95696

LA-UR-98-2649

*Approved for public release; distribution is unlimited.*

Title: Development of an Integrated System for Estimating Human Error Probabilities

Author(s): Jack L. Auflick, Heidi A. Hahn, Jerome A. Morzinski, TSA-9

Submitted to: DOE OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION (OSTI)

RECEIVED
DEC 21 1998
OSTI

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

**Los Alamos**
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form 836 (10/96)

# DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

## DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

# Development of an Integrated System for Estimating Human Error Probabilities

Jack L. Auflick*, Heidi A. Hahn and Jerome A. Morzinski

## Abstract

This is the final report of a three-year, Laboratory Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). This project had as its main objective the development of a Human Reliability Analysis (HRA), knowledge-based expert system that would provide probabilistic estimates for potential human errors within various risk assessments, safety analysis reports, and hazard assessments. HRA identifies where human errors are most likely, estimates the error rate for individual tasks, and highlights the most beneficial areas for system improvements. This project accomplished three major tasks. First, several prominent HRA techniques and associated databases were collected and translated into an electronic format. Next, the project started a "knowledge engineering" phase where the expertise, i.e., the procedural rules and data, were extracted from those techniques and compiled into various modules. Finally, these modules, rules, and data were combined into a nearly complete HRA expert system.

## Background and Research Objectives

Human reliability analysis (HRA) is defined as the prediction and evaluation of work-oriented human performance in probabilistic terms, using indices like error likelihood, probability of task accomplishment, or response time. As an empirical approach, HRA is applied to activities having a goal, a set of fixed procedures that personnel perform to accomplish that goal, and some consequence of the performance that can be used to determine success. Systematic HRAs allow analysts to examine human-machine relationships, identify error-likely situations, and provide estimates of relative frequencies for human errors on particular critical tasks, highlighting the most beneficial areas for system improvements.

The output for a given HRA, i.e., specific individual human error probabilities (HEPs) is generally combined with component reliabilities according to the configuration of the system and its operational flow. Functional sequences of events are then modeled by

*Principal Investigator, e-mail: auflick_jack_l@lanl.gov

failure paths through event or fault trees, where the HEPs are combined with the component failure probabilities to synthesize an estimated failure rate for the system of concern.

Today HRA consists of at least 40 different methodologies that are used to analyze, predict, and evaluate human performance in probabilistic terms. The HRA practitioner is confronted with a bewildering set of choices arising from the different philosophical approaches of the various techniques. As a result of this confusing array, a majority of analysts tend to rely on the most common methods, such as the Technique for Human Error Rate Prediction (THERP) [1], two methods based on Human Cognitive Reliability (HCR) [2], or the four (pre-accident screening and nominal and post-accident screening and nominal) Accident Sequence Evaluation Program (ASEP) techniques [3], despite the fact that the chosen technique may not be appropriate for that particular analysis. Lack of human performance data, unskilled analysts, and the poor selection or improper application of methodology all can produce invalid probabilistic estimates of human performance, which in turn could have severe consequences related to the occurrence of unanticipated accidents causing injury, death, or property damage.

To address these situations, Los Alamos National Laboratory (LANL) has a newly developed, HRA-centered, knowledge-based expert system that will help with the estimation of HEPs for specific human error actions during the quantification phase of a risk assessment. This expert system has several distinct advantages. First, the expert system combines the "expertise" of several HRA experts and associated data for human error rates. In doing this, it also helps disseminate that problem solving expertise to minimally trained HRA end users that have been tasked with doing HRA. In addition, this knowledge-based system helps improve the validity and reliability of a given HEP by standardizing the conclusions for a given set of data. Next, because the software already contains the various rules and data, it can free the end user from repetitive routine HRA quantification tasks (e.g., thumbing through numerous techniques and pages of data to find the most relevant HEP, given the situation of interest). Additionally, by combining the most common techniques, their rules, the human performance data, and HRA expertise, the expert system helps to codify HRA quantification techniques for future users. Finally, this knowledge-based expert system provides an effective training tool for less experienced HRA practitioners.

