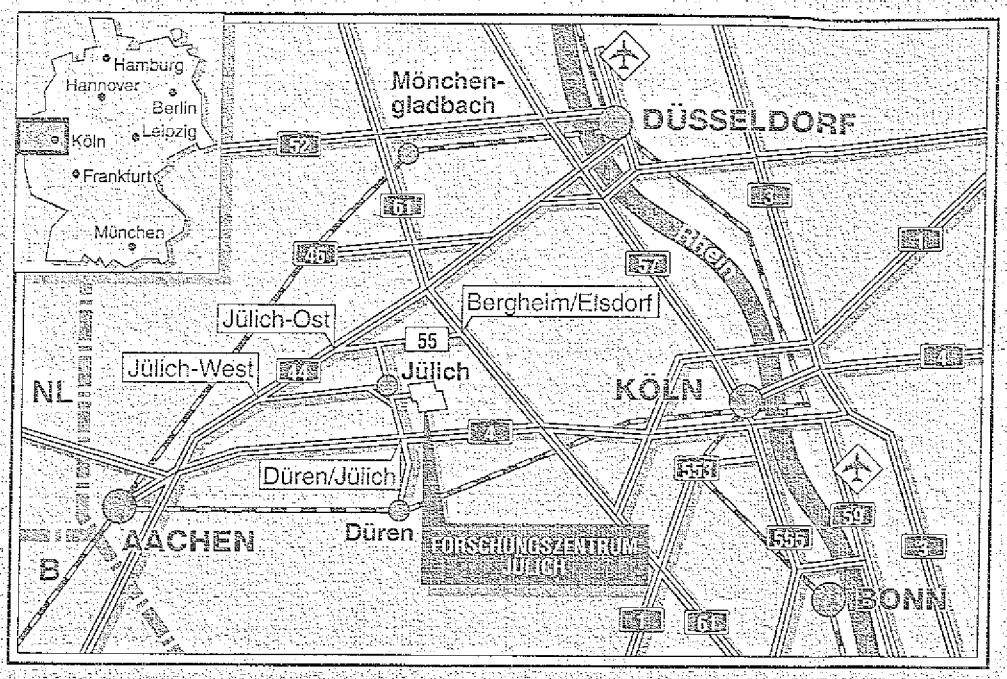
FORSCHUNGSZENTRUM JÜLICH GmbH

KFA

Institut für Sicherheitsforschung und Reaktortechnik

Evaluation of Human Reliability Analysis Methods Addressing Cognitive Error Modelling and Quantification

B. Reer O. Sträter J. Mertens



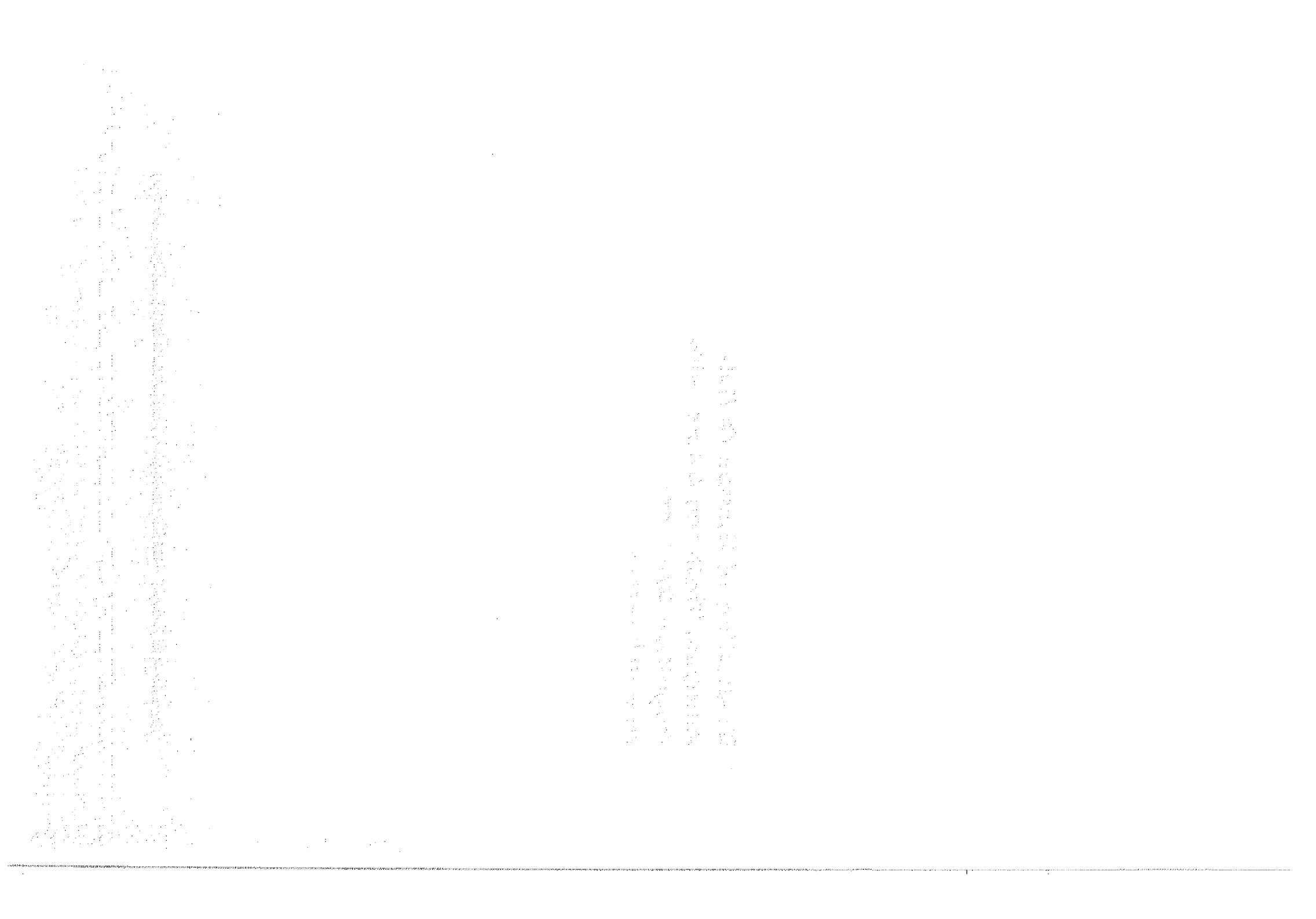
Berichte des Forschungszentrums Jülich ; 3222
ISSN 0944-2952
Institut für Sicherheitsforschung und Reaktortechnik Jül-3222

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek
D-52425 Jülich · Bundesrepublik Deutschland
Telefon: 02461/61-6102 · Telefax: 02461/61-6103 · Telex: 833556-70 kfa d

Evaluation of Human Reliability Analysis Methods Addressing Cognitive Error Modelling and Quantification

Bernhard Reer Oliver Sträter* Johannes Mertens

*Gesellschaft für Anlagen- und Reaktorsicherheit mbH (GRS), D-85748 Garching, Germany



Forschungszentrum
Jülich GmbH
ISR

Jül- 3222

May 1996

EVALUATION OF HUMAN RELIABILITY ANALYSIS METHODS ADDRESSING COGNITIVE ERROR MODELLING AND QUANTIFICATION

Bernhard Reer

Oliver Sträter*

Johannes Mertens

ABSTRACT

Actions and errors by the operating personnel, which are of significance for the safety of a technical system, are classified according to various criteria. Each type of action thus identified is roughly discussed with respect to its quantifiability by state-of-the-art human reliability analysis (HRA) within a probabilistic safety assessment (PSA). Thereby, the principal limits of quantifying human actions are discussed with special emphasis on data quality and cognitive error modelling. In this connection, the basic procedure for a HRA is briefly described under realistic conditions. With respect to the quantitative part of a HRA - the determination of error probabilities - an evaluating description of the standard method THERP (Technique of Human Error Rate Prediction) is given using eight evaluation criteria. Furthermore, six new developments (EdF'sPHRA, HCR, HCR/ORE, SLIM, HEART, INTENT) are briefly described and roughly evaluated. The report concludes with a catalogue of requirements for HRA methods.

* Gesellschaft für Anlagen- und Reaktorsicherheit mbH (GRS), D-85748 Garching, Germany

Forschungszentrum
Jülich GmbH
ISR

Jül- 3222

Mai 1996

AUSWERTUNG VON METHODEN FÜR DIE ANALYSE MENSCHLICHER ZUVERLÄSSIGKEIT HINSICHTLICH MODELLIERUNG UND QUANTIFIZIERUNG KOGNITIVER FEHLER

Bernhard Reer

Oliver Sträter*

Johannes Mertens

KURZFASSUNG

Handlungen und Fehler des Betriebspersonals, die für die Sicherheit eines technischen Systems von Bedeutung sind, werden nach verschiedenen Kriterien klassifiziert. Zu jedem dadurch identifizierten Handlungs- bzw. Fehlertyp wird in Ansätzen diskutiert, inwieweit er sich nach dem derzeitigen Kenntnisstand in der Analyse menschlicher Zuverlässigkeit (AMZ) im Rahmen einer probabilistischen Sicherheitsanalyse (PSA) quantifizieren läßt. Hierbei werden auch die prinzipiellen Grenzen bei der Quantifizierung menschlicher Handlungen diskutiert - mit besonderem Schwerpunkt auf der Datenqualität und auf der Modellierung kognitiver Fehler. In diesem Zusammenhang wird das prinzipielle Vorgehen bei einer AMZ kurz und praxisnah erläutert. Hinsichtlich des quantitativen Teils einer AMZ - der Ermittlung von Fehlerwahrscheinlichkeiten - erfolgt eine bewertende Beschreibung der Methode THERP (Technique of Human Error Rate Prediction) anhand von acht Beurteilungskriterien. Darüberhinaus werden sechs Neuentwicklungen (EdF'sPHRA, HCR, HCR/ORE, SLIM, HEART, INTENT) kurz beschrieben und grob bewertet. Abschließend wird ein Katalog von Anforderungen an Methoden zur AMZ erstellt.

* Gesellschaft für Anlagen- und Reaktorsicherheit mbH (GRS), D-85748 Garching

CONTENTS

1 Introduction into human reliability analysis (HRA) in technical systems

The analysis of human reliability within the framework of a probabilistic risk assessment (PRA) comprises the detection (identification) of safety-relevant man-machine interactions and the calculation (quantification) of the corresponding probabilities [PRA-PG]. This type of analysis offers an opportunity for concrete improvements to the man-machine conditions in a plant. Safety-relevant interactions between man and machine can be *predictively* identified *before* they cause or contribute to causing damage (see Figure 1). However, this advantage is at the same time a challenge for the analyst. Reliable and practically useful results can only be expected from a high-quality predictive analysis. This quality also depends on the application of relevant insights from psychology and ergonomics and on the application of conclusions drawn from incorrect actions in the past, for example by evaluating operational experience or by means of simulator experiments. In the case of *correct* application, it will then be possible to *realistically* distinguish between relatively likely and relatively unlikely damage opportunities.

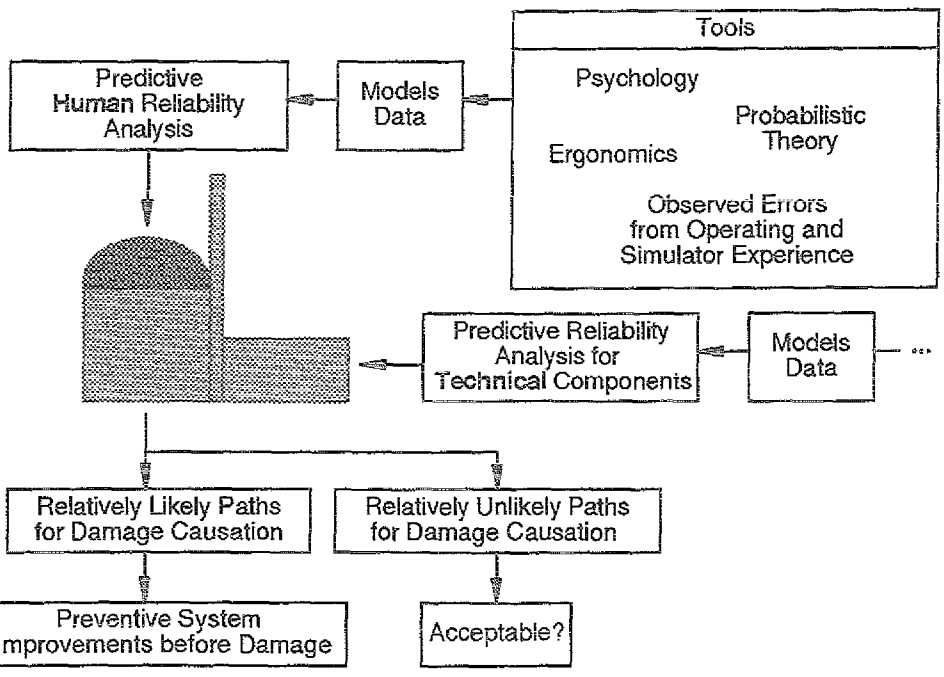


Figure 1: Interdisciplinary procedure in predictive human reliability analysis

This report describes the current status of development in the field of human reliability analysis; priority is given to the methods of assessing error probabilities during the course of an accident. Section 2 deals with the man-machine interactions, which are normally covered explicitly by a human reliability analysis within a PRA. Section 3 outlines the basic procedure

in human reliability analysis. The identification of safety-relevant opportunities for human error is the subject of discussion in Section 4. Section 5 gives a survey of different methods of quantification (for the assessment of error probabilities). Section 6 describes and evaluates the established standard method THERP (Technique of Human Error Rate Prediction). Promising new developments are discussed in Sections 7 and 8. Requirements for an HRA method are outlined in Section 9. The report ends with conclusions in Section 10.

This report is related to part of the work already carried out for [Hennings 95] and [Sträter 94b].

2 Classification of human actions and errors

Human reliability analysis (within a PRA) only comprises a subaspect from the total spectrum of human activities in a technical system, for example a nuclear power plant (NPP), i.e. actions by the *operating* personnel *after* commissioning the system. Human errors *prior to* commissioning (design, installation etc.) are assumed to be covered by the reliability characteristics of technical components, see Table 1.

Man-machine interaction	Subject of human reliability analysis?
Prior to commissioning: – design – installation – ...	No
After commissioning: – scheduled operation (incl. startup/shutdown and revision) – repair, maintenance, test, calibration – control room activity (monitoring, startup/shutdown) – accident – required and unrequired actions – actions with and without predefined procedure – actions within and beyond design limits	Yes, with restrictions, see further tables

Table 1: Human activities prior to and after commissioning a plant

For the time after commissioning a plant, activities are analysed both in undisturbed normal operation and under accident conditions. The typical results of such activities, which are of interest for a PRA, are listed in Table 2.

Situation	Man-machine interaction positive	Man-machine interaction negative	Subject of human reliability analysis?
normal	malfunction of a component detected	malfunction of a component overlooked	YES, with restrictions according to Table 3
	tested / serviced / repaired component left in proper condition	malfunction of a component brought about	
abnormal (acci-dent)	required manual action for accident control successful	required manual action for accident control fails	YES, with restriction according to Table 9
	unrequired and dangerous intervention in automatic system responses for accident control omitted	unrequired and dangerous intervention in accident control made	SELDOM (requirement of methods)

Table 2: PRA-relevant results of human activities

2.1 Activities in normal situations (prior to the occurrence of an accident)

2.1.1 Latent errors contributing to the unavailability of safety-relevant systems

Not all activities in the absence of an accident are the subject of human reliability analysis. Normally, only those activities are analysed in which errors (e.g. faulty maintenance) can occur leading to failure of a component included in the fault trees. These errors correspond to the man-machine interaction classified as type 1 errors in [SHARP] or type A in [IAEA 89]. They thus occur primarily during repair and maintenance work. Their analysis is not required if the contribution of such activities as, for example, maintenance, is already covered by the failure statistics of the component concerned. However, corresponding evidence is not always easy to furnish, especially if

- the procedure belonging to the activity, generally a written instruction for action, is new,
- the activity is very plant-specific or
- failure of the component concerned is very seldom.

A contribution to the failure of a fault tree component, quantified by human reliability analysis, is generally a latent (hidden) error (cf. [Reason 90], Chapter 7; see also Table 3): a system or component is in an improper standby condition, possibly for a prolonged period of time. Table 4 contains two examples of latent errors from operating experience with nuclear plants. In the most favourable case (Gundremmingen) such an error is detected during later inspection or checking, in the most unfavourable case (TMI) only during an accident when the system affected is urgently needed.

2.1.2 Active errors initiating accidents

Human reliability analysis rarely deals explicitly with errors initiating accidents (Table 3). Table 5 contains two examples from nuclear operating experience. Such errors are practically always assumed to be covered by the frequency statistics of accident-initiating events. Rough estimates – as, for example, in [KFA 81] for errors leading to a loss of the main heat sink – are the exception. Based on recent insights [EPS 1300] into the possible risk relevance of zero power level conditions, there is increasing interest in a precise analysis of accident-initiating errors. Zero power level conditions in nuclear power plants require a large number of additional manual actions which do not occur during power operation.

Recent American studies confirm that most (30 out of 39) of the errors recorded in zero and partial power operation resulted in the initiation of an accident ([Barriere 94], page 2-8).

Degradation of plant safety	Plant state		
	zero power, revision	partial power, startup/shutdown	full power operation
latent error (after maintenance, repair, test or calibration or control room activity): component / system undetected unavailable	covered by human reliability analysis for relevant fault tree inputs		
active error (during maintenance, repair, test or calibration or control room activity): accident is initiated	not covered by HRA with a few exceptions	not covered by HRA with a few exceptions, e.g. [KFA 81]	

Table 3: Errors covered by human reliability analysis (HRA), broken down according to plant state and degradation of plant safety

Error	Degradation of plant safety
NPP Gundremmingen, 1975 [BMI 77]: primary isolation valve closed during calibration and not re-opened	measuring channel group for differential pressure unavailable
NPP TMI-2 (Harrisburg), 1979 [Kemeny 79]: block valves closed during maintenance and not re-opened	auxiliary feedwater system unavailable

Table 4: Examples of latent errors caused by the operating personnel

Error	Degradation of plant safety
NPP Obrigheim, 1972 [IRS 77]: drain valve inadvertently opened when purging the pressurizer relief tank	loss of coolant
NPP Neckarwestheim, 1975 [Smidt 79]: demineralized-water feeding stopped too late during startup	unintended criticality

Table 5: Examples of accident-initiating (active) errors of the operating personnel

2.2 Activities in abnormal situations (during accidents)

2.2.1 Required and unrequired actions

The human reliability analysis of activities during accidents is generally confined to manual actions required for accident control. In exceptional cases, unrequired and hazardous interventions in automatic system responses are also studied, see Table 2. Evaluations of operational experience specify an amount of 22 % for unrequired and unsuitable actions; see Figure 2. Against this background, the unrequired shutdown of safety systems is also analysed for French nuclear power plants [EPS 900], [EPS 1300]. Such actions are also referred to in the PSA literature as unsuitable initiatives ([EPS 900], page 73), extraneous acts ([Swain 83], page 4-9) or mistakes that aggravate situations ([SHARP], page 2-6). Table 6 shows that unrequired actions can have both a positive and a negative influence on the course of an

accident. It cannot therefore be stated from the outset whether the risk is underestimated or overestimated if a PSA neglects such actions.

