	Slight fatigue (more than 8 hr per day, but not more than 12 hr day; up to 48 hrs per work week, 1 day off [break] or more, after 48 hours of work; *normal value for process industry)*	2	
	Nominal	1	
Work Processes & Supervision	Poor	2	1
	Nominal	1	
	Good	0.8	
Work Environment	Extreme (in temp, humidity, noise, lighting, vibration, etc.)	5	
	Good	1	1
Communication	Communication system/interference damaged; poor communication environment	10	3
	No standard for verbal communication rules *(normal value for process industry); use this value if coordinated task for response*	3	
	Well implemented and practiced standard	1	
		Product	6.0
*** adjustment for practice frequency**	Number of times task performed and/or practiced per year	2	16.5
		Revised Product	98.7
		Product	0.0790
		PFD to use for this Human IPL (must be equal to or less than 0.1 to allow this Human IPL)	**0.10**

Simplifying Assumptions

Not all inputs for a full HRA evaluation are likely to be readily available to the LOPA team or analyst. Some HRA inputs are often team or analyst judgments that would lead to continued variability in the results. For HRA inputs that could be reasonably held constant, human factors are set at the expected norm or standard. If the team or analyst feels that the human factor are not "up to standard" for the IPL (or related process unit) being validated, a recommendation for improvement to the particular constant human factor would be made with the validation being contingent upon the plant or site completing the recommendation.

It is ideal if all of the human factors (except for practice) can be set to 1 (no negative effect), but the analyst should use their judgment for the task, mode of operation, etc., for the value of each human factor. ***It is very rare for a site to have all human factors at a value of 1.***

As mentioned earlier, the approach shown below is based largely on the methods described in SPAR-H (NUREG/CR 6883)[8]. A somewhat similar calculation approach to human IPL validation has been used by Dow (Stack, 2010),[10] but it does not allow correction of the estimate for poor human factors.

For a given human response IPL (triggered by an alarm or some other call for action), this validation approach consists of the following steps:

1. Estimate the time required for successful response (which must be less than MART, as defined earlier). This is estimated from the following aspects of the response:
    a. Estimate the time for instrumentation or/or operator detection.
    b. Estimate the time for operator decision making (this is normally small compared to operator action).
    c. Estimate the time for operator action and verification
    d. Fill in the appropriate factor in the *USED* column in Table 5 for the "Available Time" row at the table
2. Rate or scale the other Human Factors based on the descriptions in the columns. Fill in the appropriate factor in the *USED* column for each Human Factors
3. Fill in the number of practices per year, which includes number of actual alarms responded to per year and number of times this alarm or a very similar alarm is drilled/practiced. The formula used to get to the right Practice Factor, starting with a value of 0.0008 as the lowest baseline error rate possible, is:

    = (250/(number of practices per year))^0.6+0.4*((number of practices per year)/1000)^2
4. Calculate the overall human unreliability which gives the PFD (normally 0.1) for the human IPL. If the number is > 0.1 then no PFD is allowed.

This PII method takes about 15 minutes for validation of a PFD for one human IPL. This assumes the analysis is performed by a Subject Matter Expert (SME) who is trained in this method, is a human factors expert, and has access to the site data needed, which may in turn require solicitation of expert judgment (such as for the estimate of "time to respond"). Note that

the time invested for validation by this method is comparable to validating one human IPL using Site-specific data, presented in Section 5.3, indicated that simply validating the actual responses to alarms (in drills) is likely the better approach.

NOTE: Approach and calculation method above is copyrighted to PII, 2017.

5.3 Validation of Human 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.*

This appendix provides an example of the data needed for adequately counting the human in a LOPA (and other risk assessments) using Site-specific data for validation. One key focus of the appendix 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 appendix to provide the reader with some examples of how a simple data collection and validation method can be set up within their companies.*

If a site has a very good system for reporting and investigating near-misses, then this system can be used to find site-specific data for human errors (including both IEs and failure of human IPLs). Getting high near-miss reporting rates is covered in other research and papers (Bridges 2008, 2012, etc.)[11]; note the ratio of near-misses to loss events likely needs to be higher than 15 to provide sufficient data for validation using near-miss data alone.

Another way to collect site-specific data on error rates is to measure error rates 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 practice is not commonplace in the chemical industry, but the US Nuclear Regulatory Agency (NRC) in 10 CFR 55.45 and 55.59[12] 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 appendix.