At this time, this expert system includes the well known core HRA techniques mentioned above, i.e., HCR, ASEP, THERP, and a newly developed Bayesian method [4]. Using Bayes' Theorem permits HRA analysts to combine expert opinions, probabilistic estimates, and real plant-specific failure data, producing a refined HEP that captures the operational environment within the specified facility or system. This expertise is contained in approximately 1500 if/then rules with about 200 various qualifiers and variables. Additionally, this tool contains numerous hypertext links that will enable end users to access various HRA definitions, assumptions, and rules associated with a specific technique or the derivation of the final HEP.

As each methodology was modeled and added to the core, it went through an extensive validation process that had several critical goals. First, it ensured that the knowledge engineering had extracted the correct rules from the various techniques. Next, the validation ascertained whether those rules were fired in a sequence that resulted in a valid HEP (based on the underlying assumptions within that technique). Finally, the validation process verified that the inference engine driving the expert system produced the mathematically and theoretically correct results.

As the prototype system moves toward completion, this HRA quantification tool: (1) affords users the opportunity to access available HEPs relevant to their situation, (2) helps determine a best match between available data and specific situations, and (3) provides one or more relevant HEPs with subjective confidence estimates regarding their utility in the present circumstances. Finally, as envisioned, future improvements could incorporate other viable HRA techniques as well as various risk assessment tools, such as a graphical user interface capable of drawing fault and event trees. Exactly what that suite of tools should contain remains to be determined.

## Importance to LANL's Science and Technology Base and National R&D Needs

The development of this Human Reliability Assessment (hereafter referred to as HuRA) expert system interacts with several of LANL's and DOE's core competencies. Specifically, LANL's *Analysis and Assessment* core competency integrates basic theory and experimental data across multiple disciplines to allow independent and unbiased assessments of complex systems. HuRA's creation addressed the development of new methods for performing a HRA, which was called out as a potential area of study in the Human-System Integration thrust area. In addition, HuRA's evolution led to the creation of a new, validated Bayesian HRA technique [4].

As a result, this system will be particularly useful to analysts at LANL where the scientific management of the United States' stockpile of nuclear weapons requires assessing human performance in areas where "real" data are hard to find (e.g., weapon disassembly). Having a workable expert system to produce enhanced human performance estimates and the ability to use those estimates in modeling the human's contribution to system performance can result in refined and defensible risk assessments, safety analysis reports, and hazard assessments within the LANL and DOE missions. This addresses the Laboratory's core competency in *Theory, Modeling, and High Performance Computing* . Because of its flexibility and ability to generalize this unique tool, HuRA also offers the same benefits to the United States Nuclear Regulatory Commission, the Department of Defense, or any other general industrial or manufacturing industry that uses risk and hazard assessments as part of our government's regulatory process.

## Scientific Approach and Accomplishments

An increasing number of human factors experts around the world are reaching the same conclusion, i.e., that human error is the principle cause of most catastrophic accidents within complex systems. As can be seen in Figure 1, these same experts estimate that in today's complex operational environments, fully 90% of all accidents can be attributed to human error. Because these mistakes made by people are often the drivers for risk in complex systems, it is very important to be able to identify and model error-likely situations and to provide a quantitative estimate for the likelihood of human error actions.

HRA then, is the collection of methodological tools used to analyze, predict, and evaluate work-oriented human performance in complex systems. As a diagnostic tool, HRA can be used to identify those factors in the system that lead to less than optimal human performance and can estimate the error rate anticipated for individual tasks. In a given system or subsystem, HRA can also be utilized to determine where human errors are likely to be most frequent. While HRA had its beginnings in the reliability engineering paradigm of the 1950s, today's HRA practitioner generally uses a basic eleven-step methodology [5] shown in Figure 2 and summarized below.

1. Select the risk analysis team (usually composed of risk assessment and HRA analysts) and train them on relevant plant functions and systems.
2. Familiarize the risk assessment team with the plant through the use of system walkdowns, simulator observations, etc.
3. Ensure that the full range of potential human actions and interactions is considered in the analysis.
4. Construct the initial model (using event and fault trees) of the relevant systems and interactions.
5. Identify and screen specific human actions that are significant contributors to the safe operation of the plant. This is done with task analyses, time-line analyses, observations of operator performance in the plant and in the simulator, evaluations of the human-machine interface and quantitative screening analyses.