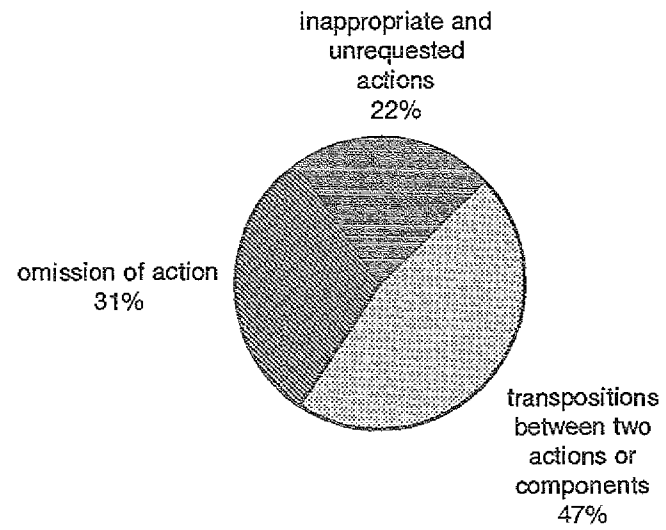


Figure 2: Percentages of different types of human errors for 51 events in nuclear power plants [Ghertman 85]

Action	Unrequired because ...	Course of action	Influence on accident sequence
NPP Obrigheim, 1972, after loss of coolant through open drain valve (TE1A10): bridging the torque limit switch of TE1A10 [IRS 72]	...automatic safety injection would have controlled the loss of coolant [Hoffmeister 74]	positive: bridging was successful, so that valve (TE1A10) could be closed	positive: loss of coolant stopped
NPP Davis Besse, 1985, after failure of the main feedwater pumps: manual start of emergency feedwater pumps [NUREG-1154]	...the emergency feedwater pumps would have started automatically later on	negative: starting fails, steam generator inadvertently isolated	negative: failure of both emergency feedwater pumps
NPP TMI-2, 1979, after loss of coolant through open pressurizer relief valve: manual disconnection of the high pressure injection (HPI) pumps [Kemeny 79]	...automatic HPI would have controlled the loss of coolant	positive: disconnection of HPI pumps successful	negative: core damage

Table 6: Examples of unrequired operator actions from operational experience with nuclear power plants

Table 7 contains examples of required manual actions during accidents, whose failure is studied within the scope of a PRA. However, their requirement (with a view to avoiding core damage) does not always result exclusively from the accident-initiating event. In many cases, an action (e.g. isolation of a defective steam generator) is only required in combination with a further malfunction (e.g. failure of the water feed system replenishing any water lost through steam generator leakage) in order to prevent core damage. In rare cases, a manual action is required exclusively due to the initiating event; e.g. shutdown after a small leak in the main coolant pipe of older plants in which the shutdown process is not automated.

Required action	Accident-initiating event			
	small leak in a main coolant pipe	leak in a heating tube of a steam generator	failure of main feedwater system	failure of the main heat sink
plant shutdown	X	X		
start of emergency system		X	X	X
isolation of defective steam generator		X		
restart of volume control system		X		

(after [DRS-B], Section 5.2)

Table 7: Important operator actions included in the DRS-B accident sequence diagrams (event trees) and quantified

2.2.2 Actions with and without predefined procedures

Not every action (or sequence of actions) suitable for controlling an accident sequence is predefined by a written instruction. Actions without predefined procedure are referred to by [SHARP] (page 2-6) as type 5 actions: »By improvising, plant personnel can restore and operate initially unavailable equipment to terminate an accident«. Such actions are basically repairs (in the widest sense) to failed components. The specific course of a repair depends on the failure cause of the component concerned, in one case only a fuse has to be replaced or an actuator pressed (e.g. manual start on failure of automatic starting device), in another the whole component must be replaced. Since there are often several failure causes, some of them unknown, *concrete* repair actions are seldom laid down in procedures and are therefore not *explicitly* included in risk studies. Several risk studies [KFA 81], [KFA 84] for plants, in which there is much time (more than about 5 h) available for repairing a failed component during an accident, explicitly include repairs. The error probability is estimated by evaluating a repair statistics documenting the repair times required for failures which previously occurred on the

component concerned. Methodological approaches to evaluation and critical comments on the informative value of such statistics can be found in the literature [Reer 94a].

The extent to which manual actions described by procedures are included depends on the scope of analysis defined in the relevant PRA. A basic criterion is the question of the extent to which recovering interventions by the operating personnel are analysed in beyond design basis accident situations ("accident severity"); this will be discussed in the following section.

2.2.3 Actions within and beyond design

Actions for accident control are classified into two groups in Germany (cf. [Roth-Seefried]; [DRS-B]):

1. actions within design basis accident control, e.g. start of emergency system after failure of steam generator main and auxiliary feeding,
2. emergency actions beyond design basis accident control, e.g. steam generator feeding by a mobile fire fighting pump after failure of main feeding, auxiliary feeding and emergency systems.

The first group is related to the *planned* operating and safety systems for accident control under the licensing procedure, so that the term "planned actions" (in German: "geplante Handlungen") is also used. Most of these actions are described in the operating procedures. There are, however, also actions within design basis accident control, which are not contained in the operating procedures, for example the manual start of a system upon failure of the automatic starting device.

There are often still safety margins available even for a failure of design basis system functions. In these cases, the accident can still be controlled by emergency actions, although this is already an *unplanned accident* according to licensing design criteria. For this reason, the terms "unplanned actions" (in German: "ungeplante Handlungen") and "accident management" are also used in this connection. The action sequences are described in emergency operation procedures, inasmuch as they can be predefined in detail. In international usage, such actions are subsumed under the term "recovery" ("... all human interactions that are not initially integrated into plant logic models") [Wakefield 92] (page 2-2).

Although emergency actions do not fall within the scope of design basis accident control, considerations have recently been under way to the effect of including such actions in the regular simulator training programme [Hoffmann 89]. According to [Swain 83] (page 12-23), a reliable diagnosis (i.e. the timely recognition of the need for an emergency action) depends on such training. The emergency exercises recently introduced in many nuclear power plants

[Hansmann 89], [Gautschi 89] have a direct influence on diagnosis reliability. Exercises of the required manual actions reduce their time requirement, so that more time is available for diagnosis and the operators can analyse the system state under less time pressure. Knowledge of the training level in the plant under consideration is thus of great significance for human reliability analysis.

The question of the extent to which such emergency actions have been included depends on the objectives of the study. In many studies (e.g. [EPS 1300]) no distinction is made (as, for example, in Germany) between emergency actions and actions within design basis accident control. A relevant evaluation of such studies therefore requires a determination of the extent to which accident sequences, for which manual operator actions for accident control were still included, have exceeded the scope of design basis accidents.

2.3 Error causation categories

Various types of errors can contribute to the failure of an action or sequence of actions. The categories of incorrect human outputs in [Swain 83] provide rough insights into error causation. Deeper insights can be found in classifications based on modern psychological error research. In this context, the error type classifications in [Reason 90] and [Hacker 86] have gained acceptance.

2.3.1 Errors of omission, errors of commission

Table 8 outlines a classification system taken from the final version of the THERP handbook [Swain 83]. Two main groups are distinguished,

- error of omission (EOM): failure to perform a task (or action), and

- error of commission (ECOM): incorrect performance of a task (or action).

It should be noted that any classification according to these groups (defined in [Swain 83]) depends on the definition of the underlying task (or action). Given a definition including a very detailed specification of the respective task, almost every error can be classified as an EOM. For example, during the Davis-Besse incident (see Table 6), an operator pushed the wrong controls labelled "SG 1-1 LOW STM PRESS" and "SG 1-2 LOW STM PRESS" while attempting to actuate the auxiliary feedwater system (AFWS), he should have pushed the controls labelled "SG 1-1 LOW WTR LVL" and "SG 1-2 LOW WTR LVL". This is an ECOM according to a task defined as ACTUATE AFWS. However, according to a task defined as

PUSH CONTROLS "SG 1-1 LOW WTR LVL" and "SG 1-2 LOW WTR LVL", this is an EOM.

To avoid such pitfalls, an ECOM should be defined as an incorrect performance of a task, given that the task is initiated.

Moreover, it should be noted that the boundaries between the ECOM categories are fluid. For example, an ERROR OF SEQUENCE could also be described by the attributes TOO EARLY or TOO LATE.

Nevertheless, the classification system in Table 8 provides valuable insights into error causation because:

- it demands that an analyst describes the task of an operator, and

- it gives an overview of the principal reasons why a task could fail.

- Errors of Omission
  - Omits entire task
  - Omits a step in a task
- Errors of Commission
  - Selection error:
    - Selects wrong control
    - Mispositions control (includes reversal errors, improperly made connections, etc.)
    - Issues wrong command or information (via voice or writing)
  - Error of sequence
  - Time error:
    - too early
    - too late
  - Qualitative error:
    - too much
    - too little

Table 8: Categories of incorrect human outputs. "Any of these incorrect human outputs may be the result of other human errors: an error of interpretation of a pattern of signals, a misreading of a display, a misprint in an emergency operating procedure, etc. In an HRA, the incorrect human outputs and human errors leading to these incorrect outputs must be analysed " ([Swain 83], page 2-16)

2.3.2 Slips, mistakes, violations

Reason's classification (Table 9) of errors - or (more general) unsafe acts - uses basic elements of the cognitive behaviour classification in [Rasmussen 79] into the following levels:

- skill-based (stored patterns of preprogrammed instructions are called from memory),

- rule-based (familiar problems are tackled with the aid of IF-THEN rules), and

- knowledge-based (novel problems are tackled by analytical approach).

This classification forms the basis in the USA for categorizing human error types in events occurring in nuclear power plants (cf. [Barriere 94], Appendix A).

With respect to these properties of human cognitive behaviour, Reason distinguishes two main groups of errors:

A. Wrong execution of a plan due to attentional failure (e.g. overlooking something) or memory failure (e.g. forgetting something): error types 1 and 2 in Table 9. These errors are summarised as "skill-based errors" and specified as "slips" and "lapses".

B. Erroneous or deliberate decision for a wrong plan: error types 3 to 7 in Table 9.

Table 9 suggests that the current methods of human reliability analysis exhibit weak points in the consideration of decisions for wrong plans. Erroneously wrong decisions are frequently only quantifiable at the rule-based level, i.e. as errors in the application of IF-THEN rules: if system state ZX, then plan PY ([Reason 90], page 43).

Knowledge-based mistakes, for example selective perception directed to a misleading signal, are to be expected in novel situations involving limited resources in conjunction with knowledge deficiencies for problem solution ([Reason 90], page 43). All methods of human reliability analysis exhibit deficiencies in the quantification of such mistakes, and models are used which greatly simplify reality (e.g. the diagnosis model in [Swain 83]).

Deliberately wrong decisions (violations) form a special category. A violation of type 5 or 6 (Table 9) is not necessarily an error in the sense that the goal envisaged by the operator is not achieved, but a deviation from the procedures (or policies) defined by designers, managers and authorities to ensure plant safety ([Reason 90], page 195).

The operator thus runs the risk in the event of an unsafe deviation from such prescribed procedures that an undesirable system state occurs with increased probability. Routine violations, such as subsequently filling in a checklist, are basically covered by normal human reliability analysis (e.g. [Swain 93], Table 20-6). There are, however, deficiencies in the quantification of violations in exceptional situations, e.g. during an accident. Some methods ([Hall 82], page 26; [Moieni 94], page 54) generally include such action tendencies by means of a so-called "reluctance factor" by which the error probability of a task involving economic

losses is increased (e.g. depressurization of the primary circuit after total failure of steam generator feeding, which results in a prolonged plant outage).

Intended sabotage is not a subject of the HRA methods described here. It is always assumed that the operator is willing and tries to avert damage from the system (be it by a wrong or correct plan).

Error types after [Reason 90] which can contribute to failure of an action (or of a sequence of actions)		Explicitly included in human reliability analysis?
action is *not* performed *as planned* (skill-based error)	(1) attentional failure (slip)	basically yes, point-type deficiencies *3)
	(2) memory failure (lapse)	
action is performed as planned, *plan* is *erroneously wrong* *1) (mistake)	(3) rule-based mistake	in part, strongly simplifying models with frequently estimated input parameters *3)
	(4) knowledge-based mistake	
action is performed as planned, *plan* is *deliberately wrong* *2) (violation)	(5) routine (frequently occurring) violation	basically yes, point-type deficiencies *3)
	(6) exceptional (seldom occurring) violation	in part, strongly simplifying models with frequently estimated input parameters *3)
	(7) act of sabotage	no

*1) Unsuitable for achieving the goal envisaged by the operator.

*2) Deviating from rules (procedures or policies). Note, in contrast to other authors, Reason does not define such deviation as a criterion for an error; cf. [Dougherty 94] and [Reason 94].

*3) The assessments concerning a consideration of these error types are vague because the boundaries – e.g. between (4) and (2) – are not always very sharp.

The error types 1 to 4 are specified by 35 subcategories outlined in [Reason 90] (pages 68-95).

Table 9: Error types included in human reliability analysis

In practical HRA application, it is difficult to distinguish between the categories in Table 9. As an example, the HCR model [Hannaman 84] differenciates between everyday actions for skill-based behaviour, actions with procedures for rule-based behaviour, and actions without procedures for knowledge-based behaviour. Nevertheless, the relation of cognitive behaviour to real objects is not a 1:1 representation and categorising into these cognitive levels is more difficult than expected. Even the application of available IF-THEN rules frequently requires knowledge-based performance, for example if the system state must be determined from several interacting process parameters. According to evaluations in French nuclear power plants, even the application of a written procedure contains knowledge-based performance elements:

> »The operator does not apply the procedure mechanically. ...Every time an operator reads a procedure he *necessarily* interprets or "re-thinks" it« ([Mosneron 92], page 631).

Considering this example, HRA has to recognize that the cognitive behaviour of human beings is not related to external properties of the situation (i.e., procedures do not always involve rule-based behaviour). Human beings are able to accommodate themselves to different situations. They learn behaviour and become skilled if they do an action several times. If an action is performed at least regularly and often, it may become highly skilled and subconscious. This enables the human being to reduce information load by performing some cognitive activity on the subconscious level and is to a wide extent independent of the observable complexity of the action. Hence, an application of a bad procedure may be a knowledge-based decision error as well as a skill-based error of habit. This depends on the situation (external cues of information) as well as on the experience and habit of the operator.

The basis for this accommodative performance is the recognize-act cycle of human information processing. If the recognition of a situational pattern does not imply an abnormality (e.g., a new situation) in the human brain, then the human will act on the subconscious or low conscious level (e.g., walking until we stumble). If the recognition does imply an abnormality then conscious processing will be involved for this abnormality. This so-called 'cognitive dissonance' is a mismatch of learned behaviour and actual situation.

This means that the level of cognitive behaviour is directly related to the amount of cognitive dissonance which is implied by the difference of the situation from a learned behaviour. The distinction of the level of cognitive behaviour is therefore more a continuum measure than a dichotomy and has to consider both the learned behaviour of the operator and the involved tasks (cf. [Wickens 84]). This aspect may be summarised in Table 10. [Sträter 95b] provides an elaborated discussion of cognitive dissonance, underlying cognitive dimensions and resulting cognitive habits.

The table indicates that the recognition of the cognitive level by the operator is normally not distinguished in HRA. The underlying assumption for this is that operators are normally well

trained (i.e., the assumption of the log-normal distribution) and differences in the operator's recognition of the cognitive level are modelled by PSFs like training and qualification.

In summary, it should be noted that the deficiencies (indicated in Table 9) considering decision-based errors in action planning do not necessarily suggest that the current HRA approach leads to unrealistic results. Not every pursuit of a wrong plan can be classified as a mistake or violation. Skill-based errors (e.g. overlooking or disregarding something) can also cause an operator to follow a wrong plan. A certain percentage of decision-based errors is therefore explicitly quantifiable by probabilities of errors in skill-based performance. The method proposed in [Moieni 94] for assessing the probability of misinterpreting a procedure step containing a diagnosis logic is based e.g. on the THERP data for rule-based actions; see Figure 13 in this report.

Moreover, it cannot be ruled out that the error probabilities used in human reliability analysis also implicitly include contributions beyond the skill-based level. A certain percentage of errors from which an error probability was calculated could be attributable, for example, to knowledge-based mistakes (such as excessive confidence in a misleading information perceived first).

Operators' recognize-act cycle situation	skill (no dissonance)	rule (dissonance)	knowledge (strong dissonance)
skill (regular and frequent everyday-action)	skill-based slip	rule-based slip	knowledge-based slip
rule (regular and frequent action with procedure)	routine violation	rule-based mistake	attentional failure
knowledge (not regular and unusual without procedure)	exceptional violation	memory failure	knowledge-based mistake

Table 10: Coherence of cognitive level and situational factors and involved error types

2.3.3 Lack of information, nonuse of information, incorrect use of information

The classification system of Hacker ([Hacker 80], revised version in [Hacker 86]) emphasises the underlying information deficiency of an error. Table 11 summarises the related categories. Similiar to the categories EOM and ECOM in [Swain 83] (see Table 8 in this report), Hacker distinguishes between NONUSE and INCORRECT USE of available information. There are also similarities (e.g., B.3 violation such as ignoring available information) to categories used in [Reason 90] (see Table 9 in this report).

In addition to the error classification systems outlined above, Hacker's system offers an important category: lack of useable information. This category allows for an advanced view of human error as man-machine interaction: human deficiencies and deficiencies of the technical systems contribute to the causation of an error. For example, it is evident that a technical deficiency (incorrect signal of valve state) was an important cause of a significant error (cutting out HPI pumps) during the TMI incident (1979); see Table 6.

(A) LACK OF USEABLE INFORMATION
(B) NONUSE OF AVAILABLE INFORMATION
(B.1) Overlooking
(B.2) Forgetting
(B.3) Ignoring (violation)
(B.4) Information reduction (stereotype response without verifying check)
(B.5) Deficiencies in information processing due to time restrictions or mental capacities
(C) INCORRECT USE OF AVAILABLE INFORMATION
(C.1) Incorrect use in orientation
(C.2) Incomplete use in goal formation
(C.3) Formation of an incorrect plan to act
(C.4) Mismatching recall of a correct plan to act

Table 11: Main headings of categories for information deficiencies as causes for human errors. Translated from [Hacker 86] (page 435). The categories C.1, C.2 and C.4 are specified by eight subcategories, see [Hacker 86] (Section 10.2) for details.

3 Basic procedure in human reliability analysis

3.1 HRA steps

Essentially, many of the proposed sequences of HRA steps have equal contents. However, according to the starting point, two different concepts appear worth distinguishing:

- starting with a human-related event (identified by fault tree analysis) [PRA-PG] or system failure of interest [Swain 83], or
- starting with a certain type of human interaction ([IAEA 89], [SHARP]).

3.1.1 PRA Procedures Guide, THERP

Table 12 outlines the basic procedure recommended in the American PRA Procedures Guide ([PRA-PG], Chapter 4) for human reliability analysis. The procedure comprises 12 steps, which are very much tailored to the THERP method (Technique of Human Error Rate Prediction) described in Swain's handbook [Swain 80], [Swain 83]. Since other methods are also discussed here, this chapter concentrates on a simplifying classification of the essential steps (1 to 10) into two groups:

- error identification, and

- error quantification.

Note that in [Swain 83] (Fig. 5-6) a more detailed classification is given: familiarisation (steps 1-2), qualitative assessment (3-5), quantitative assessment (6-10), and incorporation (11-12).

Step	Level of analysis	
	error identification	error quantification
1) plant visit	X	
2) review information from fault tree analyst	X	
3) talk-through	X	
4) task analysis	X	X
5) develop HRA event trees		X
6) assign human error probabilities (HEPs)		X
7) estimate the relative effects of performance shaping factors (PSFs)		X
8) assess dependence		X
9) determine success and failure probabilities		X
10) determine the effects of recovery factors		X
11) perform a sensitivity analysis, if warranted		
12) supply information to the fault tree analyst		

Table 12: Rough classification of steps recommended for human reliability analysis in the PRA Guide [PRA-PG]. The steps are to apply to each "human-related event" identified by the fault tree analyst.

Section 4 (corresponding to steps 1 to 4 in Table 12) gives a brief survey of the methods of predictive identification of safety-relevant errors. Section 5 (steps 4 to 10) contains a rough evaluation of various probabilistic error quantification methods.

The classification selected here emphasizes the central significance of task analysis for the evaluation of human reliability. It comprises ([PRA-PG], Chapter 4) a decomposition into subtasks (subactions) and the identification of safety-relevant error opportunities.

Step 4 referred to as task analysis is closely connected with the other steps. A task analysis is based on the plant visit, review of human-related events identified by fault tree analyses and discussion of the procedure steps of the manual action to be analysed together with the

competent operators. The most important factors to be included are compiled in [DRS-B] (pp. 399-400):

- task description, i.e. a brief description of the tasks to be fulfilled by the personnel,
- location, i.e. specification of the location at which the action is performed,
- time after the occurrence of the accident, i.e. time at which the action must be initiated at the latest,
- duration of the action, i.e. average time required for performing the action,
- written instructions or aids, i.e. stating whether and, if any, what written instructions or working documents are available to the personnel for performing the action,
- possibility of identifying the need for manual actions, i.e. what means of information are available to the personnel indicating the need for performing the action (only basic information sources are included),
- substitution for automatic action, i.e. stating whether the action described must be performed in the absence of an action otherwise initiated automatically,
- cause, i.e. stating the reasons requiring the action,
- action feedback, i.e. stating the means of information which provide the personnel with a feedback of successful task performance,
- relation to other actions, i.e. information about functional relations with other actions.«

This provides important information for a realistic quantification in steps 5 to 10 (Table 12); e.g.

- functional relations between subactions for realistic modelling in step 5, or
- conditions aggravating or facilitating task performance for a realistic matching of HEPs and conditions in the plant to be analysed (steps 6 and 7).

The evaluation of individual operator performances is not the subject of a task analysis.