EXAMPLE OF SITE-SPECIFIC DATA FOR HUMAN RESPONSE IPLs

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 Sections 5.1 and 5.2 of this paper for a discussion of these alternate approaches.

On the following pages are the options for validating human IPLs by direct measurement including (1) 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.

Approach to Using a 100% Individual Test Plan 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 walk-throughs 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). 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 response other 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 10^{-1} value for the PFD of all of these identified human response IPLs:

1. *Determine the number of tests 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 discrete tests (20 operators X 130 human response IPLs = 2600 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 U.S. NRC regulation for control room operators (10 CFR 55.45 and 55.59)[12] 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 for the fact that response to one alarm is normally similar to response to other alarms, so the operators are 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), since the basic control room regimen of testing each operator with each action each year was essentially the same in the 1970s as it is 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.) (Note that testing only a sample of these responses, discussed later, would reduce the load considerably.)*

ACTUAL EXAMPLE of Using Tests/Drills (Site-Specific Data Collection) to Validate Human Response IPLs

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 in fact validate human response using this method. Recent research (Bridges, 2010 and 2011)[13, 14] by three chemical companies documented the effort required to validate human response IPLs using Site-specific data. The following is an excerpt of the research results.

Validation Setup: A simple test was used 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.

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 tests 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. Below is an example of such an index card.

Figure 3 Example of card used to administer validation of a single human IPL

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	

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.

*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 one minute to record the data. The validation was performed by multiple operators on multiple shifts. 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 of their other duties. The total time for the test varied, but the two 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 (i.e., time to administer the test).

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 LOPA, the PFD is limited to a value of 0.1).

Table 6 provides a sample of the site-specific data for several similar human IPLs from the three sites (one in Malaysia, one in Canada, and one in the USA). For the Malaysia and Canada sites, the data was from the same operating area consisting of multiple operators across four rotating shifts of eight hours per shift. For the USA 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 humans 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 responses can be grouped together into general types of response IPLs. Then a lesser percentage will need individual validation drills.

Table 6: Site-Specific Validation of Human Response IPLs

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 one or two 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 valve wrench or other tools are available, etc.).

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

ADJUSTMENT for STRESS: As mentioned, these data (as with all drills) are 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 (Gertman, 2005)[8] 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.

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

It is important to ensure that a 10^{-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. U.S. 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).

There is resistance by some to the idea of only testing a sample of the human response pairs, with the argument that not all humans are the "same." Of course all humans are not "exactly the same," but every site eventually agrees that an individual is competent enough to be allowed to perform a task on their own (without direct supervision). This is the level of sameness that is sufficient for sampling; in other words, all operators in a unit who are competent enough to be allowed to operate independently can likely perform troubleshooting and others aspects of a human IPL within 90% of the skill of other operators.

Another argument is that valuable drills could be missed by many individuals if the sample plan is too small. This argument is focused on the training benefit derived while doing the validation drills for a human IPL. But the learning achieved by doing many drills is much greater than the training of the individuals/alarms included in the sample size alone. Unlike machines, humans learn by (1) doing similar activities, (2) giving the tests to others and watching/scoring them, and (3) by watching a test if they are neither the official observer nor the one being tested.

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[15], 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 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 that hardly moves above 350 in the 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 the equation below:

Sample Size (SS):

Equation 1: 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., ±5%)

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

Equation 2: Validation SS (finite population) = $$\frac{SS}{1 + \frac{SS - 1}{population}}$$

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." For human IPLs, the group is a combination of similar operators (likely all "qualified, independent" operators can be treated equally for this purpose, since the company has likewise made the judgment that the operators are "qualified") and similar responses to similar triggers. Creating this grouping takes careful consideration by a multi-disciplinary team (with heavy emphasis on the operators in 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/responses 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 is 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 text, such as Walpole, 2006, for this derivation). Another recognized standard for determining the sample size and the pass/fail targets is to use ANSI Z1.4[16]; this standard is now incorporated by reference with U.S. 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 2600.