6. Develop a detailed description of the important human interactions and associated key factors necessary to complete the plant model. This description should include the key failure modes, an identification of errors of omission/commission, and a review of relevant performance shaping factors.
7. Select and apply appropriate HRA techniques for modeling the important human actions.
8. Evaluate the impact on system performance using the significant human actions identified in Step 6.
9. Estimate error probabilities for the various human actions and interactions, determine sensitivities, and establish uncertainty ranges.
10. Review results (for completeness and relevance).
11. Document all information necessary to provide an audit trail and to make information understandable.

Accurate identification, modeling, and quantification of human error actions can identify root causes of human error and point to measures to improve system safety and performance. Such improvements may reduce system risk by several orders of magnitude. Given this, HuRA (LANL's HRA-centered expert system) was designed to be an effective tool for analysts during steps five through nine of the HRA process described above. Depending on the way a user answers queries from HuRA, the expert system's inference engine helps analysts describe, characterize, and quantify individual human error actions as well as determines an appropriate HRA technique (based in part on the amount of available human factors data and the expertise of the analyst). HuRA's resulting human error probability (HEP) also considers the operational characteristics of the system, the task to be performed, operator characteristics, and various performance shaping factors that either increase or decrease the likelihood of human error in situations of interest.

HuRA's expertise derives from its extensive HRA knowledge base and the data that was derived during the knowledge engineering phase of development. As mentioned earlier, this knowledge engineering activity collected seven of the most common and widely used HRA techniques, as well as creating the new Bayesian method, and then extracted the various procedural rules, assumptions, and human error data from each technique. This information was then reformed into the assorted qualifiers and variables within the knowledge base. As it exists today, that knowledge base has approximately 1500 if–then rules that generate the queries which are, in turn, inputs that drive the inference engine.

HuRA's reasoning ability stems from the interaction between user input, the knowledge base, and the internal inference engine. In the existing prototype, the inference engine uses answers from queries (i.e., the if-then rules derived from the qualifiers and variables) to select a technique and calculate an acceptable HEP. Drawing inferences, or reasoning, is driven in a backward chaining mode. This simply means that the inference engine first tries to reach one of eight goal states, e.g., the correct calculation of a HEP for each of HuRA's eight techniques. If there is insufficient information to reach a goal, then HuRA "backs up" (i.e., backward chains) to the first qualifier and begins to sequentially fire each of the 1500 rules using the answers from the queries to collect data. When sufficient information can be extracted from the answers to the queries, HuRA reaches a conclusion where it calculates a practicable HEP and provides an estimate of "confidence" in the result. As used here, confidence is a subjective measure of correctness or accuracy ranging on a scale of 0 (no confidence) to 10 (very high confidence from almost perfect data), given the operational characteristics of the system, the task to be performed, operator characteristics, and various performance shaping factors. When a user has more human performance data coupled with human factors and risk assessment expertise, confidence in the result is higher (in a range from 7-9) than when the analyst has little data and little or no human factors and risk assessment experience. In the event that the qualifiers don't provide sufficient information, HuRA asks the user for specific input, which will then be used to reach a goal state.

In the development of this knowledge-based expert system, HuRA also had to deal with several problem areas within the domain of HRA. In the past, HRA has been criticized with respect to questions about its predictive validity [6], i.e., that HRA is able to predict, within tolerable margins of uncertainty, the actual number of accidents that could occur in specific situations. This criticism stems in part from the many of the HRA techniques' lack of operational data from real situations. On the other hand several reports show that the use of THERP-based techniques (i.e., THERP and ASEP) and carefully structured sessions of expert elicitation can produce acceptable levels of predictive validity [7,8]. HuRA deals with this situation by including the new approach that utilizes Bayes' Theorem in conjunction with 1) a derived HEP using one of the seven core techniques, 2) the data from expert elicitation, and 3) real, plant-specific, operational failure data [4].