3.1.2 IAEA, SHARP

The THERP handbook ([Swain 83], Chapter 4) had set up 10 steps for man-machine system analysis (MMSA). These steps can be arranged as a subset of the HRA process described in [IAEA 89], which is similar to the approach in [SHARP]. If the MMSA steps in [Swain 83] are taken as sub-steps of the IAEA structure, the HRA process is divided into the following steps which have to be repeated to optimize MMS reliability. Note that steps 9 and 10 of Swain are excluded because they are relevant for the design process but not for the HRA process.

1. Definition: Analysis of the different types of human actions (types A-C in [IAEA 89], corresponding to types 1-5 in [SHARP]).

2. Screening: Identify the human interactions that are significant for the operation and safety of the plant.
   1. Describe the system goals and functions of interest.
      In this step the points of interaction have to be identified.
3. Qualitative Analysis: Detailed description of the important human interactions and definition of the key influences.
   2. Describe the situational characteristics.
      In this step the external PSFs have to be identified.
   3. Describe the characteristics required of the personnel.
      In this step human-related PSFs (internal PSFs) have to be identified.
4. Representation: Modelling of human interactions in logic structures.
   4. Describe the jobs and tasks performed by the personnel.
      In this step a task analysis has to be performed. Note that a task analysis for HRA purposes is generally not performed in as much detail as for ergonomic design.
5. Impact Integration: Exploration of the impact of significant human actions.
   5. Analyse the jobs and tasks to identify error-likely situations and other problems.
      In this step the task has to be brought into relation to the situational and personnel characteristics.
6. Quantification: Assign probabilities for the interactions.
   6. Estimate the likelihood of each potential error.
      In this step the dependence of subtasks has to be estimated.
   7. Estimate the likelihood that each error will be undetected or uncorrected.
      In this step, recoveries and consequences have to be estimated.
   8. Estimate the consequences of each undetected or uncorrected error.
      In this step a sensitivity analysis has to be performed.
7. Documentation: Making the analysis traceable, understandable, and reproducible.

Summarising, one can find two different requirements concerning the type of data and the level of data. The first requirement is related to the qualitative type of HR data while the second one is related to the quantitative type of HR data. They are summarised in Table 12.

3.2 Data acquisition for HRA

3.2.1 Requirements of the HRA methods

In [IAEA 90], the requirements of the different methods have been summarised. In the document one can find the following requirements:

The HCR model [Hannaman 84] needs data about

- the working media or the man-machine interface
- influence factors e.g. stress, skill level, fatigue, environment and organisational effects
- the time available for diagnosis and correct execution of a task (time window for action).

The SLIM approach [Embrey 84] needs data about

- the description of the task and actions
- influence factors e.g. stress, skill, fatigue, environment and organisational effects
- the number of demands for two tasks for calibration of the quantitative estimations.

The THERP method [Swain 83] needs data about

- the description of the task and the actions
- influence factors e.g. stress, skill level, fatigue, environment and organisational effects
- recovery factors for the different tasks
- the working media or the man-machine interface
- the persons or teams which have to perform the task
- the time available for diagnosis and correct execution of a task (time window for action)
- the available procedures
- the dependence of the different tasks and influence factors.

The ASEP method [Swain 87] needs data about

- the technical system
- the available procedures
- the time available for diagnosis and correct execution of a task (time window for action)
- recovery factors for the different tasks
- the dependence of the different tasks and influence factors
- the error type.

3.2.2 A framework to acquire HRA data

The approach presented in the following is a bottom-up approach that is able to provide data for different current HRA methods (for discussion see [Sträter 94a]). It has the advantage that the data obtained can be used for most of the present HRA purposes. The framework is outlined in Figure 3.

The framework is based on the MMS (man-machine system) approach (see [Bubb 93]). A MMS can be described in the following way: A task description will be evaluated by an operator. The operator processes the task with respect to the actual system state (feedback loop). By performing an action via the control elements he changes the system behaviour. The system change is then displayed and reported to the operator. The whole MMS is embedded in an organisation structure and an environment. The organisation is related to shift or revision plans or general strategies to improve safety for instance. Here everything can be noted, which is not directly related to the work of the person. In some situations the operator also has to report the task transaction to the supervisor or has to document his transaction in a test protocol or written procedure.

To extend this MMS approach to an error modelling approach it is necessary to add influences to the different stages of the MMS as indicated by the dotted arrays in Figure 3. These influences are known as performance shaping factors (PSFs). Internal PSFs can be found in the human-related arrays and external PSFs in the other arrays.

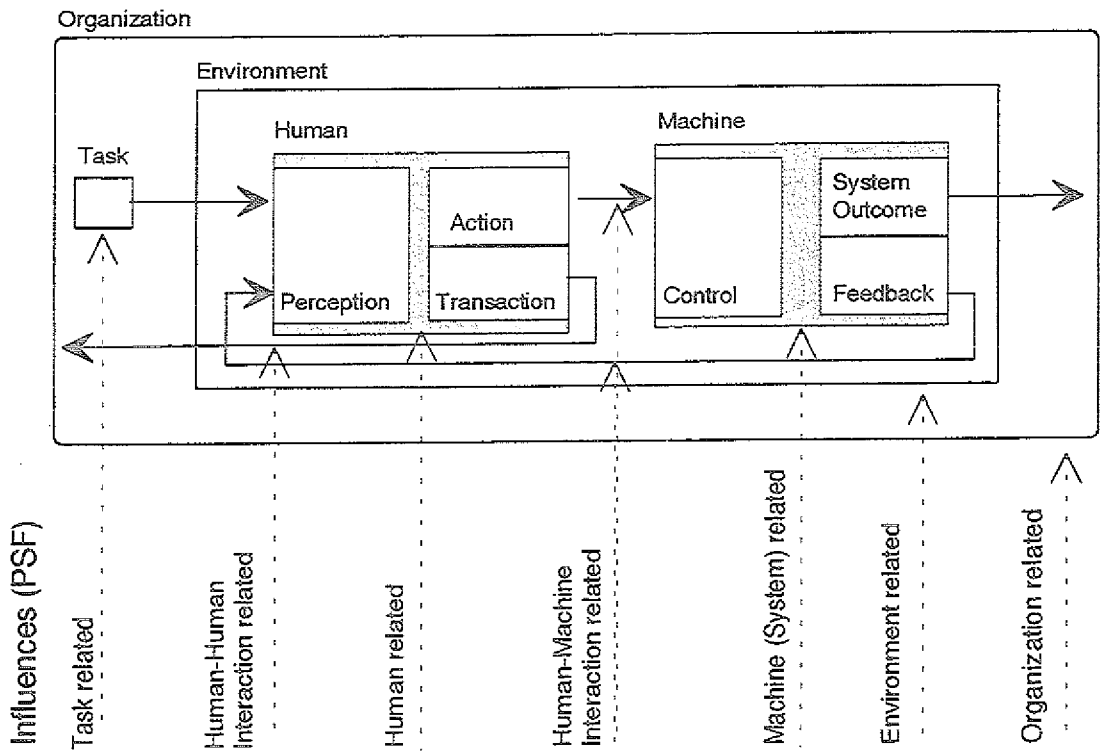


Figure 3: A framework to collect human reliability data for various HRA methods.

From the framework, nine question marks can be derived which are outlined in Table 13. Answering these questions leads to complete information about an event, if the following properties are noted according to the different stages of the MMS: factual information, error taxonomies (like EOM/ECOM in Table 8), and influences. If a taxonomy is used to fill up the framework the acquired data become highly reliable. For more details on this framework see [Sträter 95a].

1	Task	What was the task of the person ?
2	Person	Who was the person involved in the event ?
3	Action	What has the person done ?
4	Feedback	What were the information sources to check what has to be done ?
5	System	What part of the system has been manipulated ?
6	Environment	Where has the event happened ?
7	Transaction	What had the person to do to inform others about his work ?
8	Organisation	What are the contributing organisational factors ?
9	Time and Duration	When did the event happen and how long did it last ?

Table 13: Nine question marks on data for HRA.

With this scheme it is possible to acquire data for HRA purposes from events or simulator data or other types of low level information (e.g., experiments). The framework introduced in this chapter is able to provide most of the information which is necessary for the HRA methods. How to use this information is described in the methods.

4 Identification of safety-relevant errors

4.1 Analytical expertise

The methods of hazard and error identification summarised in [Madjar 93], [ESCIS 86] and [Whalley 89] can be used, in principle, for the identification of safety-relevant error opportunities contributing to a system failure of interest or to the failure of an action (or sequence of actions). Error identification in human reliability analysis within a PRA is generally inductive and deductive without any rigid scheme being given (»no hard-and-fast rules« [Swain 83], page 4-9). The prediction of safety-relevant error opportunities is mainly based on the analyst's analytical expertise. It is thus a highly knowledge-based activity.

Important preliminary information on safety-relevant operator errors is provided by the deductive approach of fault tree analysis. The systematic search for failure causes of a system function enables the detection of human errors related to these failure causes; see Figure 4 and Table 2.

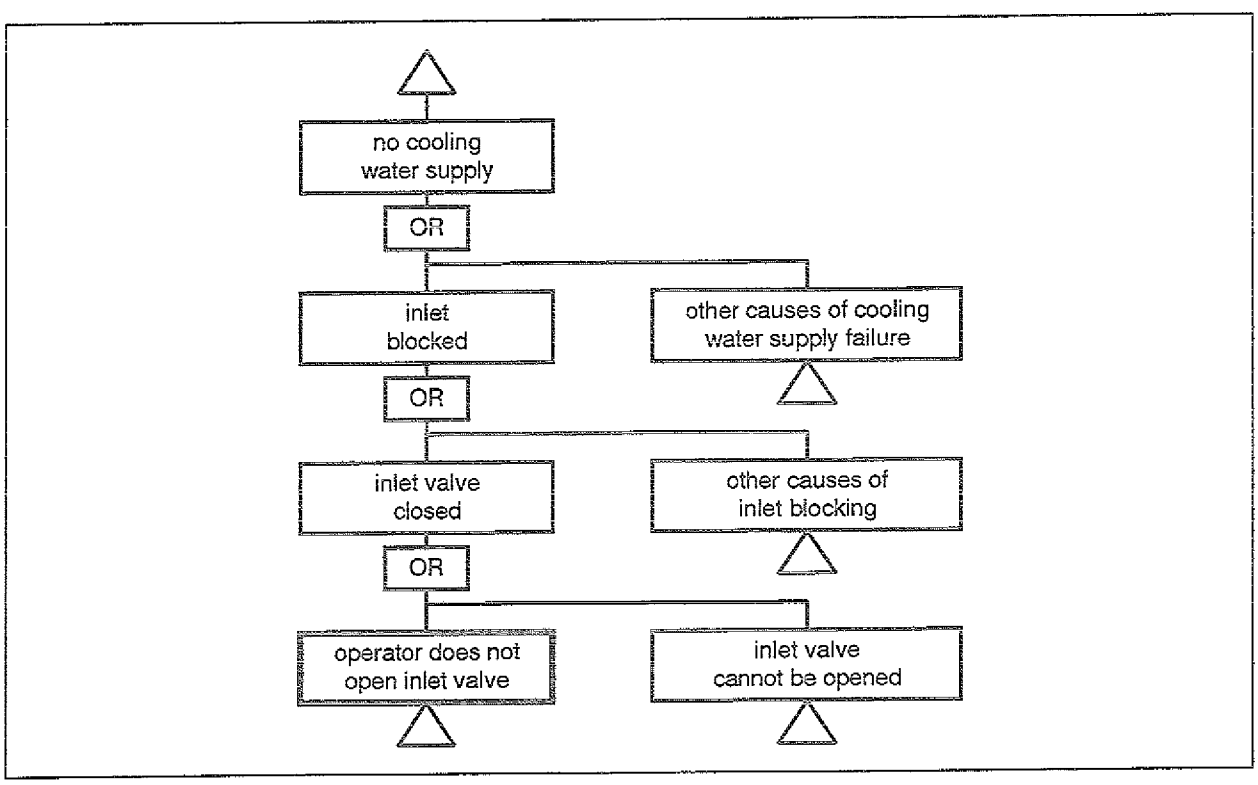


Figure 4: Deductive procedure for the identification of operator errors contributing to system failures

As a rule, the review of available fault trees (step 2 in Table 12) is not sufficient to ensure a high degree of completeness in error identification. For this reason, the single subtasks, into

which a task has been decomposed, are considered in relation to man-machine interactions. [Swain 83] "decomposes" man into the components of perception, mediation and response; see Figure 5. On the one hand, man is *exposed* to external impacts but, on the other hand, he can also *act on* the external environment. For the identification of errors that may occur in these interactions, Swain and Guttmann listed a number of factors characterizing these five components (external input, perception, mediation, response, human output).

Incorrect human outputs, i.e. the externally observable aspects of human errors, are divided in Swain's handbook into the groups: errors of omission and errors of commission. The subcategories (Table 8) of these main groups are suitable as tools for inductive error identification, for example:

- Category "too much": What happens if an excessively high shutdown gradient is adjusted?
- Category "too little": What happens if less than four steam generators are depressurized?
- Category "too early": What happens if the steam generators are depressurized too early?

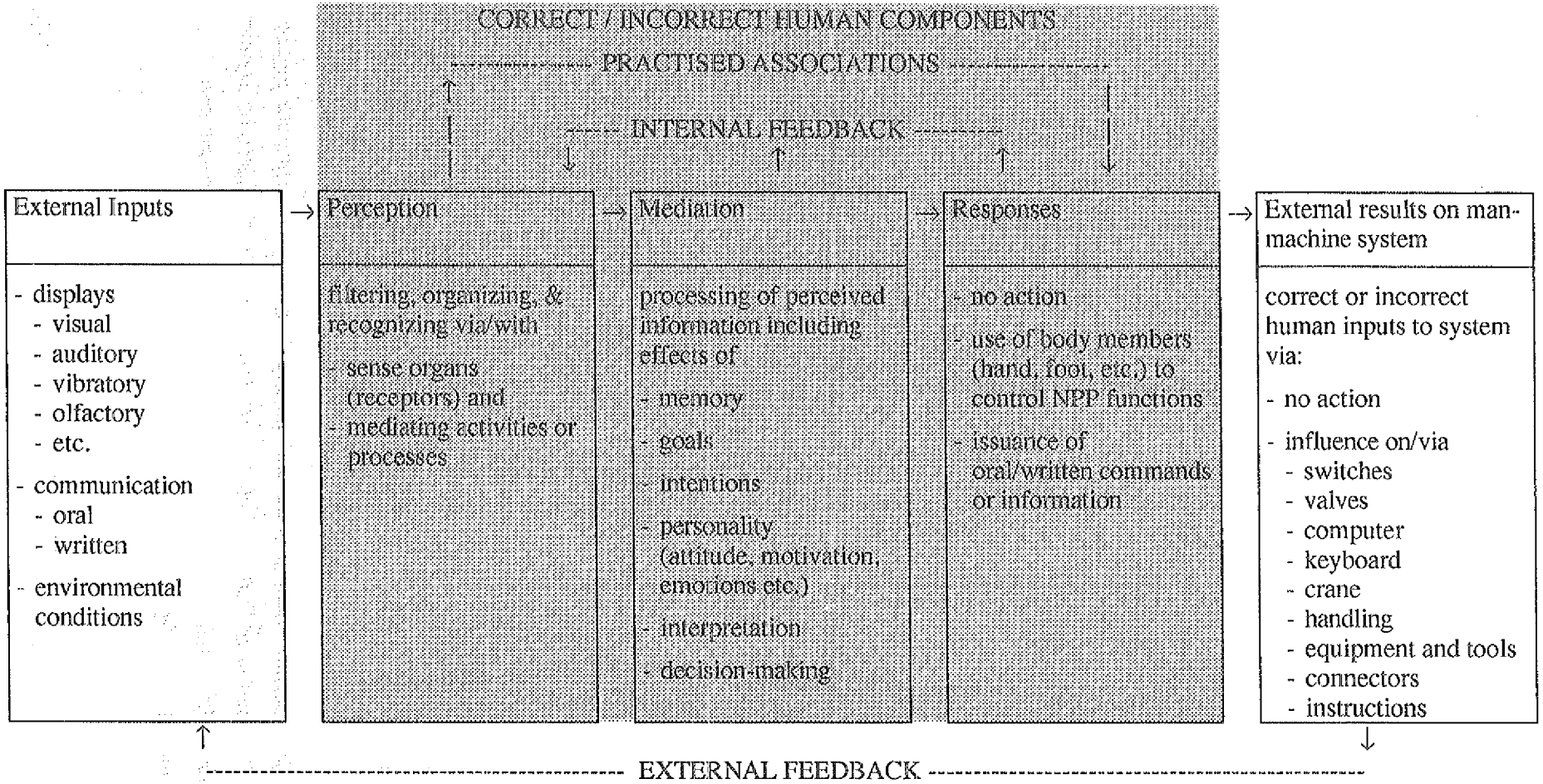
Deductive (in fault tree generation, see Figure 4) and inductive (checking possible deviations from the correct performance of a task with the error categories listed in Table 8) analyses thus ensure a largely complete identification of errors acting on the system as obvious (basically observable) modes of human behaviour. In addition, the man-machine model outlined in Figure 5 and the categories listed in Tables 9 and 11 also help to identify factors contributing to the causation of such errors. Knowledge of these factors plays a role in error quantification, especially in the assignment of human error probabilities (HEPs) and in the identification of performance shaping factors (PSFs) which have an increasing or decreasing effect on these HEPs, see steps 6 and 7 in Table 12.

In all cases background knowledge about

- basic findings in error causation (e.g.: [Swain 83], [Wickens 84], [Reason 90]) and
- observable circumstances (constellations of external PSFs, context) of errors in real incidents (cf.: [Flanagan 54], [Dougherty 93])

would be very helpful for predictive error identification (see Figure 1).

Recent model concepts consider man-machine interactions in a larger context, see Figure 6, including also organisational in addition to technical features. It should therefore be examined in a separate research project whether and how these extended models lead to *new* insights into error identification and are quantifiable.



From [Swain 83] (page 4-10), slightly simplified

Figure 5: Simplified model for the incorporation of human components into a man-machine system

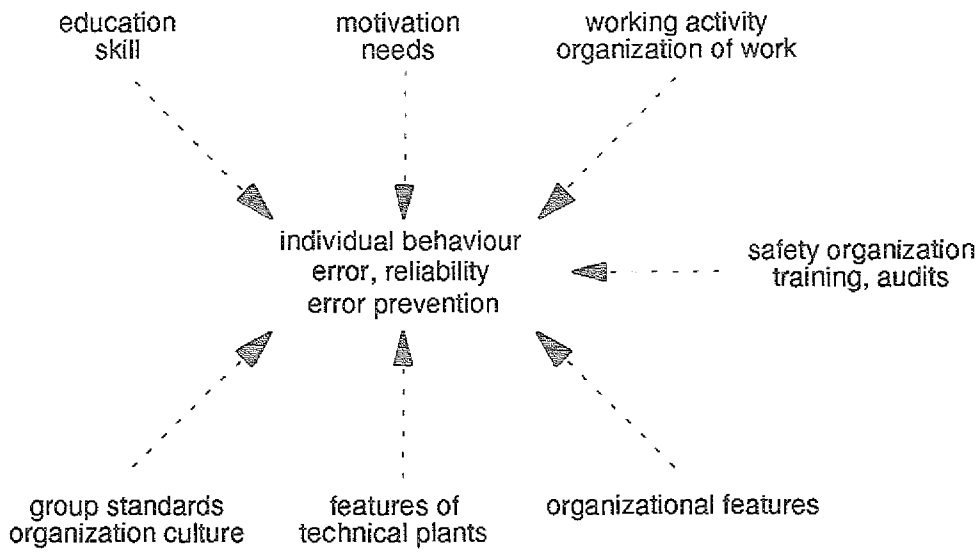


Figure 6: Survey of factors influencing individual behaviour (translated from [Grote 93], page 6)

4.2 Computer programs as tools

As already mentioned above, the degree of completeness in error identification greatly depends on the analyst's expertise, among other aspects on his experience from previous analyses performed, on his knowledge of plant-engineering details, on his knowledge of cognitive error mechanisms, on his knowledge of previous incorrect actions in technical systems and on his ability to draw analogy conclusions for the system analysed.

Attempts have recently been made to reduce this dependence by developing special computer programs. The computer is to serve as an aid for the analysis of complex man-machine systems. The basic idea of the ESAP computer program (expert system for the systematic analysis of operator actions [Degen 94]) is to provide support in the application of THERP in order to reduce subjective influences, and also to create a knowledge base from previous analyses; this knowledge base can then be used for current analyses. The development of ESAP is not yet completed.