Case A: For one case, assume that all actions and all operators are equivalent, 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 accounting 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 A that operators are about 95% equivalent with respect to alertness and to experience with such response actions. 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, so the validation would likely only take .35% of a staff year per operator, or six hour per year per operator (a small investment in time for validation of IPLs). For this sample of 91 validation tests, if six 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 accounting 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 7 provides various sample sizes versus confidence values and relates these to the maximum number of failures (pass/fail target) using standard statistical methods (such as Equation 1, 2, and those from standard statistical handbooks). This table includes the data for Case A and B above.

Similar values can be obtained using the look-up tables and graphs in ANSI/ASQC Z1.4[16]. This approach is based on a binary distribution, and the operating curves and acceptable quality levels have slightly different meanings than confidence levels and confidence intervals for normal distributions, such as used for Case A and B above.

Also, using MIL-STD-105E (1989) adaptation of ANSI Z1.4, a repeat of Case A requires similar steps and produces identical results (not shown here).

Table 7: 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

6. Conclusion

Operator response to an alarm can be a valid safeguard and even an IPL, of the approach for identifying the necessary action is valid, and if there is a procedure (trouble-shooting guide) and training. But in addition, the proper response action and speed must be proven in the field as well; otherwise the alarm response will not happen as hoped. The steps outlined in this paper will ensure the probability of proper operator response is high.

7. Acronyms Used

AIChE– American Institute of Chemical Engineers
ASM – Abnormal Situation Management consortium
CCPS – Center for Chemical Process Safety (a division of AIChE)
CMA – Chemical Manufacturer 's Association, now American Chemical Council (ACC)
HAZOP – Hazard and Operability Analysis
IPL - Independent Protection Layer
JSA – Job Safety Analysis
LOPA – Layer of Protection Analysis
MART – Maximum Allowable Response Time
MIL – Military
MOC – Management of Change
NRC – Nuclear Regulatory Commission (USA)
NUREG – Nuclear Regulation, US Nuclear Regulatory Commission
OSHA – Occupational Safety and Health Administration, US Department of Labor
PHA – Process Hazard Analysis
PFD – Probability of Failure on Demand
PII – Process Improvement Institute, Inc.
P&ID – Piping & Instrumentation Diagram
PSI – Process Safety Information
PSM – Process Safety Management
PST – Process Safety Time
SOP – Standard Operating Procedure
STD – Standard

8. References

1. "Layer of Protection Analysis (LOPA) Guideline," CCPS/AIChE, 2001.
2. "Guidelines for Initiating Events and Independent Protection Layers and Initiating Events," AIChE/CCPS, 2015.
3. Bridges, W., & Dowell, A., "Identify SIF and Specify Necessary SIL, and other IPLs, as part of PHA/HAZOP," *12th Global Congress on Process Safety*, AIChE, 2015.

4. *Online Database of IPLs and IEs*, Process Improvement Institute, Inc., 2017 [pending].
5. Bridges, W., "LOPA and Human Reliability," *6th Global Congress on Process Safety*, AIChE, 2010.
6. Bridges, W and Tew, R., "Best Practices for Writing Operating Procedures and Trouble-Shooting Guides;" *13th Global Congress on Process Safety*, AIChE, 20106
7. "Guidelines for Chemical Process Quantitative Risk Analysis," AIChE/CCPS, 2000.
8. 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.
9. 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.
10. Stack, R. and Delanoy, P., "Evaluating Human Response to An Alarm for LOPA or Safety Studies," *6th Global Congress on Process Safety*, AIChE, 2010.
11. Bridges, W., "Gains in Getting Near Misses Reported," *8th Conference, ASSE-MEC*, Bahrain, 2008; and *8th Global Congress on Process Safety*, AIChE, 2012.
12. "Training Requirements for Nuclear Power Plant Operators," 10 CFR 55.45 and 55.59, US Nuclear Regulatory Commission.
13. Bridges, W., "LOPA and Human Reliability – Human Errors and Human IPLs*;" 6th Global Congress on Process Safety*, 2010.
14. Bridges, W., Clark, T, "LOPA and Human Reliability – Human Errors and Human IPLs (Updated)*;" 7th Global Congress on Process Safety*, 2011
15. Walpole, R, et. al., *Probability & Statistics for Engineers & Scientists (8th Edition)*, Prentice Hall; March 5, 2006.
16. MIL-STD-105E. US Department of Defense; Military Standard: *Sampling Procedures and Tables for Inspection by Attributes*, 10 May, 1989. MIL-STD-1916 supersedes MIL-STD-105E, 1996.