Bayes' Theorem uses deductive reasoning to weigh prior information with empirical evidence allowing HuRA users to calculate probabilities of causes based on the observed effects. Specifically, Bayes' Theorem states that the odds in favor of a given hypothesis, after a piece of data is acquired (the posterior probability), should be equal to the prior odds (our knowledge about the "hypothesis" before obtaining plant-specific knowledge), multiplied by the likelihood ratio (the conditional probability of the additional data given the hypothesis). Using Bayes' Theorem within the context of HRA and HuRA provides an optimal model of how revisions to probabilistic HEP estimates should be carried out as plant specific operational data are factored into the analysis. HuRA's Bayesian approach has been validated and presented at the Energy Facility Contractors and Owners Group (EFCOG) Conference [9] and at the International Probabilistic Safety Assessment and Management Conference [4], as well as in a prototype demonstration at the 40th Annual Meeting of the Human Factors and Ergonomics Society [10].

In addition to worries about predictive validity, there are also generic concerns about HRA's convergent validity, a measure of correlation between the results of a HRA carried out by using different methods or between different analysts using the same methods. For example, The Human Factors Reliability Benchmark Exercise [11] found that quantitative results from its validation exercise differed by a factor of 10 or more between teams using the same technique or between teams using different HRA methods. Apparently, unskilled analysts and the poor selection or improper application of methodology can produce invalid or questionable probabilistic estimates of human error. In its review of HRA convergent validity studies. The European Advisory Committee on the Safety of Nuclear Installations (ACSNI) addresses this problem by stating: "It seems likely that the discrepancies between different assessors could be reduced by training, practice, and feedback…" [6].

HuRA's design and implementation also address these issues. Specifically, HuRA contains a large number of hypertext links that will, in the case of questions, take users to moderately detailed explanations of underlying assumptions, methodology, warnings, limitations, or HRA jargon. In addition, most of the confusing assumptions and differing HRA philosophies for each technique were extracted during the knowledge engineering phase of development and were then incorporated into the qualifiers, variables, and the 1500 if-then rules in the knowledge base. This functionality was intended to provide essential information, if needed, for the less-than-experienced end user that may not have the requisite training and experience to be using a particular HRA technique. Although not yet implemented, HuRA has the ability to develop context-sensitive Help screens for each variable and rule in the knowledge base. When implemented, this functionality could provide further on-line guidance with respect to the information required to answer specific queries generated by the inference engine. Finally, as part of the implementation phase, most of the confusing assumptions and bewildering sets of rules were implicitly incorporated into the qualifiers, which in turn are often explicitly presented to the user in a simple binary Yes-No fashion or in various tables. As the user answers the Yes-No questions, or selects an alternative from a list of choices, the inference engine begins to "reason", based on the user's inputs.

In its essence then, the expert system combines the "expertise" of several HRA experts and passes that problem solving ability to minimally trained (or even expert) HRA end users that have been tasked with doing HRA. Depending on the way that users answer the queries, HuRA determines which technique is appropriate and proceeds toward the goal of deriving an appropriate HEP. If the user answers "I don't know", i.e., indicating that they have insufficient task analysis and human factors information to answer the query, HuRA by default, tells the user that they don't have the requisite information to be using the selected technique and then returns a screening HEP (i.e., a very conservative, high numeric value of 0.005, 0.05, or 0.5) depending on whether the human error action was either skill-, rule-, or knowledge-based behavior [12, 13]. By trying to reduce end user confusion and error, and by providing implicit HRA expertise, this knowledge-based system helps to improve the validity and reliability of a given HEP by standardizing the conclusions for a given set of data. Additionally, because the software already contains the various rules and data, it frees the end user from repetitive, error prone, HRA quantification tasks (e.g., thumbing through numerous techniques and pages of data, trying to understand the rules and assumptions, then applying those rules and assumptions with the various performance shaping factors from the situation of interest to find the most appropriate HEP). Additionally, by combining the most common techniques, their rules, the human performance data, and HRA expertise, HuRA helps to codify its core HRA quantification techniques for future users and as a result of this, it provides an effective training tool for less experienced HRA practitioners.