The computer system CES (»Cognitive Environment Simulation« [Woods 88], [Woods 90]) is also still at the development stage. Work has progressed to a level where the first test applications are available. CES simulates the process of decision-making; see Figure 7. Information about dynamic plant behaviour under accident conditions is processed in a manner providing the CES user with a list of intentions to act which could be pursued by the

personnel. In simulating operator decisions, insights from psychology and from a number of accidents already occurred are utilized. According to the CES model concept, a decision-based error is caused by a mismatch between problem solution requirements (accident variant, available data, ...) and problem solution capability (knowledge of plant behaviour, preferred strategy in decision-making, ...). Consequently, CES is also in a position to simulate complication factors, i.e. deviations from the text book case of an accident situation. This will be illustrated by an example ([Woods 90], Chapter 4) describing the simulation of decision-making after an accident with steam generator tube rupture:

> The CES output showed for the text book case that the personnel diagnoses the accident without any significant delay. In a second case, complications concerning the accident variant were entered *by the CES user*, for example a leak in the seal of a main coolant pump as an additional failure. The CES output showed an increased diagnosis time. Knowledge deficiency was then entered as a further complication factor, i.e. the personnel attributes the falling pressurizer level caused by the steam generator heating tube leak to the sealing leak. In this case, the CES output showed that accident diagnosis fails.

In this example, the CES user thus decided which complication factors were to be simulated in accident diagnosis. The analyst's expertise is therefore also of decisive significance for the application of such computer programs. According to the present state of the art, these tools can at best support the analyst's knowledge-based skills, but by no means replace them.

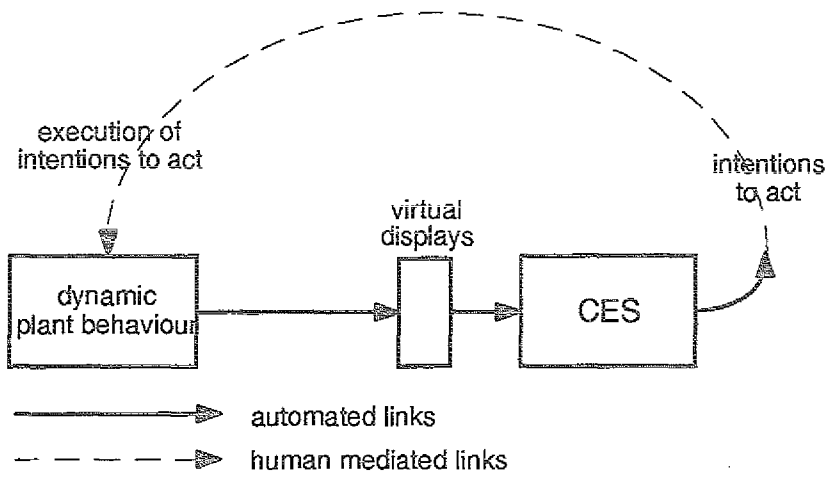


Figure 7: Simulation of human intention formation using CES [Woods 88]

5 Methods for quantitative human reliability analysis (HRA)

5.1 Overview of the variety of HRA methods

This section gives an overview of probabilistic methods for quantitative human reliability analysis. Such methods provide guidelines for the analyst assessing the failure probability of a manual action or the probability of an operator error identified as safety-relevant.

In the meantime, there exists a large variety of quantification methods. PRA-oriented evaluations of individual quantification methods can be found, for example, in [Swain 89] or [Kosmowski 94]. Evaluations from a psychological point of view are to be found in [Reason 90] and [Zimolong 90]. Table 14 gives a rough survey of existing quantification methods and some important evaluations concerning these methods.

5.2 Selection of significant HRA methods for further evaluation

In addition, two recent methods are also worth mentioning:

- the method based on French operating and simulator experience [Mosneron 90];
- the method derived from the EPRI-sponsored ORE programme (HCR/ORE) [Moieni 94].

A first rough review led to seven methods (Table 15) that appear worth evaluating in more detail. This preselection is governed by the following three criteria:

1. topicality: new methods for which only few or no evaluations are available (HEART, INTENT, EdF'sPHRA, HCR/ORE),
2. PRA relevance: methods frequently used in risk studies (THERP, HCR, SLIM, HEART, EdF'sPHRA),
3. cognitive relevance: promising methods for the quantification of nontrivial operator decisions (INTENT, EdF'sPHRA, HCR/ORE).

An evaluating description of the standard method THERP (Technique of Human Error Rate Prediction) is given in Chapter 6. The other six methods (EdF'sPHRA, HCR, HCR/ORE, SLIM, HEART, INTENT) are briefly described and roughly evaluated in Chapters 7 and 8.

Method		Evaluation			
		[Swain 89]	[Kosmowski 94] (Chapter 2)	[Reason 90] (Chapter 8)	[Zimolong 90]
THERP	Technique of human error rate prediction [Swain 83]	plausible results for competent application; speculative for knowledge-based behaviour during accidents	database available; insufficient consideration of decisions after diagnosis	efficient for competent application; overemphasis of externally observable errors	neglects psychological processes in performing tasks
ASEP	Accident sequence evaluation program [Swain 83]	short version of THERP; application inexpensive	enables fast preselection of important operator actions	n. e.	n. e.
OAT	Operator action tree [Hall 82]	pioneering achievement for diagnosis and time dependence; PSA steps unclear	operator decision analysable	time reliability curves not sufficiently differentiated from a psychological aspect	n. e.
AIPA	Accident initiation and progressing analysis [Fleming 75]	only of historical relevance	n. e.	n. e.	n. e.
HCR	Human cognitive reliability model [Hannaman 84]	interesting unproven approach; further development necessary	simple handling; incomplete concerning PSFs and time-independent HEPs	time reliability curves not sufficiently differentiated from a psychological aspect	n. e.
SAINT	System analysis of integrated networks of tasks [Kozinski 84], [Wortman 78]	complex interactions can be modelled; models partially difficult to understand; as yet unsuitable for PSA	n. e.	n. e.	n. e.
PC	Paired comparison [Seaver 83]	best expert estimation procedure for single HEPs	difficult for HEPs of complex tasks	n. e.	n. e.
DNE	Direct numerical estimation [Seaver 83]	requires good reference values; unsuitable for complete HRA	encourages useful discussions among experts	n. e.	n. e.

n. e.: not evaluated

Table 14: Methods for human reliability quantification and excerpts from their evaluations

Method		Evaluation			
		[Swain 89]	[Kosmowski 94] (Chapter 2)	[Reason 90] (Chapter 8)	[Zimolong 90]
SLIM	Success likelihood index methodolgy [Embrey 84]	flexible; poorly validated; sophisticated	good theoretical background; sophisticated expert estimations	less overemphasis of externally observable errors than in THERP; selection of reference HEPs is critical	PSF interactions are left out of consideration
STAHR	Socio-technical approach to assessing human reliability [Phillips 85]	contributes towards understanding human actions; as yet unsuitable for PSA	not particularly user-friendly at present	n. e.	n. e.
CM	Confusion matrix [Potash 81]	pioneering achievement for problems of error in diagnosis	greatly dependent on expert estimations	primarily restricted to qualitative applications	n. e.
MAPPS	Maintenance personnel performance simulation model [Siegel 84]	suitable rather for optimization than for PSA assessment of maintenance work	sophisticated; results difficult to understand	n. e.	n. e.
MSFM	Multiple-sequential failure model [Samanta 85]	dependence model, not yet mature	n. e.	n. e.	n. e.
SHARP	Systematic human action reliability procedure [SHARP]	useful method for planning HRA in a PSA	useful frame for the integration of several methods	facilitates method selection	n. e.
HEART	Human error assessment and reduction technique [Williams 88]	n. e.	simple; excessively isolated consideration of single tasks	n. e.	n. e.
INTENT	Method for estimating HEPs for decision-based errors [Gertman 92]	n. e.	promising approach to the quantification of operator decisions	n. e.	n. e.
TESEO	Tecnica empirica stima errori operatori model [Gertman 92]	n. e.	simple; theoretical background questionable	simple; no hard data base	n. e.

n. e.: not evaluated

Table 14 (continued): Methods for human reliability quantification and excerpts from their evaluations.

1.	THERP ("Technique of Human Error Rate Prediction"), described in [Swain 83], short version (ASEP) in [Swain 87]
2.	EdF'sPHRA (Method of Electricité de France for Probabilistic Human Reliability Analysis), described in [Mosneron 89], summarised in [EPS 1300] and [Mosneron 90]
3.	HCR ("Human Cognitive Reliability Model"), described in [Hannaman 84], summarised in [Hannaman 85] and [Hannaman 88]
4.	HCR/ORE ("Human Cognitive Reliability/Operator Reliability Experiments"), based on operator reliability experiments, described in [Spurgin 90] and summarised in [Moieni 94] and [Parry 91]
5.	SLIM ("Success Likelihood Index Methodology"), described in [Embrey 83] and [Embrey 84]
6.	HEART ("Human Error Assessment and Reduction Technique"), described in [Williams 88]
7.	INTENT ("Method for Estimating Human Error Probabilities for Decision-Based Errors"), described in [Gertman 92]

Table 15: Significant methods for human reliability quantification

5.3 Classification of significant HRA methods

HRA methods may be classified into different classes according to the following criteria:

- the level of detail into holistic and decompositional methods,
- the level of data scale, which is used in the method, into absolute scale, relative scale and ordinal scale,
- the key parameters used by the method into error related, time reliability related and PSF related methods.

5.3.1 Classification into decompositional and holistic methods

The classification into decompositional and holistic methods is made according to the procedure, which is used in the assessment process (see [Zimolong 91], [Heslinga 93], [Gerdes 93]). Decompositional methods quantify the reliability of human actions by decomposing a situation into sub-situations up to a defined degree of resolution of the action tree. They assess each single sub-situation, and the reliability for the entire task is inferred by combining the reliability for the sub-tasks. Holistic methods perform a quantitative assessment by assessing the entire situation without distinguishing between the different tasks in a given situation. Table 16 allocates the methods to this classification.

Decompositional Methods	THERP ASEP
Holistic Methods	EdF'sPHRA HCR HCR/ORE SLIM HEART INTENT

Table 16: Classification of the considered methods into decompositional and holistic methods.

5.3.2 Classification according to the data levels

Another possibility for the classification of HRA methods is the underlying level of the data scale. The level of the data scale may be subdivided into absolute scale, relative scale and ordinal scale.

Absolute scale means that the method does provide real HEP in the range from 0 to 1 where 0 means no error and 1 sure failure. Relative scales are only able to provide information like "if HEP_1=0.1 and HEP_2=0.2 then event 2 is twice as probable as event 1". Ordinal scales only can provide information like event 2 is more likely than event 1. It cannot be stated to what extent it is more likely.

A simple rule to distinguish the methods is that absolute scales need real probabilities given by the general probability axiom n/N where n is the observed frequency of events (of interest) and N all observable events. Relative scales need only a subset of the observed events N. The more events observed, the closer the scale is related to the absolute scale. Ordinal scales only need

frequencies and no number of observed events. Judgement methods also only provide ordinal scaled data. Normally, methods using this type of scale try to enhance the data level by calibrating the frequencies with some mathematical approach.

Surely, every HRA method intends to achieve the absolute data scale, because it is needed for PSA application. Unfortunately, most of the methods do not provide such data but behave as if they had this data level. A typical example is the SLIM method which uses so-called anchor tasks to calibrate other tasks to generate HEPs from data that is of an ordinal scale nature.

Absolute scales are sparse and can be excluded for HRA methods. No HRA method can achieve this data level. Even the THERP catalogue for instance only provides data on the relative scale, because absolute scales need the number of all observable events. Concerning the ordinal scale, it is not usually possible to enhance the data level of methods, because either the underlying assumptions of the mathematical approach or the mathematical approach itself is not proven. Table 17 allocates the methods according to the data level. For methods like SHARP for instance, which do not provide data at all, another type of data level is added.

Relative Scale Methods	THERP ASEP EdF'sPHRA HCR/ORE
Ordinal Scale Methods	HCR SLIM HEART INTENT
No Scale at all	SHARP

Table 17: Classification of the considered methods according to their data scale.

5.3.3 Classification according to key parameters

A very practical classification is the distinction of key parameters mainly used by the methods. They may divide the methods into error related, time reliability related and PSF related methods. Table 18 gives an overview of the classification.

Error related	THERP ASEP
Time reliability related	EdF'sPHRA HCR HCR/ORE
PSF related	SLIM HEART INTENT

Table 18: Classification of the considered methods according to their key parameters.

The last classification approach will be used in the detailed description of the following three sections.

6 The standard method: THERP

After publication of the preliminary version [Swain 80] the final version of Swain's handbook [Swain 83] on human reliability analysis (HRA) appeared in 1983. These handbooks document a level of development which began in the early sixties and was referred to as "Technique of Human Error Rate Prediction" (THERP).

Human reliability analysis using the THERP version described in Swain's handbook is comparatively sophisticated. The American PRA Guide [PRA-PG] specifies approximately two to three man-months per NPP-PRA for the effort required (in the authors' experience, this figure is too low rather than too high). For this reason, Swain published a short version of THERP [Swain 87] under the title "Accident Sequence Evaluation Program" (ASEP), which permits a time-saving rough assessment ("screening") which, in most cases, leads to pessimistic results. Moreover, ASEP also contains explanatory and complementary notes on methodological details in Swain's handbook (1993). [Swain 89] (page 3-41) therefore recommends the use of both sources ([Swain 83], [Swain 87]) for a comprehensive HRA.

The descriptions and evaluations carried out here are based on the 1983 THERP handbook [Swain 83] since THERP and ASEP are based on the same methodological principles.

THERP was used in the following studies (cf. [Swain 89], page 3-37):

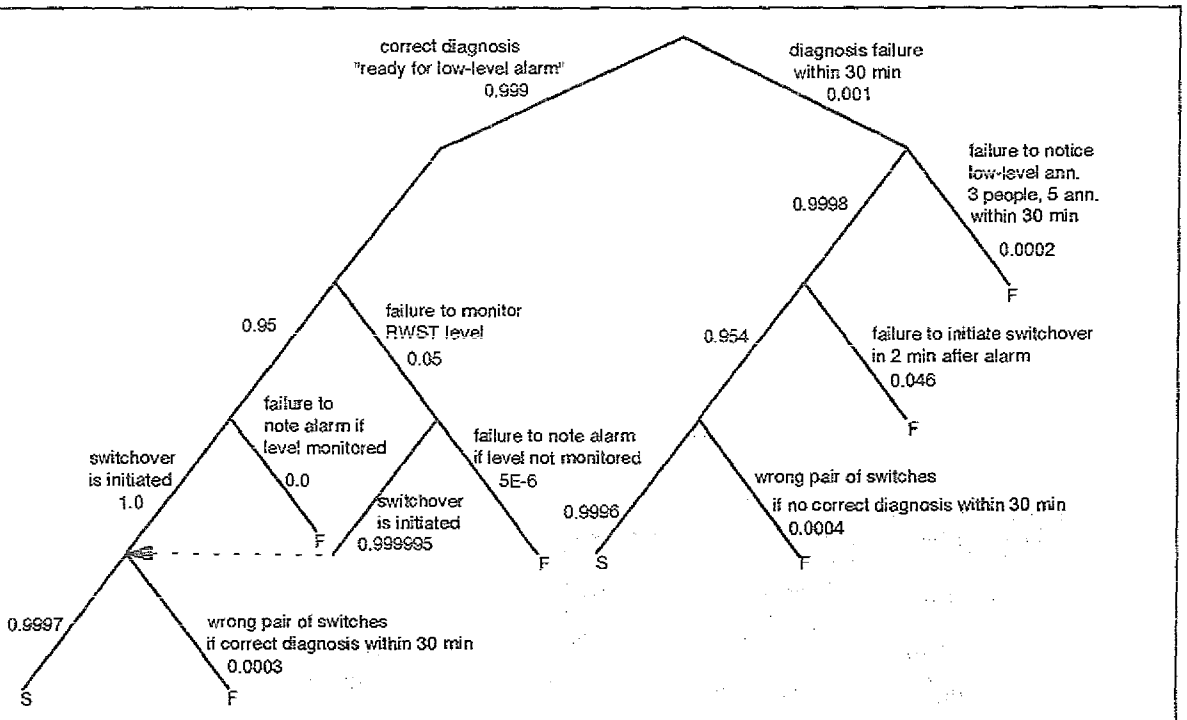
- Zion PRA [Zion 81];
- Indian Point PRA [Indian Point 82];
- Arkansas Nuclear One PRA [Kolb 82];
- Peach Bottom PRA [Kolaczkowski 86];
- Grand Gulf PRA [Drouin 87]
- and in various analyses of non-nuclear systems [Miller 87].

Furthermore, applications are reported in Scandinavian [Hirschberg 90] and German [DRS-B] studies.

6.1 Error model

DESCRIPTION

Boolean algebra is the main error model of the THERP method. In order to calculate the (initially unknown) failure probability of an operator task, the principle of action or task decomposition is used. It is assumed that this failure probability using the operation rules of Boolean algebra can be calculated from the single failure probabilities of the subtasks into which the task was previously decomposed. The 'tool' for task decomposition and subsequent error quantification is an event tree ("HRA event tree"), which models the logical and time-dependent correlations between the individual success and failure events. Figure 8 shows such an event tree. It is an error model for a task (switching to the circulation mode) to be performed by the personnel in the course of a major loss-of-coolant accident in order to prevent failure of the pumps required for decay heat removal.



From ([Swain 83], p. 21-22) with additional explanations reflecting the analysis described by Swain & Guttmann; F: failure, S: success

Figure 8: THERP event tree modelling for the failure of the task of changing from feed to circulation mode as soon as the alarm warns of an excessively low water level in the refuelling water storage tank (RWST) during a large loss-of-coolant accident

The analysis example illustrates two principles to be applied in decomposing the task in the sense of THERP:

- the consistent consideration of dependences between subtasks; cf. [Swain 83], Chapter 7;
- decomposition into meaningful subtasks, so-called perceptual units; cf. [Swain 83] (pages 11-5, 13-5, 15-14, etc.)

For the assessment of single error probabilities (HEPs) required for the quantification of an event tree for HRA, THERP uses different "submodels". The most important submodels are:

- time-reliability correlation to assess the probability for a diagnosis failure as a function of the time t available for diagnosis:
  $HEP(t) = pr(T>t)$,
  where the time T required for diagnosis is modelled as a random variable;
- multiplicative model for adapting (modifying) a nominal error probability (NHEP) to conditions (e.g. stress) to be encountered in the plant or action situation and having a reliability-increasing or reliability-decreasing effect. The adaptation is carried out with the aid of influence factors, called "performance shaping factors" (PSFs) in THERP:
  $HEP = NHEP \cdot PSF_1 \cdot PSF_2 ...$,
  where rarely more than two PSFs are assumed. Numerous PSFs (e.g. arrangement of switches) are already considered in selecting the NHEP.

EVALUATION

On the whole, the THERP error model, which essentially consists of action decomposition, time-reliability correlation and multiplicative PSF model, is to be evaluated positively.

The decomposition of action provides an important contribution towards understanding the man-machine situation in a complicated technical system. The authors currently know of no better method for an outsider to obtain analytical access to human reliability in technical systems. In decomposing an action into meaningful subactions perceived by the operator as such, and in thoroughly considering any dependences involved, important information is obtained about logical and time-dependent correlations of significance for the successful performance of an action.

Moreover, thorough decomposition of an action provides subactions which are suitable for experimental verification with respect to their reliability. As a general rule, the determination of experimental reliability is the simpler,

- the higher is the error probability (HEP), i.e. relatively few trials are required in order to obtain a reliable estimate, and
- the smaller is the number of PSFs to be checked for each trial.

The time-reliability correlation as a tool for quantifying the diagnosis is a pragmatic, but meaningful supplement to action decomposition. According to the THERP model concept ([Swain 83], Chapter 12), diagnosis is a highly dynamic (i.e. knowledge-based) activity which is difficult (and with considerable effort) to access by decomposing the action. Figure 5 illustrates the dynamic component of a diagnosis activity after the occurrence an accident. It is to be expected that both the external and the internal control loop (feedback) are 'passed' several times. Simulator experience confirms this model concept:

> »Of particular significance from the operators' point of view, and thus their main problem, is the fact that, after the beginning of a major accident, ... a completely unstructured decision problem is initially encountered in the control room. The operator does not know for the moment what happens....« (translated from [Hoffmann 89], page 5.2-6).

For the practical feasibility of an HRA it is therefore meaningful to model the diagnosis as *one* action, although it comprises several subactions. Whether this is meaningful from a psychological point of view will have to be investigated in more detail. The quantification problems arising in connection with time-reliability correlations will be discussed later.

The multiplicative PSF modification of NHEPs is also to be classified as meaningful from a pragmatic rather than psychological point of view. It would be ideal to have an empirically well-founded failure probability under comparable boundary conditions (PSFs) for each subaction to be quantified in the event tree. A realization of this ideal case would require a considerable amount of data. An additional data requirement would have to be expected for every new HRA. Modification models are therefore indispensable for the practical feasibility of an HRA. The use of multiplicative models is meaningful in so far as a probability is in the first place a multiplicative variable. The probabilistic implementation of all operations of Boolean algebra can be attributed to multiplications by failure probabilities (P) and success probabilities (Q) in the sense of non-failure probabilities (Q = 1 - P).

In strict mathematical terms, the direct application (HEP = NHEP · PSF) of multiplicative models for probabilities is only approximately correct, for a probability is not a variable whose multiplicity is defined on a ratio scale; the mathematical principles of the scale theory can be found, for example, in [Coombs 75]. The reason for this is obvious. An HEP cannot assume values greater than 1. For example, NHEP = 0.5 and PSF = 5 would give HEP = 2.5 - a nonsensical result which would have to be interpreted as HEP = 1. For NHEP = 0.5 and PSF = 10 the result would also have to be interpreted as HEP = 1, although this case (PSF = 10) is to be classified as twice as 'susceptible to failure' as the case with PSF = 5 - a violation of the prerequisite for the application of a ratio scale.

It is mathematically correct to perform the multiplicative PSF modification for the quotient, $y = b/a = (1\text{-NHEP})/NHEP$, of the number b of successes and the number a of failures. This quotient is also referred to as the odds ratio. The multiplicity of this quotient is defined on a ratio scale ($0...\infty$) in the same way as that of its reciprocal value (a/b). The mathematically

correct modification of a nominal error probability should therefore take place according to the equation

$$HEP = \frac{1}{1+\frac{y}{PSF}} = \frac{NHEP \cdot PSF}{NHEP(PSF-1)+1}$$

and this equation would also have to be used for the empirical PSF determination:

$$PSF = \frac{y_1}{y_2}$$

where y_1 is the odds ratio in situation 1 and y_2 the odds ratio in situation 2. The PSF expresses the extent to which these two situations differ with respect to their susceptibility to failure.

It is shown in [Reer 93] (p. 128) that the results of the mathematically correct modification procedure, HEP = 1/(1 + y/PSF), differ from the results of the usual PRA modification procedure (HEP = NHEP · PSF) in proportion to the closeness of NHEP to 1. On the whole, the usual PSA procedure is acceptable, since it tends to overestimate rather than underestimate the HEP.

In principle, the THERP error model is to be classified as positive. However, the authors do not agree with the quantitative implementation carried out in the analysis example (Figure 8), since the recovery effect of the level alarm of Swain & Guttmann is assessed as too optimistic. According to psychological findings, as summarised in [Semmer 94], a dependence between diagnosis failure and failure of the respective control mechanisms has to be assumed. This point of criticism is discussed in more detail in Section 6.7.

6.2 Database

Description

The data contained in Swain's handbook (1983) mainly consist of:

- descriptions for approx. 100 NPP-specific errors with associated boundary conditions and
- numbers for point values and uncertainty bounds of the corresponding error probabilities.

Table 19 gives a rough survey. In addition, Swain's handbook specifies a few factors (PSFs) for the situation-specific modification of error probabilities, e.g. due to stress or lack of operator experience.

Evaluation

The validity of the THERP data can be classified as reasonably satisfactory.

It is correct that only few HEPs are directly based on empirical data from nuclear power plants. On the other hand, however, many HEPs do not have any empirical database at all, but were determined from expert estimates by the authors of the handbook. Table 20 gives a rough survey concerning the sources of THERP data.

It should be noted, however, that the expert estimates are based on sound knowledge concerning man-machine interactions in NPPs. Swain's handbook comprising roughly 700 pages is itself the best support for this assessment. It is still one of the most comprehensive collections of knowledge on human reliability with the emphasis being placed on safety-relevant situations in NPPs.

Moreover, some empirical studies confirm the validity of essential HEPs from THERP within the usual PRA uncertainties. The diagnosis failure probabilities determined by EdF simulator experiments [Villemeur 86] are within the uncertainty bandwidths of the THERP diagnosis model; most of the plots (EdF) are very close to the nominal THERP curve. For critical actions after diagnosis, HEPs were determined in simulator studies [Beare 84] which deviate from the corresponding THERP HEPs within a factor of 2.

However, the validity of the THERP HEPs evaluated here as satisfactory only holds in connection with the associated uncertainty factors (UF = 2...30, depending on HEP). A satisfactory validity (with respect to input data quality) of the final result can only be expected if the HEP uncertainties are also considered when using the HEPs.

The uncertainty factors recommended in Swain's handbook in connection with error probabilities are based on conservative conclusions from experiments [Wechsler 52] on the variability of human performances.

The variables used by Wechsler to measure human performance are not failure or success probabilities, but concrete quantitative results from working activities, such as the number of words typed per minute. The authors cannot evaluate at present to what extent the derived uncertainty factors (in human variability) can be transferred to error probabilities. Investigations are recommended here based on the methodological principles of mathematical statistics concerning variance.

In any case, Swain & Guttmann hold the view that the THERP uncertainty factors cover the variability causes listed in Table 21 of this report ([Swain 83], p. 7-9). The above-mentioned simulator studies ([Villemeur 86], [Beare 84]) suggest that this applies at least to the uncertainty types listed under point 1 (lack of NPP-specific data) and point 5 (individual and time-dependent variability). The authors are not able to evaluate at present the extent to which the other uncertainty types are covered. Further relevant studies are recommended here. Results from so-called benchmark exercises [Poucet 88] suggest that Swain & Guttmann underestimate the variability causes listed under points 3 (incompleteness) and 4 (analyst).

The authors therefore assess the THERP uncertainty factors as sufficiently large on condition that THERP is competently applied, and they are aware of the problem of determining whether an application is competent or not. The problem of a non-competent application is also known to [Swain 89] (p. 3-41).

Operator error	NHEP(*/UF) [Swain 83]
1. incorrect accident diagnosis	0.00001(*/30) ... 1.0
2. misreading or failure to note information in observing system state displays	0.001(*/3) ... 0.1(*/5)
3. failure to respond to an audible alarm	0.0001(*/10) ... 0.25(*/10)
4. performance of a procedure without using the relevant written document	0.005(*/10) ... 0.3 (*/5)
5. incorrect use of a checklist	0.5 (*/5)
6. omission of a step in performing a procedure	0.001(*/3) ... 0.05(*/5)
7. error of selection in actuating a component (switch, hand valve) or in referring a system state display	0.0005(*/10) ... 0.05(*/5)
8. incorrect performance of a dynamic (knowledge-based) action under extremely high stress	0.25(*/5)

NHEP: nominal human error probability (median)

UF: uncertainty factor (error factor)

For German translations, see [Reer 88]

Table 19: Important operator errors contained in the THERP database

· Direct estimates of the handbook authors	
· Based on empirical data	* nuclear power plant experience * simulator studies * industrial and military studies * "artificial" experiments * source not clear
– Empirical probabilities directly adopted	
– Empirical probabilities modified with factors factors are estimates of the handbook authors factors are based on empirical data	

Table 20: Sources of Swain's handbook error probabilities

1. Dearth of data on human performance in NPPs that is in a form suitable for HRA.
2. Inexactness of models of human performance that purport to describe how people act in various situations and conditions.
3. Inadequate identification of all relevant performance shaping factors, of their interactions and effects.
4. Limitation in the skill and knowledge of the human reliability analyst.
5. Variability in performance within an individual and among the performances of different individuals.

Table 21: Variability causes to be covered by the uncertainty factors in Swain's handbook. From [Swain 83] (pages 7-9 to 7-10)]

6.3 Performance shaping factors (PSFs)

DESCRIPTION

An analysis carried out with THERP takes a number of performance shaping factors (PSFs) into consideration, which have a reliability-decreasing or -increasing effect. The PSFs specified in THERP can be roughly classified into four groups:

1. PSFs already included in selecting the nominal error probability (NHEP), e.g.: in a long written procedure the error of omission probability is 0.01 per step, in a short one 0.003 ([Swain 83], Table 20-7).
2. PSFs specified in the form of factors by which an NHEP is to be multiplied, e.g.: for medium-high stress the failure probability of a rule-based action increases by a factor of 2 ([Swain 83], Table 20-16).
3. PSFs specified as rules for modifying an HEP within their uncertainty limits, e.g.: use the upper bound of diagnosis failure probability if the diagnosis of the analysed accident is not regularly trained ([Swain 83], Table 12-5).
4. PSFs which are mentioned, but for which no quantitative data are given ([Swain 83], Chapter 3).

The THERP user is generally urged to verify the transferability of the THERP HEPs to the respective analysis, using the PSFs discussed in Chapter 3 of the handbook, and to make modifications within the uncertainty bounds, if required ([Swain 83], page 3-10).

EVALUATION

The THERP method is to be evaluated positively with respect to PSF quantification.

THERP is open for the incorporation of numerous plant-specific details. An evaluating compilation of such details can be found e.g. in [Reer 94d].

The detailed discussion of PSFs of significance for human action in NPPs is one of the principal items of Swain's handbook. Concrete studies as well as psychological and ergonomic findings are quoted for numerous PSFs. The authors do not know of any other method for HRA in which PSFs are discussed more comprehensively than in THERP.

Many statements in Swain's handbook are highly topical even today. As early as in 1983, for example, the benefits and disadvantages of symptom-oriented (as an alternative to event-oriented) procedures in accident diagnosis were discussed ([Swain 83], p. 3-82). In this context, it is definitely not advisable to readily assume increased diagnosis reliability in the presence of symptom-oriented procedures. The authors recommend thorough reading of the studies quoted in Swain's handbook ([von Herrman 83a], [von Herrman 83b]), in which also problems with symptom-oriented procedures are discussed.

Psychological and cognitive factors of significance in making nontrivial decisions are described in THERP, but the associated quantification problem – especially with respect to knowledge-based performance – is only inadequately solved. All methods of HRA known to the authors exhibit deficiences in this respect (see Section 2.3.2). For this reason, the quantification of knowledge-based performance in decision-making will be discussed as a separate evaluation criterion.

6.4. Operator decisions at the knowledge-based level

DESCRIPTION

The THERP diagnosis model ([Swain 83], Chapter 12) is suitable for quantifying knowledge-based performance in accident diagnosis. The identification of an abnormal event to the effect of knowing with which component or system the problem caused by the event can be eliminated or reduced is defined as diagnosis in THERP. The probability for a failure of the correct diagnosis is modelled in a time-dependent manner.

Actions after diagnosis are classified by Swain & Guttmann into the categories "step by step" (corresponding to rule- or skill-based) and "dynamic" (corresponding to knowledge-based). An increased error probability is recommended for knowledge-based actions, higher by at least a factor of 2.5 than for rule-based actions ([Swain 83], Table 20-16).

In addition, Swain's handbook also contains probabilities for errors whose origin is also influenced by knowledge-based performance elements, e.g. failure to notice that a measuring instrument displays a wrong reading (due to a technical defect), although the wrong reading can be detected by using further system state parameters ("cross check") ([Swain 83], Table 11-13). The THERP dependence model also allows the consideration of errors which greatly depend on knowledge-based performance, e.g. "... neither of the operators expects the other to make mistakes" ([Swain 83], p. 10-18).

EVALUATION

The quantifiability of knowledge-based performance by THERP is to be classified as satisfactory.

In principle, THERP allows the quantification of a number of errors related to knowledge-based performances. However, the assumed quantification models, e.g. time-reliability correlation for diagnosis, simple PSF of 2.5 for subsequent actions) are too general and mostly based exclusively on expert estimates. There are no indications for the quantitative implementation of many PSFs related to knowledge-based performance.

[Swain 83] (Chapter 12) admits that the THERP diagnosis model is of a very speculative nature and that THERP exhibits deficiencies in the quantification of operator decisions after diagnosis.

In [Swain 88] the event category of diagnosis failure (i.e. no correct diagnosis) is divided into the subcategories

- no diagnosis and
- wrong diagnosis.

This further development is to be classified as positive. In quantitative application, however, difficulties are to be expected, especially due to a lack of probability data on the occurrence and correction of a wrong diagnosis.

6.5 Time reliability correlation (TRC)

DESCRIPTION

In principle, a failure probability (HEP) of a manual action during accidents, calculated with THERP, is a time-dependent variable. The time dependence results from the THERP diagnosis model in which the estimated course (Table 22) of the diagnosis failure probability Pr(T>t) is described as a function of the time t available for diagnosis. The total failure probability follows:

$$HEP(t) \approx Pr(T>t) + Pr(A)$$

where Pr(A) is the failure probability of post-diagnosis action(s). For the assessment of the complementary distribution function Pr(T>t), the diagnosis time modelled as random variable T, the THERP user may choose among three curves taking the education- and training-related knowledge of the control room personnel as a guideline (Table 23): the better the knowledge about the accident to be analysed, the faster a correct diagnosis is to be expected.

The available diagnosis time t depends on the dynamic plant behaviour and on the time requirement t_A of the actions after diagnosis. This time requirement – in contrast to the diagnosis time requirement T – is not modelled as a random variable in THERP, but as a deterministic variable. Exercises or surveys already carried out are to be used for estimating t_A [Swain 83] (p. 6-11). ASEP [Swain 87] (Table 8-1) specifies estimates which can be used for determining t_A if no empirical time data are available. The following relation then often results for the available diagnosis time (t)

$$t = t_{total} - t_A,$$

where t_{total} is the total permissible time from the start of an accident to success of the required manual action.

For certain cases of analysis, the dynamic plant behaviour is the main criterion for estimating t. An example is shown in Figure 8 analysing a change to the circulation mode. This action must be taken 2 min, at the latest, after the alarm warns of an excessively low refuelling water storage tank level. The alarm is to be expected 30 min after the occurrence of an accident. These 30 min are assumed in [Swain 83] (p. 21-12) as the available diagnosis time, which means that 2 min are assumed to be sufficient for the required switching action.

t	1 min	10 min	20 min	30 min	60 min	1500 min
MDN(t)	1.0	1.0E–1	1.0E–2	1.0E–3	1.0E–4	1.0E–5
LB(t)	1.0	1.0E–2	1.0E–3	1.0E–4	3.3E–5	3.3E–6
UB(t)	1.0	1.0	1.0E–1	1.0E–2	3.0E–3	3.0E–4

MDN = median; LB = 5 % lower bound; UB = 95 % upper bound; t = available diagnosis time. The parameters between two t-values are calculated by a loglinear interpolation. After [Swain 83] (Table 20-3).

Table 22: Time-dependent parameters of diagnosis failure probability

The probabilities in Table 22 are applicable if only *one* abnormal event occurs. If a second or third event occurs, THERP assumes that the time required for diagnosis is extended by 10 min for each additional event; this estimate is based on simulator studies by [Woods 82]. For two events, for example, a median (MDN) of 0.1 (instead of 0.01 for one event) would result for t = 20 min, for three events the median would be MDN = 1.0. This 10-minute rule is further illustrated in ASEP [Swain 87] (Table 8-1).

Moreover, it should be noted that, in additon to TRC, THERP gives guidelines to quantify some other aspects of time, especially in periodically scanning or resuming attention to an alarm; see [Swain 83] (Tables 20-24 and 20-25).

EVALUATION

THERP permits a satisfactory consideration of the time factor in the HRA.

Curves are available with which the diagnosis failure can be quantified as a function of the available diagnosis time and the general level of operator knowledge. It is positive to note that the time required for diagnosis is modelled as a random variable, since this is a highly dynamic task (see Figure 5). Another positive aspect is that the diagnosis failure probability HEP(t) is modelled as relatively slowly decreasing and with a high uncertainty factor (UF = 30) with increasing time (t). This reflects the extrapolative nature of the model in the range of high t-values; in comparison: the empirical diagnosis times published in [Villemeur 86] are all below 25 min.

1. Use upper bound (UB) if
   a. the situation is not covered in training or
   b. the situation is covered but not practised except in initial training of operators for becoming licensed, or
   c. the talk-through and interviews show that not all the operators know the pattern of stimuli associated with the situation.
2. Use lower bound (LB) if
   a. the situation is a well-recognized classic incident and
   b. the operators have practised this situation in simulator requalification exercises, and
   c. the talk-through and interviews indicate that all the operators have a good recognition of the relevant stimulus patterns and know which actions (procedures) to follow.
3. Use median (MDN) if
   a. the only practice for coping with the situation is simulator requalification exercises and all operators have had this experience, or
   b. rules 1 and 2 above do not apply.

From [Swain 83] (p. 12-23). The guidelines relate to the time dependences outlined in Table 22.

Further guidelines are contained in ASEP ([Swain 87], Table 8-1), e.g. the conditions under which a reduced diagnosis probability (lower bound) may be assumed in the case of symptom-oriented procedures.

Table 23: Guidelines for assessing the time dependence of diagnosis failure probability.

A negative feature is that THERP assumes the same diagnosis failure time dependence for all accidents requiring response to a single abnormal event. Simulator experience confirms that accident situations differ with respect to the required scope of diagnosis [Hoffmann 89].

A positive aspect is that THERP allows the quantification of an increased scope of diagnosis due to the occurrence of several abnormal events. The practical application of this option is unclear. Table 24 shows, for example, that there may be several criteria for successful diagnosis within an accident sequence. As a general rule, the scope of diagnosis increases in proportion to the severity of the accident. However, Swain's handbook does not contain any clear guidelines for the appraisal of conditions (e.g. failure to start the emergency system) increasing accident severity as additional events in the sense of the 10-minute rule (diagnosis is extended by 10 min per additional event). The authors recommend that the study [Woods 82], quoted by [Swain 83] (p. 12-14), from which the 10-min rule was derived, should be verified.

The inadequate consideration of the time factor for post-diagnosis actions is a further weak point of THERP. There is no clear indication that the time required for unnecessary preceding

actions should also be taken into consideration in addition to the time required for the (important) actions necessary to control the accident. A necessary action is mostly the end of a series of activities which also contribute to the total time required. Furthermore, it is advisable that the time required for the primarily rule-based actions after diagnosis should also be modelled as a random variable. The assumption of a fixed (non-distributed) value t_A in THERP means that the probability for significantly exceeding or remaining below t_A is neglected. This is contradicted by some data which confirm that considerable fluctuation widths with respect to the times required are also to be expected in performing rule-based actions; cf. the time data published in [Haas 82] or [Hannaman 88], e.g. a factor of 20 between the earliest and latest time of performance. The probabilistic modelling of the time dependence of diagnosis and subsequent actions also has the advantage that the success probability of subsequent corrections as a function of the residual time available can be quantified [Reer 1994b] in a more realistic way.

Event sequence (PWR)	Action to be diagnosed
◦ failure of steam generator feeding ◦ failure of automatic actuation of emergency feedwater system	manual start of emergency feedwater system
◦ manual start of emergency feedwater system fails	start of emergency system
◦ start of emergency system fails	steam generator emergency feeding with mobile fire fighting pump

Table 24: Hypothetical example concerning the dependence of the scope of diagnosis on accident severity

6.6 Dependences between errors

DESCRIPTION

The model recommended in THERP for the quantification of dependences between events relating to human action corresponds to the beta factor model known from technical reliability analysis. The following applies to the conditional occurrence of an event B:

$$pr(B|A) = \left[\frac{1+(n-1)pr(B)}{n}\right] = \beta + (1-\beta)pr(B)$$

where:

pr(B|A) = probability for B on condition that A has occurred;

pr(B) = probability for the occurrence of B independently of A, also denoted as basic human error probability (BHEP);

$\beta = 1/n$ = level of dependence (beta factor), corresponds to the probability that the cause for the occurrence of A will also lead to the occurrence of B;

$n = 15$ ($\beta = 0.05$) for low, $n = 7$ ($\beta \approx 0.15$) for moderate, $n = 2$ ($\beta = 0.5$) for high and $n = 1$ ($\beta = 1.0$) for complete dependence between A and B.

All levels of dependence recommended in the model are not based on empirical data, but on a subjective discretization of the value range from $\beta = 0$ (no dependence) to $\beta = 1.0$ (complete dependence). Swain & Guttmann give a number of guidelines for selecting the level of dependence. The most important criteria are summarised in Table 25.

The quantification of personnel redundancy in an accident sequence is one of the most frequent applications of the dependence model. THERP recommends here the assumptions summarised in Table 26.

EVALUATION

THERP permits a satisfactory consideration of dependences that may occur between operator actions.

The dependence model is uncomplicated and its principle methodologically confirmed, since it is based on the beta factor method recognized in the PRA specialist community. The beta factor method is frequently considered to be too pessimistic for high redundancy, but this weak point does not apply to THERP handling. If there is more than one recovery possibility (i.e. for the AND operation of more than two non-success events) the THERP method provides for a repeated consecutive application of the dependence model. This corresponds to the MGL method ("multiple greek letter"), a method also recognized in the PRA community.