At this point HuRA, LANL's knowledge-based HRA expert system appears to be a unique, flexible, generalizable, state-of-the-art risk assessment tool within the risk assessment and human factors community. This HRA quantification tool provides end users with the ability to access available HEPs relevant to their situation, providing pertinent HEPs with confidence estimates regarding their utility in the present circumstances. In doing so, HuRA also helps determine a best match between available data, specific situations, and an appropriate HRA methodology. As a result, this system will be particularly useful to analysts at LANL where the scientific management of the United States' stockpile of nuclear weapons requires assessing human performance in areas where "real" data are hard to find (e.g., weapon disassembly). Having a useable expert system to produce enhanced human performance estimates and the ability to use those estimates in modeling the human's contribution to system performance augments the validity and reliability of estimates of human error within a system. As a result of this, HuRA will enhance, refine, and help defend various risk assessments, safety analysis reports, and hazard assessments within LANL and DOE as well as in other general industrial or manufacturing industries that are required to use risk and hazard assessments as part of our government's regulatory process.

## Publications

1. Auflick, J., Hahn, H., and Morzinski, J., "On-going Development of an Expert System for Estimating Human Error Probabilities," Symposium on Probabilistic Risk Assessment for Nuclear and Environmental Safety, the American Society of Mechanical Engineers, San Diego, 1998 (in press).

2. Auflick, J., Hahn, H., and Morzinski, J., "Development of a Prototype Knowledge-Based System for Estimating Human Error Probabilities," European Safety and Reliability Conference, Trondheim, Norway, 1998 (in press).

3. Auflick, J., Hahn, H., and Morzinski, J., "HuRA! – The HRA Knowledge-Based System for Quantifying Human Errors," Fourth International conference on Probabilistic Safety and Management (PSAM), New York, 1998 (in press).

## References


[1] Swain, A. and Guttman, H., "Handbook of Human Reliability Analysis with the Emphasis on Nuclear Power Plant Operations," NUREG/CR-1278-F, U. S. Nuclear Regulatory Commission, (1983).

[2] Hannaman, G. W. et al., Human Cognitive Reliability Model for PRA Analysis, NUS-4531. Electric Power Research Institute, Palo Alto, CA, (1984).

[3] Swain, A., Accident Sequence Evaluation Program Human Reliability Analysis Procedure, NUREG/CR-4772. U. S. Nuclear Regulatory Commission, Washington, (1987).

[4] Auflick, J., Morzinski, J., Houghton, F., and Eide, S., "An Efficient Human Reliability Assessment Using Soft Data and Bayesian Updates," Probabilistic Safety Assessment and Management, Vol. 1, 529-534, London: Springer (1996).

[5] Kelly, D., Auflick, J., and Haney, L., "Assessment of ISLOCA Risks – Methodology and Application: Westinghouse Four-Loop Ice Condenser Plant," NUREG/CR-5744, U. S. Nuclear Regulatory Commission, Washington, (1991).

[6] Broadbent, D., et al, "Advisory Committee on the Safety of Nuclear Installations, Second Report: Human Reliability Assessment – a critical overview," London: HMSO (1991).

[7] Kirwan, B., "A Comparative Evaluation of Five Human Reliability Assessment Techniques," in Sawyers, B., (ed.), Human Factors and Decision Making, London: Elsevier (1988).

[8] Williams, J., "Validation of Human Reliability Assessment Techniques," Reliability Engineering, Vol. 11, 149-162, (1985).

[9] Auflick, J., Morzinski, J., Houghton, F., and Eide, S., "A Streamlined HRA for LANL's TA-55 FSAR," LA-UR-274, Proceedings of the Energy Facilities Owners and Contractors Group, Savannah River, GA, (1996).

[10] Hahn, H., Auflick, J., and Morzinski, J., "A Demonstration of a Prototype Knowledge-Based System for Estimating Human Error Probabilities, LA-UR-2137" Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting, Vol. 2, pp. 869, (1996).

[11] Poucet, A., et al, "Human Factors Reliability Benchmark Exercise," Euratom Report, PER 1482/88, (1988).