The discretization of the steady range from $\beta = 0$ to $\beta = 1$ into five discrete points in THERP facilitates application. The resultant inexactness is to be regarded as insignificant in view of the data problem and the PRA-typical uncertainties. Particularly helpful for the THERP user are the guidelines which THERP gives to assess the level of dependence (Table 25).

In technical reliability analysis it is a standard procedure to empirically determine the parameters of the dependence model used. This standard is not reached for the THERP

dependence model, so that the appraisal is here 'only' satisfactory despite its user-friendly features.

ZD (zero dependence)	Rare, unusual, time (> 1 min) and spatial differences between 2 tasks. Example: verification of 2 instruments at different locations in the control room during normal operation.
LD (low dependence)	If there is any doubt as to an assessment of independence (ZD); slight time (approx. 1 s to 60 s) and spatial differences between 2 tasks. Examples: between 2 operators newly acquainted with each other; between shift supervisor and new operators; between shift supervisor and other operators, if the system status must be assessed after a transient.
MD (moderate dependence)	Obvious relationship between 2 tasks; slight time (approx. 1 s to 60 s) and spatial differences. Example: between shift supervisor and other operators for tasks interacting with each other.
HD (high dependence)	Substantial relationship between 2 tasks; very slight time (approx. 1 s) and spatial differences. Examples: operator A significantly differs from operator B with respect to prestige and authority, so that the behaviour of A has a great influence on the behaviour of B; between shift supervisor and reactor operator.
CD (complete dependence)	Rare, 2 tasks are practically performed simultaneously in space and time. Example: operation of paired switches.

After [Swain 83] (Chapters 10 and 18)

Table 25: THERP guidelines for assessing the level of dependence between two tasks or acting persons.

The creation of an empirical database through the dependent occurrence of incorrect actions is therefore an important task for future research in the field of HRA. Due to the flexibility of human behaviour a high research effort must be expected to satisfy PRA needs.

In addition to this, basic investigations should be initiated because the modelling of dependence is important from various aspects of an HRA, e.g. dependence between (among)

- errors (within a given sequence of actions),
- failures of two sequences of actions,
- PSFs,
- times in action performances, or
- required and available time to perform a given task.

Time after first response to an abnormal event (accident)		(a) Personnel available to cope with the accident	(b) Dependence levels
(1)	0 to 1 min	one reactor operator (RO)	
(2)	≈ 1 min	one reactor operator (RO) shift supervisor or deputy	 HD with RO
(3)	≈ 5 min	one reactor operator (RO) experienced senior operator shift supervisor one (several) assistant operator(s)	 HD with RO LD to MD with others
(4)	≈ 15 min	one reactor operator (RO) experienced senior operator shift supervisor shift technical advisor one (several) assistant operator(s)	 HD with RO LD to MD with others LD to MD with others for diagnosis and major accidents HD to CD for detailed operations

Table 26: THERP assumptions on personnel available to cope with an accident and existing levels of dependence. After [Swain 83] (Table 20-4)

6.7 Error correction (recovery)

DESCRIPTION

The THERP method permits the quantification of a number of possibilities for correcting an operator error [Swain 83] (Figure 20-1):

1. personnel redundancy: the acting operator is checked by another operator;
2. subsequent procedure step or task;
3. annunciating display (audible alarm);
4. periodical scanning inside control room (CR) or walk-around inspection outside CR.

The THERP model assumes that safety-relevant deviations from the nominal status can be detected by CR scanning or walk-around inspections during normal operation. To this end, a number of situation-specific error probabilities are listed which, among other aspects, depend on the respective rounds strategy [Swain 83] (pages 11-25, 11-29, 19-15, 19-22). For the quantification of personnel redundancy in normal operation THERP provides a number of situation-specific control error probabilities ranging between 0.001 and 0.5 [Swain 83] (Table 20-22). In the event of an accident, THERP recommends that personnel redundancy should be quantified through the dependence model [Swain 83] (Table 20-18). This model should also be used for the quantification of recoveries by subsequent tasks; see [Swain 83] (page 10-7).

According to THERP, the response to an audible alarm is a recovery factor for the failure of diagnosis in an accident. Swain & Guttmann recommend a so-called alarm-response model in which the error probability depends on the total number of alarms and the place where the alarm considered as recovery factor ranks [Swain 83] (Table 20-23). Several alarms may have to be combined to so-called perceptual units. Moreover, the failure to notice an alarm is to be quantified considering the personnel redundancy existing in the control room. Figure 8 shows an example of the quantification of the recovery potential of an alarm, as recommended in THERP. The probability (0.0002) that the diagnosis failure (0.001) is not recovered results from the product of the probabilities

- 0.003 for an operator's failure to notice a particular alarm out of a total of 5 alarms,
- 0.5 for failure to check by a second operator with high dependence on the preceding error and
- 0.15 for failure to check by a third operator with moderate dependence on the preceding errors.

The diagnosis failure probability is thus reduced from 0.001 to $0.001 \cdot 0.0002 = 2E\text{-}7$ due to the alarm.

Evaluation

Although the THERP method exhibits both positive and negative aspects for a quantification of the recovery of an error made, the authors rather consider an overall negative evaluation to be appropriate due to grave misjudgements of control situations in accidents.

A positive feature is the systematic consideration of different recovery possibilities (personnel redundancy, alarm, subsequent procedure step, scanning, walk-around). The recommendations given in THERP for their quantitative application deserve a negative appraisal, especially for recovery actions in accident situations. The dependence on the psychological causes of an error to be recovered is not adequately taken into account. Psychological findings, which can be substantiated by simulator and accident experience, confirm the significance of this dependence type.

In the analysis examples [Swain 83] (Chapter 21) the existence of an alarm indicating the manual action to be initiated is quantified as an independent recovery factor for the failure of diagnosis. This is a considerable underestimation of the diagnosis problem in stress situations. On the one hand, this optimistic modelling contradicts in a certain way the finding quoted by [Swain 83] (p. 3-42) himself that operators tend to ignore information under stress. On the other hand, psychological findings indicate a strong dependence between diagnosis failure and failure of the respective control mechanisms. The failure of diagnosis is a highly knowledge-based error. Such failure can be denoted as a mistake because the operators do not implement a correct plan. According to [Reason 90] (p. 62) recovery mechanisms are only effective to a very limited extent in the presence of mistakes. Semmer (1994) gives a very appropriate judgement:

> »Once we have made a diagnosis,
>
> - we notice above all information supporting our opinion;
> - we frequently ignore information contradicting our opinion;
> - we look for explanations fitting our opinion in case we cannot overlook the contradicting information.
>
> ...
>
> All these tendencies are enhanced by ... stress« (translated from [Semmer 94]).

These findings can be convincingly substantiated by simulator and accident experience. Not a single error out of 19 in identifying the system status in simulated accidents was recovered by the control room crew themselves [Woods 84]. In the Oyster Greek NPP (1979) the operators did not notice an excessively low water level for 30 min, although this was acoustically signalled by a level alarm [Pew 81]. In the Davis-Besse NPP (1985), too, an acoustic monitor proved ineffective, although it had been specifically added due to the TMI accident. The operators did not notice that the pressurizer relief valve did not close any more after the third opening operation, although closure failure was signalled by the acoustic monitor [NUREG-1154]. The authors therefore recommend that the recovery effect of an alarm should be quantified with the aid of a dependence model (Section 6.6) assuming at least a low dependence.

It should be considered positive that THERP recommends the dependence model for the quantification of personnel redundancy in an accident. However, the personnel recovery

possibilities and levels of dependence recommended in Table 26 show a slightly optimistic tendency. According to the evaluation of the Davis-Besse accident in [Reer 93] (Chapter 2), three of a total of five critical actions were performed without personnel redundancy. In the above-quoted simulator study of [Woods 84] only 50 % of the errors in executing actions after identification of the system status were found to be recovered. Moreover, the assumption made under point 4b in Table 26 that a lower dependence should be assumed for the failure to recover diagnosis errors than for the failure to recover other incorrect actions contradicts the above-quoted psychological findings ([Reason 90], [Semmer 94]). The authors repeat their recommendation that a database should be created for the dependent occurrence of operator errors.

6.8 Uncertainties

Description

Probabilities for the evaluation of human reliability are modelled by THERP as uncertain variables. For this reason, an uncertainty interval (lower bound (LB); upper bound (UB)) is also specified for a probability in addition to a point value (HEP). The interval should contain the true value of the probability with a high certainty (90 %). The following uncertainties are to be covered by these intervals according to Swain & Guttmann [Swain 83] (p. 7-9):

1. random fluctuations in human reliability;
2. imperfect knowledge of the THERP user.

These two main groups are specified in more detail; see Table 21 in this report.

The uncertainty is mathematically modelled by a lognormal distribution. An estimated point value of an error probability is defined as the median (MDN) of this distribution. The uncertainty is expressed by a scattering factor (K) and - according to the lognormal distribution - quantified by a multiplicative model:

$MDN = LB \cdot K$, $UB = MDN \cdot K$, with the restriction that the value 1.0 must not be exceeded. Details on the mathematical operations can be found in Appendix A of Swain's handbook [Swain 83].

Evaluation

The mathematical treatment of the uncertainty (of an error probability) recommended in THERP can be classified as satisfactory.

The assumed lognormal distribution model is easy to handle and sufficient for PRA purposes. For probabilities, however, it is only mathematically correct to a first approximation.

In principle, a lognormal distribution is suitable for quantifying the uncertainty of a multiplicative variable whose multiplicity is defined on a ratio scale. However, as already criticized in Section 6.1 concerning the multiplicative PSF model, a failure probability (P) or a success probability (Q = 1-P) is only approximately multiplicative.

It would be mathematically correct to model the uncertainty of P via the odds ratio (Y = Q/P), because this is a true multiplicative variable:

$$p = \frac{1}{(1+y)} = \frac{1}{(1+e^{x})}$$

where:

Y = lognormally distributed variable with y = (1-p)/p as the median and K as the uncertainty factor;

X = normally distributed variable with μ = ln y and σ = (ln K)/1.645;

p = point value of the failure probability from Swain's handbook;

K = uncertainty factor of p from Swain's handbook.

This will be illustrated by an example. With p = 0.1 as the point value of the failure probability (P), the odds ratio has a point value of y = 9 : 1 = 9. The following values are obtained with an uncertainty factor (for Y) of K = 5 for the lower bound (LB) and the upper bound (UB) of P:

$$LB(P) = \frac{1}{1+y \cdot K} = \frac{1}{1+9 \cdot 5} \approx 0.022$$

$$UB(P) = \frac{1}{1+\frac{y}{K}} = \frac{1}{1+\frac{9}{5}} \approx 0.36$$

The authors therefore recommend that the point value of a failure probability from Swain's handbook should be interpreted as the expected value of a beta distribution, with LB as 5 % quantile and UB as 95 % quantile, LB and UB being calculated from the above equations. In this way, the following inconsistencies arising from the treatment of uncertain failure probabilities recommended in THERP can be eliminated:

- Swain & Guttmann define the point value (p) of a failure probability as the quotient of the number (a) of errors and the number (n) of error opportunities [Swain 83] (p. 2-17). This point value is interpreted in THERP as the median. According to the probability theory, however, this is an expected value (mean). Since the mean is always greater than the median according to the lognormal distribution assumed in THERP, the mean obtained is systematically too high because the empirical point values are based on the estimate

p = a/n. Example: for an uncertainty factor of 10, the mean of a lognormal distribution is always higher by a factor of 2.7 than the median.

- Swain & Guttmann define the success probability Q (i.e. the probability for the complementary failure event) by the following relation: Q = 1-P [Swain 83] (p. 2-18). All point calculations carried out with this relation are performed with medians in THERP (see Figure 8). For lognormally distributed failure probabilities, however, a point calculation, in which the relation Q = 1-P must be used, only leads to a mathematically correct result if the calculation is carried out with expected values. For a beta-distributed failure probability, this relation can also be used for quantile estimates, since the following symmetry relation is valid:

$$Q_{1-\alpha} = 1 - P_{\alpha}$$

From the above calculated values for the upper bound (UB = 0.36) and lower bound (LB= 0.022) of the failure probability (with p = 0.1 as the point value and K = 5 as the uncertainty factor relative to the odds ratio), it is thus possible to directly calculate the values for the upper and lower bounds of the success probability:

UB(Q) = 1 - LB(P) = 1 - 0.022 = 0.978

LB(Q) = 1 - UB(P) = 1 - 0.36 = 0.64

This symmetry relation usually also applies to the mean value:

q = 1 - p = 1 - 0.1 = 0.9

The success probability would thus have an expected value of 0.9 and is within the interval (0.64; 0.978) with 90 % certainty.

However, the discrepancies arising from the lognormal distribution do not present a serious weakness of the THERP method because they have an effect towards the safe side and thus rather lead to pessimistic results.

6.9 Summary evaluation

On the whole, the THERP method can be considered as satisfactory to good, if applied in a competent manner. The weak point discussed in Section 6.7 with respect to a too optimistic tendency in quantifying recoveries of errors can be eliminated by utilizing the scope for discretion which THERP expressly permits the user.

Apart from the obvious requirement to improve the database, a requirement that could be fulfilled in the medium term can be derived from the evaluations in this report. In order to achieve more realistic conditions with respect to the quantification of diagnosis activities, decisions at the knowledge-based level and time dependences, it should be examined how

subaspects from more recent methods ([Gertman 92], [Moieni 94], [Reer 93], [Mosneron 90]) can be derived for improvements.

7 New developments with emphasis on time-reliability correlations in accident diagnosis

The time curves of diagnosis failure probabilities in the THERP model [Swain 83] reflect expert estimates by Swain & Guttmann. These expert estimates are based on a consensus achieved at the workshop described in [Oswald 82].

In the meantime, the data situation in the field of diagnosis reliability has improved. Time curves are available which are based on the simulator studies initiated by EdF (Electricité de France) and EPRI (Electrical Power Research Institute, USA). These studies have led to the development of self-contained methods: EdF [Mosneron 90], HCR [Hannaman 84] and HCR/ORE [Moieni 94], HCR/ORE being a methodological further development of HCR, so that the 1984 HCR model version is only of historical significance by now.

These methods will be briefly described and roughly evaluated in the following.

7.1 EdF'sPHRA

BRIEF DESCRIPTION

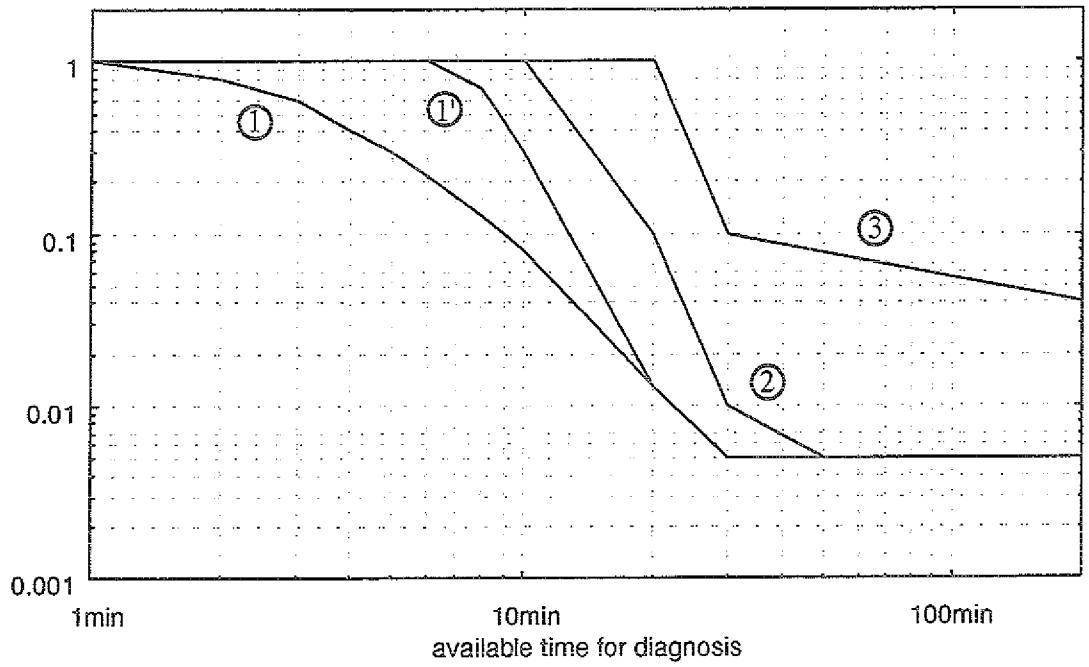
The method of EdF for Probabilistic Human Reliability Analysis (PHRA) is essentially based on French simulator studies and on models from THERP (pessimistic diagnosis curves, dependence model) and ASEP (maintenance error model). Note: EdF recommends to use the wording PHRA [Mosneron 94] (instead of only HRA) if HEP quantification is involved.

This chapter only deals with the quantification of errors in an accident sequence. A relevant detailed description (in French) of the EdF method can be found in [Mosneron 89], brief descriptions and illustrations (in English) in [Mosneron 90], [EPS 900], [EPS 1300] and [Mosneron 94]. The EdF method has been used in [EPS 900] and [EPS 1300].

EdF distinguishes in the same way as THERP between diagnosis and execution of post-diagnosis actions. In contrast to THERP, EdF considers the possibility that the accident can be successfully controlled even with an incorrect diagnosis; however, the corresponding conditional error probability is rated very high. THERP quantifies this error probability with 1.0: if no correct diagnosis, then failure of accident control by the personnel. In order to assess the diagnosis failure probability as a function of available time, EdF distinguishes between four time curves (Figure 9):

- two pessimistic time curves based on the THERP diagnosis model (curves 2 and 3),
- two less pessimistic time curves based on French simulator data (curves 1 and 1'),

where the complexity of the accident situation is to be taken as the selection criterion.



comparatively easy diagnosis (French simulator data):
- situations without safety injection (curve 1)
- situations with safety injection (curve 1')

comparatively complicated diagnosis (derived from [Swain 83]):
- average situations (curve 2)
- most difficult diagnosis (curve 3)

Figure 9: Time-dependent curves of diagnosis failure probability recommended in EdF'sPHRA [EPS 900] (page 80)

According to the EdF model concept, there is a so-called "residual probability" in the diagnosis; this is the probability that the correct diagnosis will never be made by the crew itself without any help (from the safety engineer or the technical support center). This probability is 0.03 (in the case of extremely difficult accident diagnosis, curve 3 in Figure 9) or 0.005 (in other cases, curves 1, 1' and 2), see [EPS 1300] (page 112).

The EdF method furthermore provides for the possibility of quantifying some kinds of misdiagnoses leading to unrequired actions (unwarranted shutdown of safety injection after LOCA; unwarranted isolation of steam dump to the atmosphere after steam generator tube rupture) [Mosneron 94]. For these specific cases, the execution model (see below) can be used (though it is generally used for "pure" execution, not including diagnosis) [Mosneron 95].

Every important (critical) action after diagnosis is to be quantified with the so-called "execution model" [Mosneron 90] (Section 2.3.3.3):

$$P_E = P_B \cdot K_F \cdot P_{NR}$$

where:

P_E = probability that execution of action fails;

P_B = basic probability for failure;

K_F = modification factor with respect to the action context;

P_{NR} = recovery failure probability.

For the model parameters (P_B, K_F, P_{NR}) situation-specific numbers are given. These numbers are based on a combination of simulator data and expert judgement.

If a reliable estimate on the average time requirement is available for the entire action (including diagnosis), then the time-dependent error probability is to be determined according to standardised (i.e. normalised) curves outlined in Table 27. In this case, the following additional contributions (see above) to failure of the manual action are to be quantified:

- residual probability that the diagnosis will never be made;
- probability that the execution after diagnosis fails.

$t/T_{0.5}$	curve 1	≈0.3	≈1	≈2.5	≈3.2	≈4.8
	curve 2			≈3.1	≈4	≈7.6
pr(T>tla)		1.0	0.5	0.1	0.05	0.01

Table 27: Some typical values of the probability pr(T>tla) that the action is not performed within t on condition that the action is performed at all (derived from [Mosneron 90] (Figure 3). In selecting the curve, the complexity of the situation and the availability of experimental results must be taken into consideration [EPS 1300] (page 112).

These two approaches - (1) TRC via diagnosis curve, (2) TRC via standardised curve - may be illustrated by the following example (taken from [EPS 1300], pages 111-113).

Accident Shutdown state, break in primary system, safety injection system (SIS) fails to start up automatically.

Action Operators bring SIS into operation.