[12] Hannaman, G., et al, "Systematic Human Action Reliability Procedure (SHARP)," EPRI NP-3583, Electric Power Research Institute, (1984).

[13] Rasmussen, J. Human Errors: A Taxonomy for Describing Human Malfunction in Industrial Installations, Riso-M-2304. Riso National Laboratory, Roskilde, Denmark, 1981.

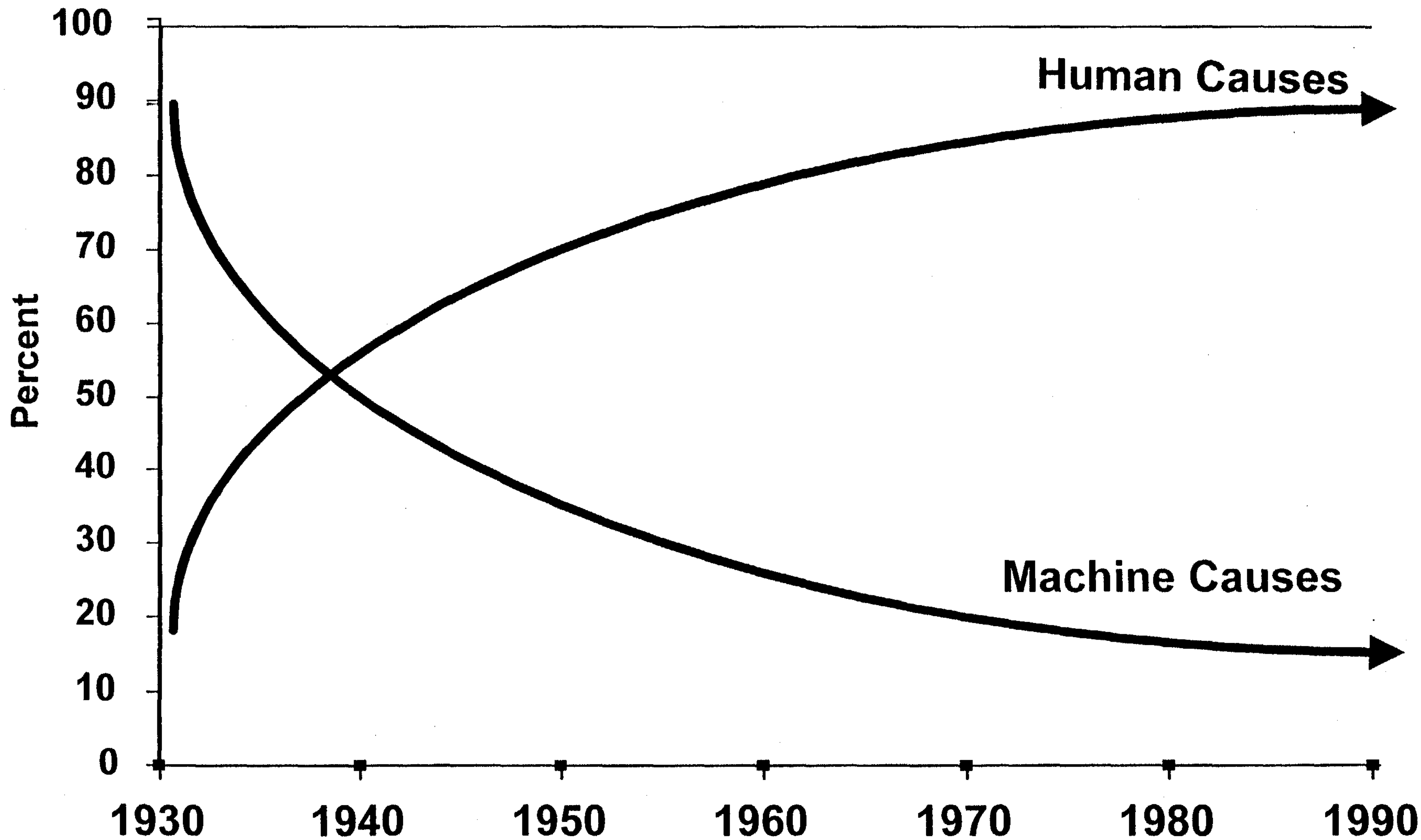


Figure 1. Causes of system failures: human versus hardware.

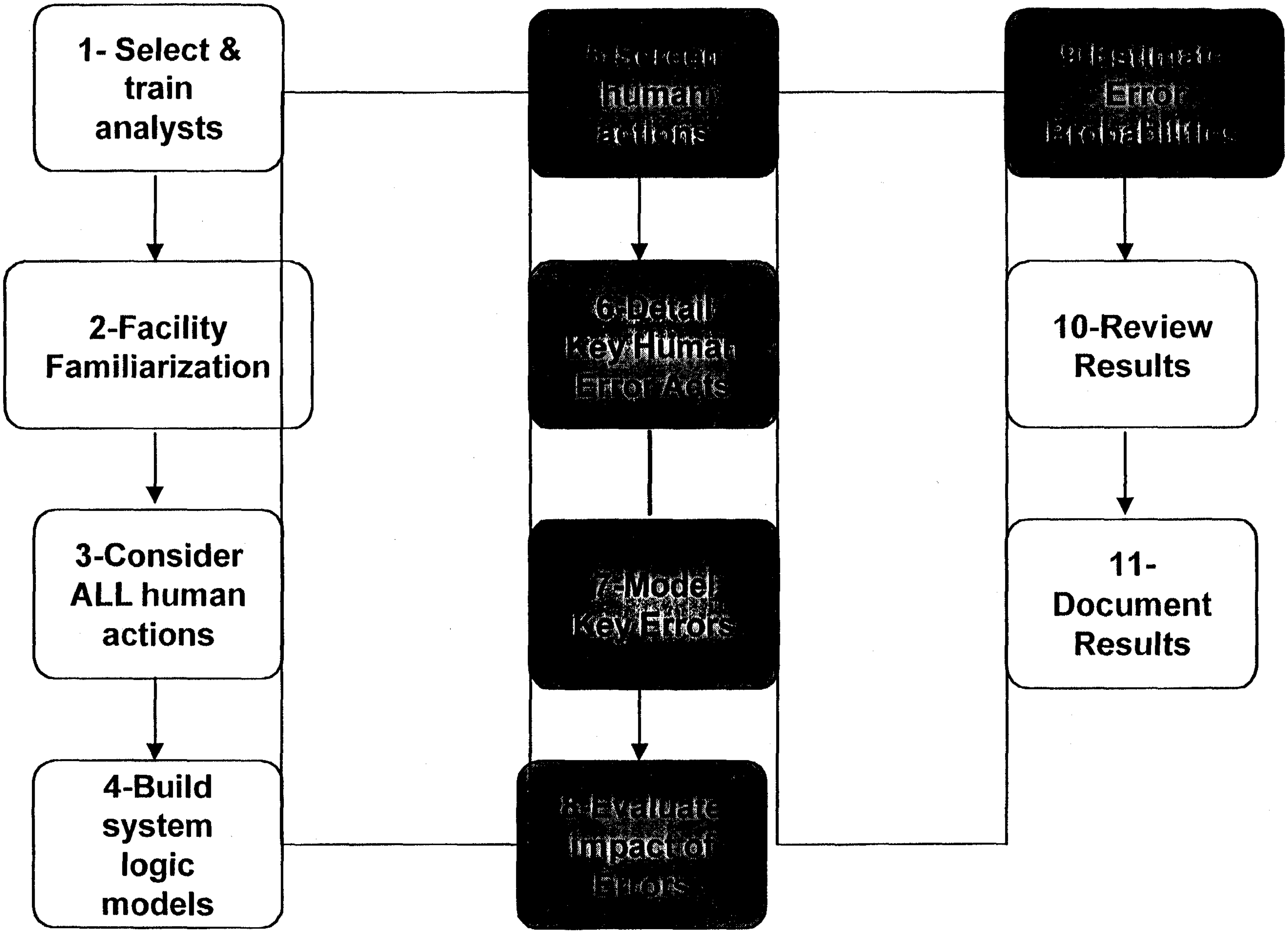


Figure 2. The 11-step HRA process.