Analysis (1) Incorporation of TRC via diagnosis curve

$p = p_D\, p_{EID} + (1 - p_D)\, p_{EId} = 0.005 \cdot 1 + (1-0.005) \cdot 0.006 \cong \underline{0.011}$,

where:

p = total failure probability;

$p_D = 0.005$ = probability of diagnosis failure; curve 2, t = 50 min available;

$p_{E|D} = 1$ = conservative assessment of the probability that execution of action fails, given diagnosis failure;

$p_{E|d} = p_B \, K_F \, p_{NR} = 0.06 \cdot 1 \cdot 0.1 \cong 0.006$ = probability that execution of action fails, given correct diagnosis;

$p_B = 0.06$ = basic execution failure probability; to choose between 0.02 and 0.06, according to the type of error;

$K_F = 1$ = factor for contextual correction; to choose between 1, 1/3, and 3, depending upon the context in question;

$p_{NR} = 0.1$ = probability of non-recovery; to choose between 0.03, 0.1, 0.3, and 0.6 depending on the recovery factors (explicit human redundancy and/or alarm signal) and available recovery time (<30min or >30min).

(2) Incorporation of TRC via standardised (normalised) curve

$p = p_{RA} + (1 - p_{RA}) \, p_{A(t)|a} = 0.007 + (1 - 0.007) \cdot 0.002 \cong \underline{0.009}$,

where:

$p_{RA} = p_{RD} \, p_{RE|D} + (1 - p_{RD}) \, p_{RE|d} = 0.007$ = residual probability that no action will be implemented;

$p_{RD} = 0.005$ = residual probability that accident diagnosis will never take place; to choose between 0.03 (difficult diagnosis) and 0.005 (other cases);

$p_{RE|D} = 1$ = conservative assessment of the residual probability that the action will never be implemented, given that the diagnosis has failed;

$p_{RE|d} = p_B \, K_F \, p_{NR} = 0.06 \cdot 1 \cdot 0.03 \cong 0.0018$ = residual probability that the action will never be implemented, given that the diagnosis has been successful; where $p_{NR} = 0.03$ (instead of 0.1) since it is a residual probability;

$p_{A(t)|a} = pr(T > t|a) = 0.002$ = probability of failure to implement action within time limit t, given that action will eventually be implemented; from standardised curve 1 at $t/T_{0.5} = (50 \text{ min})/(7 \text{ min}) \cong 7$.

The intervention by an outside expert, the so-called "safety engineer", is quantified by EdF as an additional opportunity for recovery. The safety engineer is not permanently present in the control room, but can (and should) be called in the event of an accident. The failure probability of an error recovery by the safety engineer is modelled dependent on time and situation [EPS 1300] (Chapter 5.3.3.5). The required data are based on simulator tests and on the dependence model from [Swain 83]. Finally, the intervention of the technical support centre is very roughly modelled considering that no important error occurs later than four hours after the beginning of the accident ([EPS 1300], page 107].

EdF's model of recovery by the safety engineer may be illustrated in the following example (taken from [EPS 1300], page 115).

Accident	Shutdown state, break in primary system, safety injection system (SIS) fails to start up automatically, operators (without safety engineer) fail to bring SIS into operation.
Action	Safety engineer recovers operators' failure.
Analysis	$p = A(t) + [1 - A(t)]\ p_{Ela} = 0.15+(1-0.15)\cdot 0.05 \cong \underline{0.2}$, where: p = probability that recovery by the safety engineer fails; A(t) = 0.15 = probability of absence of the safety engineer at a given time t = 50 min (time required for corrective action is assessed as negligible) beyond which it is too late to implement corrective action; A(t) is obtained from the simulator tests and plant-specific survey; p_{Ela} = 0.05 = probability that recovery fails when the safety engineer is present; to choose between 0.05, 0.1, and 0.3, depending on the type of written instructions for the corrective action and the time available to implement the corrective action; the numbers are derived from simulator tests and the dependence model in [Swain 83].

Furthermore, EdF'sPHRA includes a method for assessing the uncertainty of an estimated HEP point value, see [EPS 1300], Section 5.3.1.6.

Finally, the reader should take notice that EdF's methodology has been updated in 1995 to account for symptom-based procedures and the new organisation of the crews [Mosneron 96]. The main changes are:

- The standardised curves (see Table 27 in this report) will be no longer used.
- The guidelines for assessing the probability (p_{NR}) of non-recovery (by the crew) have been adapted to the structure of symptom-based procedures (with new recovery factors).
- The HEPs (A(t), p_{Ela}) of the safety engineer model have been reduced to account for improvements in the organisation.

ROUGH EVALUATION

First at all, the reader should recall that the authors' evaluation (in Chapters 5 to 8) mainly concerns the quantitative component of an HRA. Therefore, especially for EdF'sPHRA, only a limited view will be presented, because EdF's methodology gives priority to the qualitative component of an HRA:

»Quantification models are not so important for us. We think that knowledge about operator behaviour, habits, attitudes, is the core of PHRA (see [Mosneron 94]). This knowledge is extracted from:

- simulator tests;
- interviews with operators, trainers and operation engineers;
- on site surveys.

So, only a specialist who has acquired this knowledge can perform a good PHRA. Then, quantification models only help him to structure his judgement« [Mosneron 96].

Concerning HEP quantification, the EdF method can be referred to – with reservations – as a positive new development. This is a judgement with reservations because the authors lack detailed information on the method. In particular, there is a lack of more concrete instructions concerning the procedure to follow in applying the method and there are no details concerning the procedure applied by EdF in deriving model data (primarily probabilities) from simulator data (observations concerning the behaviour of the personnel). Actually, this information exists, but is not publicly available.

In comparison to the THERP diagnosis model, the EdF diagnosis model is more advanced since it is based, at least in part, on genuine empirical data. However, the method description does not refer to a clear distinction between diagnosis and execution time. Consequently, some methodological aspects of EdF'sPHRA are not totally clear. The case study presented in [EPS 1300] (pages 111-113) demonstrates such a pitfall. Two approaches are presented to quantify the failure to bring the Safety Injection System (SIS) into operation in an accident during shutdown:

1. Given no data about the median time $T_{0.5}$ required to implement the action, the time-dependent diagnosis failure probability is determined by diagnosis curve 2 at $t = 50$ min available; cf. Figure 9 in this report.
2. Given a data-based estimation of $T_{0.5} = 7$ min, the time-dependent probability of failure to implement the action is determined by the standardised curve 1 at $t/T_{0.5} \cong 7$, also with $t = 50$ min as the input value; cf. Table 27 in this report.

Although it is not explicitly stated in [EPS 1300], the authors would conclude from this that the post-diagnosis execution time is considered as negligible in the example.

However, the $T_{0.5} = 7$ min estimation (i.e. a 0.5 HEP at 7 min) seems to be unrealistic with respect to diagnosis curve 2 because in this curve the failure probability is 1.0 for times below 10 min. The authors would conclude from this that the first approach (TRC via diagnosis curve) is more pessimistic than the second approach (TRC via standardised curve).

The EdF execution model itself does not refer to modelling of the possible errors (after diagnosis) in an event tree in the sense of THERP (development of an HRA event tree based on a detailed task analysis). Instead, a simple execution model (consisting of three factors: p_B K_F, p_{NR}) is presented to apply for every critical action. This approach approximately

corresponds to the screening methods outlined in THERP [Swain 83] (Table 20-2) or ASEP [Swain 87] (Table 8-5), see also the comparison between EdF'sPHRA and ASEP presented in [Mosneron 93].

Furthermore (or consequently), EdF's level of action decomposition is more holistic, i.e. less detailed than THERP's. THERP's decomposition of tasks is guided by the model outlined in Figure 5 in this report (Figure 4-2 in [Swain 83]). This requires the identification of task-participating man-machine interfaces (displays, controls, etc.) and elements of cognition (e.g. interpretation of a pattern of signals). However, for post-diagnosis execution, the latter is usually neglected or modelled speculatively, see [Swain 83] (page 12-9) and Section 6.4 in this report. EdF'sPHRA accounts for this variety of PSFs in a simplified manner: choosing between two basic HEPs (p_B), three factors for contextual correction (K_F) and four recovery HEPs (p_{NR}) per critical post-diagnosis action.

In this respect (modelling post-diagnosis execution), THERP should be preferred to the EdF method, because a detailed event tree is an important tool for the identification of

- functional and time correlations between success and/or non-success events, and
- problems with certain man-machine interfaces or elements of cognition.

However, this judgement (preference of THERP in modelling post-diagnosis execution) is preliminary, because:

- execution modelling is closely related to error identification, and for error identification, EdF combines systematic analysis and qualitative simulator data (unfortunately not demonstrated in [EPS 1300], but partially demonstrated in [Mosneron 94]) [Mosneron 95];
- the standard event tree in [EPS 1300] (page 105) may be expanded for certain cases (unfortunately not demonstrated in [EPS 1300]) [Mosneron 96].

The use of qualititative data for error identification is an advanced approach in HRA - especially for the incorporation of unrequired actions. In this context, EdF's approach is a directive development.

For screening, qualitative data can be used exclusively to obtain information about the most likely errors. However, it should be noted that screening in error identification inherently produces incomplete results. Not every error-likely situation (e.g. an adverse combination of hardware failures, see [Woods 90] or [Reer 93]) is covered by the base of information (e.g. simulator data) which is used for screening.

7.2 HCR model

BRIEF DESCRIPTION

The HCR model ([Hannaman 84]; summarising descriptions in: [Hannaman 85] and [Hannaman 88]) is the result of the second phase of a research project sponsored by EPRI (Electric Power Research Institute) for the improvement of the PRA methodology with a view to including operator actions.

The first phase comprised the generation of a framework, which was referred to as "Systematic Human Action Reliability Procedure" [SHARP] and published in 1984. [SHARP] recommends an action-type-dependent use of quantification methods (on human reliability); e.g. THERP event tree ("HRA event tree") for rule-based actions in maintenance activities.

The development of the HCR model is a response to the requirement of methods ascertained in [SHARP] for the quantification of the behaviour of control room personnel in the course of an accident.

The HCR model has so far been used in the following studies [Swain 89]:

- La Salle PRA [Hannaman 86];
- Loviisa PRA [Vuorio 87];
- TMI-1 PRA [Wakefield 87];
- PSA for French 900 MWe reactors [Lanore 87].

Moreover, applications can be found in Scandinavian studies [Hirschberg 90].

The time-reliability correlation is the main error model in the HCR model: the longer the time available for performing a task, the higher is the reliability (success probability) and the lower is the failure probability. Success is defined in the HCR model as the performance of a given task, failure as the "non-response ... similar to non-completion of a predefined task" [Hannaman 88]. The following failure causes are specified as examples:

- "slow in performing";
- "used another method" (for performing the task);
- "was not alerted";
- "required additional information".

The HCR model calculates the failure probability, HEP, for a task with the aid of a three-parameter Weibull distribution Q(t). The following two estimates are needed to determine the three parameters uniquely determining the course of Q(t):

1. Cognitive demand on the personnel by the task broken down into skill-based, rule-based and knowledge-based requirements after [Rasmussen 79].

2. Point estimate $T_{0.5}$ (median) of the average time required for performing the task. If no empirical time data are available for the determination of $T_{0.5}$, then $T_{0.5}$ is to be estimated by expert judgement (supported by plant visits and operator interviews). A value determined for $T_{0.5}$ has still to be modified upwards (at most by 300 %) or downwards (at most by 40 %) using three PSFs (see Figure 11) in order to include the boundary conditions to be expected in the action situation.

The failure probability will then result by substituting the time t available for performing the task in the complementary distribution function, $P(t) = 1 - Q(t)$, of the Weibull distribution determined before.

The HCR model is to be seen in connection with the SHARP method. This method is essentially based on action decomposition, as is THERP. In contrast to THERP, action decomposition in SHARP follows a slightly coarser pattern similar to the OAT method ("Operator Action Tree") [Hall 82], and greater emphasis is placed on modelling time-dependent error probabilities.

Figure 10 illustrates that the HCR model is primarily conceived for the quantification of cognitive processes in accident diagnosis:

> »In particular, the model considers different levels of thinking or cognitive processing associated with the way the crew supervises the course of an accident sequence« [Hannaman 85] (page 343) .

In concrete terms, the HCR model is to furnish a probability for the error designated A3 in Figure 10: no detection or slow cognitive processing relative to the requirements of the event. For the other errors associated with the diagnosis,

- (A2) misdiagnosis: faulty detection/cognitive processing,
- (B4) selection of a non-viable option despite correct identification of the accident-initiating event,
- (B6) selection of a non-viable option in the case of misdiagnosis and
- (C) recovery of a preceding diagnosis or decision-based error fails,

the HCR model refers to other methods and data sources. Errors in connection with the performance of an action (after correct diagnosis and decision) are not dealt with in the model in Figure 10. As stated elsewhere, such errors belong to the "other" aspects (not covered by HCR) of an analysis:

> »The non-response probability« (from the HCR model) »can be incorporated into the logic representation defined in step 4« (representation) »of SHARP (e.g. EOATS, HRA trees) to include other aspects of human reliability such as turning the wrong switch, given that the intent was to turn the correct switch « [Hannaman 85] (page 348).

An error probability calculated with the HCR model is thus a contribution to the failure probability of diagnosis and decision-making after the occurrence of an accident. Guidelines for assessing the uncertainty bounds are given in [Moieni 86].

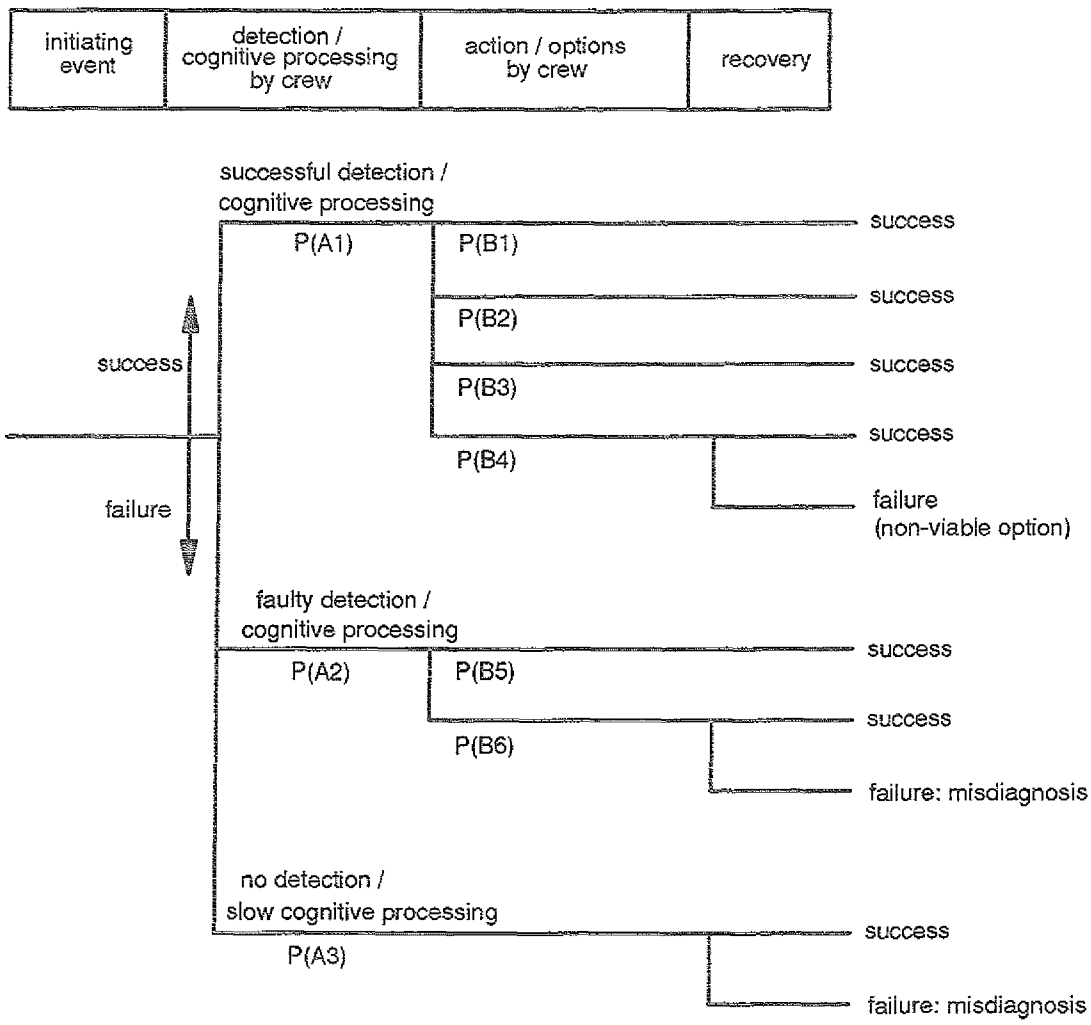


Figure 10: Expanded operator action tree [Hannaman 85] for the incorporation of the HCR model. The HCR model itself (Figure 11) serves to estimate P(A3)

EVALUATION

The HCR model itself is unsuitable for a complete human reliability analysis (HRA) within a PRA. This assessment is trivial because the authors of the HCR model explain on several occasions in the text that HCR only covers a subaspect (in the area of diagnosis) of a HRA. It is to be criticized, however, that this restriction of the application range is not clearly visible

from the publications on the HCR model. In [Hannaman 85] alone several statements can be found (on the application range) which are slightly contradictory:

- Figure 3 on page 347 suggests that the HCR model is conceived for the quantification of diagnosis reliability.
- In contrast, steps 1 and 5 of the survey on page 348 suggest that HCR is to be applied to all situations in which an evaluation of human reliability is required.
- Step 7 of the same survey, however, implies that other models should be used for inadvertent operator errors.
- On the other hand, the task example assessed as HCR on page 350 (manual reactor trip) nevertheless includes manual switching to initiate reactor trip.

All quotations are from the same source [Hannaman 85]!

Scandinavian studies indicate that – on account of the publications available on the HCR model – the application range of the HCR model can be misinterpreted. In [Hirschberg 90] (p. 315), for example, the attempt was made to apply the HCR model to actions after diagnosis; the results obtained, however, were – correctly – assessed as being obviously unrealistic.

Within the scope of the evaluation carried out here it is assumed that the application range conceived for HCR is confined to event A3 of the diagnosis/decision model in Figure 10; this corresponds to a recent explanation [Moieni 94] (p. 33) concerning the HCR model, where this model is valued as a satisfactory error model in the field of diagnosis.

It is basically positive that the HCR frame (Figure 10) distinguishes between the identification (diagnosis) of the system state required for accident control and selection of the action to be performed. This enables e.g. an explicit incorporation of decision-based errors (despite correct diagnosis) due to conflicts between action alternatives; cf. [Reer 93]. In comparison, the diagnosis in THERP comprises interpretation and decision-making. The HCR model itself (Figure 11), however, does not furnish any probabilities for the decision-based errors B4 and B6 in Figure 10.

Another basically positive feature is the fact that the event category NO CORRECT DIAGNOSIS is divided into the subcategories MISDIAGNOSIS (A2) and NO OR LATE DIAGNOSIS (A3). This allows e.g. the incorporation of unrequired operator actions [Reer 93] which have a damage-aggravating effect due to incorrect diagnosis. However, the HCR model does not furnish any probabilities in this case either.

In connection with misdiagnosis, the authors discovered a discrepancy in the "frame model" (Figure 10) belonging to the HCR model. According to the statements in [Hannaman 85] (p. 434) the following applies to the sum of probabilities for events associated with diagnosis (A1: correct diagnosis; A2: misdiagnosis; A3: late or no diagnosis):

$$P(A1) + P(A2) + P(A3) = 1,$$

where P(A3) is determined from the time-reliability correlations (the actual HCR model) in Figure 11. Consequently, the event $\overline{A3}$, $P(\overline{A3}) = 1 - P(A3)$, which is complementary to A3, would not necessarily be a success event, because it would also comprise the failure event A2 (misdiagnosis) in addition to A1. This would mean that, in addition to times until correct diagnosis, times until misdiagnosis have also been used without further differentiation to determine the time-reliability correlations. This appears illogical because a correlation favourable for reliability would then result if many misdiagnoses occur within a short time. Moreover, there is a contradiction to the time defined in [Moieni 94] (p. 33) for the determination of the HCR correlations: time between accident occurrence and first correct action. Consequently, the following holds for an error probability derived from the HCR model:

$$P(A3) = 1 - P(A1).$$

The discrepancy shown here is indicative of a basic problem in the use of time data for deriving error probabilities. The concrete causes for the more or less great deviation of a particular time required from the average time required remain hidden from the analyst who does not know the exact boundary conditions of time data acquisition.

A further basic weakness of time-dependent models (such as HCR) is the neglect of the dependence between available time t and required time T. The time perceived as available by the crew could influence the time of performance in such a way that the crew works particularly slowly or particularly fast. Comparable compensation effects can be corroborated by ergonomic findings [Hacker 86].

Because the HCR model only needs a few parameters, the method gives only a coarse picture of the reality in the plant if no detailed task analysis is performed. In this case, the HCR model does not provide enough face validity ([Fujita 92], [Swain 89], [ACSNI 91]). Moreover, the PSFs are insufficient and modelled coarsely. HCR provides only three different PSFs with a maximum of five subdivisions and the underlying situation cannot be considered adequately (see [Swain 89] and [Fujita 92]).

Another problem of the HCR model is the normalized time reliability curve, which only shows the relationship of available time and needed time. The absolute magnitude of the time reserve (e.g., hours, minutes or seconds) is not considered. Therefore, the time reliability correlation (TRC) of the model is not applicable to situations which are not covered by the underlying data of the TRC [Sträter 94b].

Furthermore, the time reliability curves for skill-based, rule-based and knowledge-based behaviour are not sufficiently appropriate for the real levels of cognitive behaviour, because the knowledge levels are more a continuum than a dichotomy (see [Wickens 84]). This problem becomes worse due to the fact that the classification into skill-based, rule-based and knowledge-based behaviour completely depends on the experience of the persons in certain

knowledge domains and may vary between tasks (see [Bubb 92], page 106). Summarising this point, the HCR model is a strongly simplifying classification of human information processing from the point of view of human engineering and cognitive psychology.

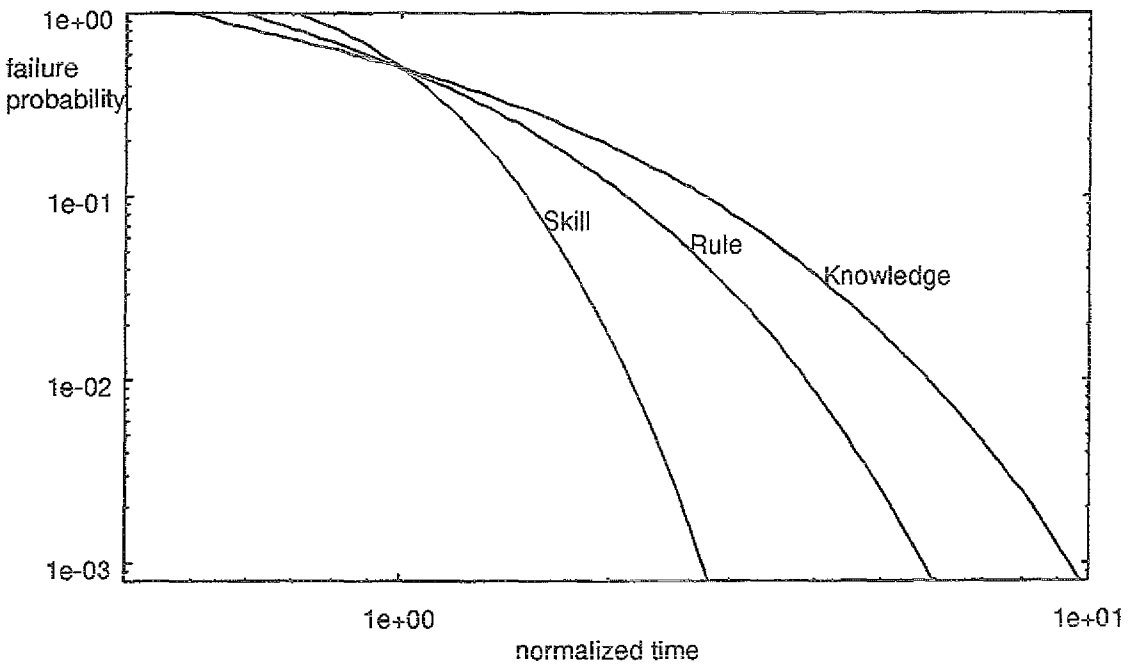


Figure 11: Dependence of the error probability p on the quotient of available time (t) and required average time ($T_{0.5}$), as quantified in the HRC model [Hannaman 84] for the performance of a task.
The error is designated "non-response" in the HCR model and defined as non-completion of a predefined task. The average time required, $T_{0.5}$ (median time), results from a nominal estimate ($T_{0.5/nominal}$) still to be modified by three PSFs:
$T_{0.5} = T_{0.5/nominal} (1 + K_1) (1 + K_2) (1 + K_3)$
where:
K_1 = coefficient for operator experience, K_1=0.22 (expert, well-trained), K_1=0 (average knowledge training), or K_1=0.44 (novice, minimum training);
K_2 = coefficient for stress level, K_2=0.44 (situation of grave emergency), K_2=0.28 (situation of potential emergency), K_2=0 (active, no emergency), or K_2=0.28 (low activity, low vigilance);
K_3 = coefficient for quality of operator/plant interface, K_3=-0.22 (excellent), K_3=0 (good), K_3=0.44 (fair), K_3=0.78 (poor), or K_3=0.92 (extremely poor).
The time-dependent error probability is:
$p(t) = \exp\{-[(t/T_{0.5} - a)/c]^b\}$
where:
(a, b, c) = behaviour-type-specific coefficients,
(0.7, 1.2, 0.407) for skill-based behaviour,
(0.6, 0.9, 0.601) for rule-based behaviour, or
(0.5, 0.8, 0.791) for knowledge-based behaviour.

The results from the ORE program [Moieni 94] have led to modified versions of the HCR frame model (Figure 10) and the actual HCR model (Figure 11) (more details in Section 7.3). The application of the HCR model version of [Hannaman 84] is therefore no longer recommended by EPRI.

Despite all the above criticism, the authors consider the HCR model to be a milestone in the quantification of time-dependent error probabilities. The methodological principle is a substantial aid for an analyst facing the problem of estimating the distribution function of a time required for performing a task. In order to comply with the multitude of possible performance situations, it is meaningful to assume a multiparameter distribution function, for example a three-parameter Weibull distribution with the parameters T_0, η and β. If no time data are available, the analyst faces the difficult task of estimating the three parameters; especially due to problems with the practical (relative to the performance situation) interpretability of η (scale parameter) and β (form parameter). In the HCR model it is sufficient to classify the type of action (with respect to its cognitive demands on the operator) and to estimate the time requirement to be expected on average. This facilitates the assessment considerably because cognitive demands and average times can be determined on the basis of a task analysis (operator interviews, plant visit, walk-through).

The methodological estimation principle of the HCR model (for the determination of distribution functions of action times) is therefore an essential contribution towards reducing the data problem in human reliability analysis. In [Reer 94b] the HCR principle was used to extend THERP so as to also model the times of actions after diagnosis as probabilistically distributed variables.

7.3 HCR/ORE method

Brief Description

The following brief description of the HCR/ORE method is based on the presentations in [Moieni 94] and [Parry 91]. Further details on the method and data base can be found in [Spurgin 90] and [Parry 92]. The data themselves are the property of EPRI and not readily available to the public; cf. [Moieni 94] (p. 39).

The method described in [Moieni 94] is not explicitly given a concrete name. The designation HCR/ORE was selected here because the use of data from the operator reliability experiments (ORE) is recommended for the determination of error probabilities and because it is a further development of the HCR model [Hannaman 84].

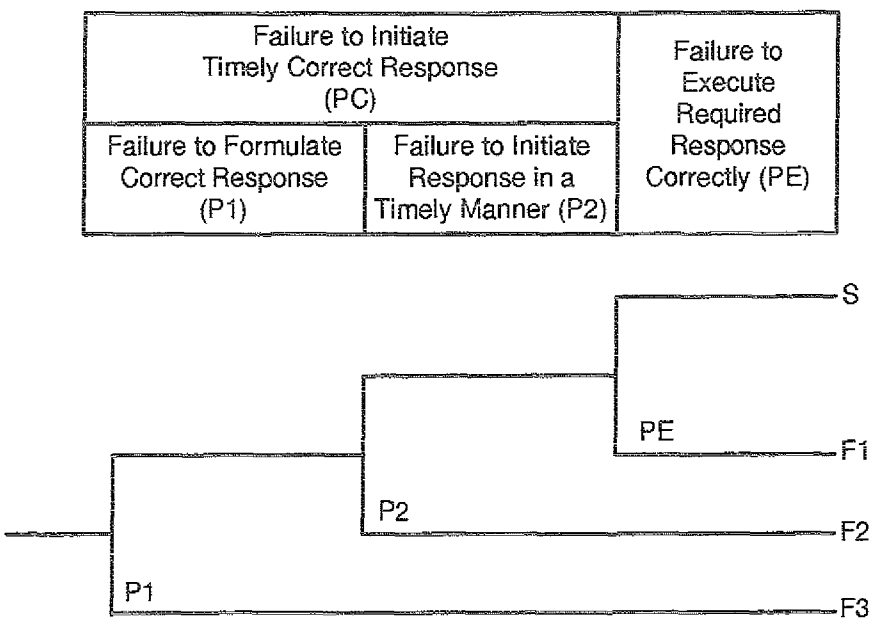
The HCR/ORE method can be used for the probabilistic quantification of actions in an accident. It is possible to quantify actions with predefined procedure ("interactions that are

dictated by procedures") as well as actions without predefined procedure ("scenario-specific recovery of failed equipment or realignment of systems") ([Moieni 94], p. 35); see also Chapter 2.2.3 in this report.

The error model for actions with predefined procedure is outlined in Figure 12. A distinction is made between diagnosis and subsequent actions. THERP [Swain 83] is referred to for the quantification of actions after diagnosis ([Moieni 94], p. 41).

Two main types of error are distinguished as contributions to diagnosis failure:

1. failure to formulate a correct response
2. failure to initiate a correct response in a timely manner (too late) on condition that a correct response was found.



S: Success

F_1: Failure: non-recovered slips in execution

F_2: Failure: non-response in given time window

F_3: Failure: non-recovered intention formation errors

Figure 12: Generalized event tree for modelling procedure-driven operator actions in an accident ([Moieni 94], Fig. 4)

Eight error mechanisms are identified ([Moieni 94], p. 40) for the first error type (failure to formulate correct response):

(a) problems due to available information;

(b) failure of attention;

(c) misreading/miscommunication;

(d) misleading information;

(e) skipping a procedure step;

(f) misinterpretation of instruction in procedure;

(g) misinterpretation of decision logic in procedure;

(h) deliberate deviation from the procedure (violation).

For each mechanism, an error probability is to be determined with the aid of structured expert estimates. The decision tree method outlined in Figure 13 is recommended for this purpose. Uncertainty is not explicitely addressed; since the HEPs obtained from the decision trees are based on THERP, they can be treated as median values of lognormal distributions with error factors ranging from three to ten [Moieni 94] (page 40). Furthermore, recovery mechanisms are to be considered for each error mechanism, see [Moieni 94] (page 41) for details.

The second error type (late diagnosis) in Figure 12 is to be quantified with the aid of time-reliability correlations from the ORE program ([Moieni 94], p. 38). The ORE data were processed according to the same methodological principle as the data of the HCR model – as complementary distribution functions of normalized times, normalizing being effected by division by the respective average time requirement $T_{0.5}$ (median). The ORE diagnosis times were approximated with the aid of two-parameter lognormal distributions:

with

$$p(t) = pr(T > t) = 1 - \Phi\left(\frac{\ln(t / T_{0.5})}{\sigma}\right)$$

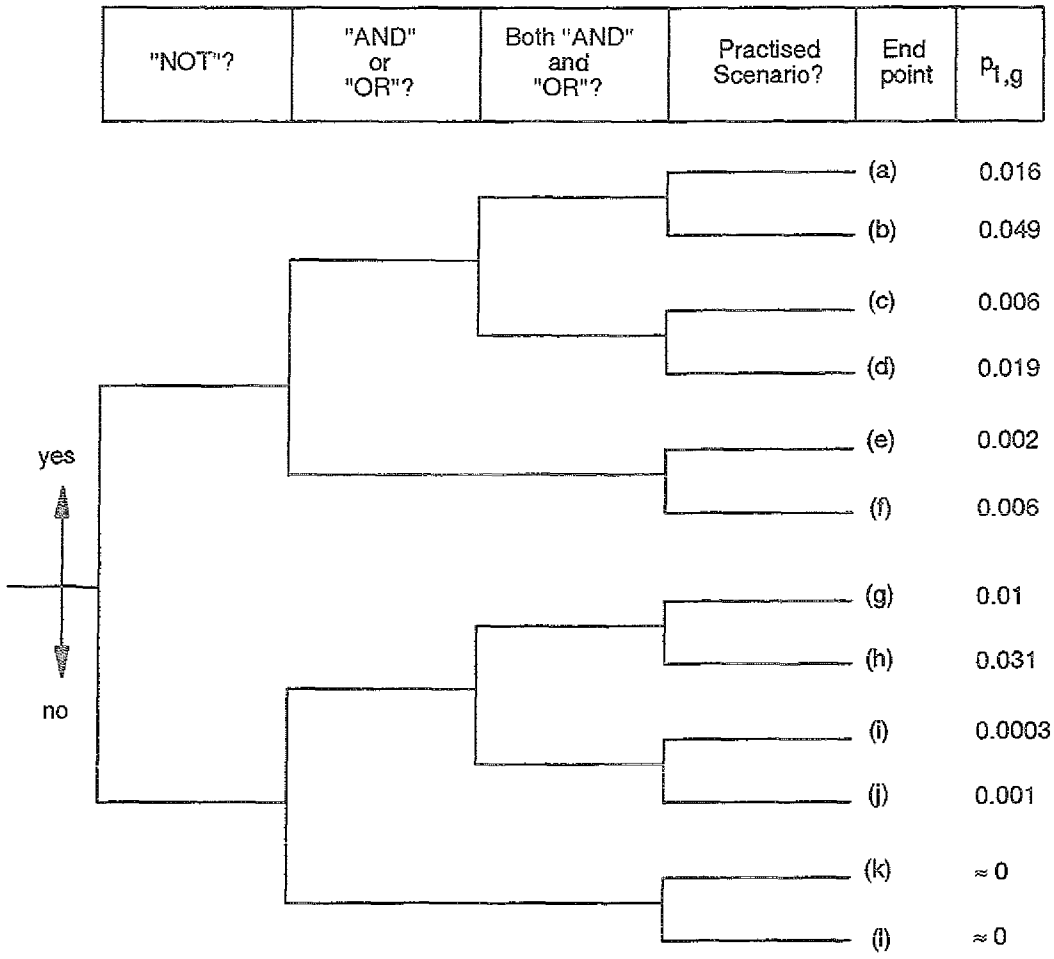
T = diagnosis time (random variable);

t = available diagnosis time

σ = standard deviation from (1n T);

$\Phi(..)$ = distribution function of the standard normal distribution

This involves a total of 1,100 diagnosis times recorded under the ORE program ([Moieni 94], p. 39). In attempting to group these data with respect to typical σ-values, the HCR grouping according to cognitive levels (skill/rule/knowledge) proved unsuitable. Instead, six typical courses were identified using other criteria. First of all, two reactor-type-specific main types were identified, one for PWR accidents and one for BWR accidents. Within each reactor type group three time profiles are distinguished depending on the accident dynamics ([Moieni 94], page 33).



"NOT": Does the procedure step contain the word "not"?

"AND" or "OR": Does the procedure step present diagnostic logic in which more than one condition is combined to determine the outcome?

Both "AND" and "OR": Does the procedure step contain a complex logic involving a combination of ANDed and ORed terms?

"Practised Scenario": Has the crew practised executing the step in a scenario similar to this one in a simulator?

$p_{1,g}$ is a contribution to failure in decision-making; see Figure 12, P1

The probabilities are based on data from THERP

Figure 13: Decision tree for determining the probability $p_{1,g}$ that a diagnostic logic in a procedure is misinterpreted ([Moieni 94], Figure 7)

Table 28 shows the six different types of accident diagnosis. The standard deviations (σ-values) of the logarithmized diagnosis times were determined here (in compiling Table 28) by the following equation

$$\sigma = \frac{\ln(K_{0.95})}{1.645}$$

where the quotient, $K_{0.95} = T_{0.95}/T_{0.5}$, of 95% percentile ($T_{0.95}$) and 50% percentile ($T_{0.5}$) was read from the time profiles printed in [Moieni 94] (p. 38). Due to the readings performed here, the σ-values listed in Table 28 are not particularly precise. The precise σ-values are the property of EPRI.

Required response dictated by the dynamics of the accident	Reactor type	σ
(1) Response following a change in the plant state that is indicated by an alarm or value of a monitored parameter (e.g. response to a spurious pressurizer spray operation in a PWR).	PWR	≈0.55
	BWR	≈0.73
(2) Response following an event that gives rise to a primary cue (as in case 1 above) that has to be achieved when a parameter is exceeded or can be seen not to be maintainable below a certain value (e.g. initiate residual heat removal when the suppression pool (SP) temperature exceeds 35 °C in a BWR). This human interaction involves a waiting period after the primary cue in order to reach a determined plant state.	PWR	≈0.36
	BWR	≈0.61
(3) Response following an event that gives rise to a primary cue (as in case 1 above) that has to be achieved before some plant parameter reaches a critical value (e.g. initiate standby liquid control system before SP temperature reaches 43 °C in a BWR). This critical value can be regarded as a soft prompt, or secondary cue.	PWR	≈0.76
	BWR	≈0.79

Table 28: Standard deviation (σ) of the logarithmic diagnosis time (ln (t)) as a function of response and reactor type. After [Moieni 94] (Figures 3 and 6). Except for PWR response type 3, the curves are plotted together with their uncertainty bounds.

For the quantification of actions without predefined procedure (recovery actions), i.e. repair actions in the widest sense, the HCR/ORE method roughly distinguishes between

- identification of the recovery action (by the crew)

  and its

- execution,

where it may be necessary to subdivide the execution even more precisely. If there are empirical time data available for these subtasks, then the recovery action is to be modelled as a stochastic process, and the result obtained is a time-dependent error probability ([Moieni 94], page 45). Essentially, this corresponds to the approach proposed in [Reer 93] (Section 2.4.1) and [Reer 94b].

If no empirical time data are available, structured estimates are to be used for quantification and the user is offered two decision trees as a tool. The first tree serves to quantify the error probability p_I in identifying (i.e. diagnosis by the crew) the recovery action, it comprises 6 PSFs; the second tree comprises 4 PSFs and serves to assess the probability p_E for failure to execute the recovery action ([Moieni 94], pages 47-49):

(I) PSF assessments needed for a simplified estimation of p_I:

1. Is the recovery action identified in emergency operating or other procedures?
2. Are there direct indications in the main control room?
3. Is the time available for diagnosis long, intermediate or short?
4. Are the crew familiar with or trained on the recovery process?
5. If reasoning is required, is it deductive or inductive?
6. Is detection by auxiliary operators possible?

(E) PSF assessments needed for a simplified estimation of p_E:

1. Is the time for execution long, intermediate or short?
2. Are the crews trained (or practised) in the execution of the recovery action and/or is there a procedure?
3. Is the recovery action simple or complex?
4. Are the environmental conditions around the equipment hostile?

Finally, the non-recovery probability is

$$p_{NR} = p_I + p_E - p_I p_E$$

ROUGH EVALUATION

The HCR/ORE method can also be valued – with reservations – as a positive new development. This evaluation is subject to reservation because the method is relatively recent and the data (being the property of EPRI) are not readily available to the public.

The division (Figure 12) of diagnosis failure into the events (1) failure to formulate response and (2) response too late is, in principle, to be regarded as a positive new development in relation to THERP where both events are considered in one model. A separate consideration of the first event (failure to formulate response), however, is required for reasons of plausibility alone, because the lognormal distribution model (HCR/ORE) would lead to unrealistically small error probabilities, compared to the THERP diagnosis model, for long available times.

Furthermore, it should be noted that the two events cannot always be clearly differentiated. The error mechanisms (e.g. misreading) specified in [Moieni 94] for the first event (failure to formulate correct response) could, in principle, also contribute only towards a delay in diagnosis without causing complete failure. Such error mechanisms would then be covered, in principle, by the time-reliability correlations and would not have to be specifically modelled. In a strict sense, only those error mechanisms should be considered for quantifying the first event (failure to formulate correct response) in Figure 12 which are irreversible, i.e. ultimately lead to failure of the requested action. It is therefore recommended in further investigations to discover the basic causes for the ultimate failure of an action, for example, early intervention into the system resulting in a failure of the hardware required for action.

It cannot therefore be stated right from the beginning whether the diagnosis modelling proposed in the HCR/ORE model is more realistic than the THERP diagnosis model.

For the quantification of the second event (response too late) in Figure 12 the HCR/ORE method recommends the use of time-reliability correlations defined on a normalized time scale ($t/t_{0.5}$). It is criticized in [Sträter 94b] (page 8) that the absolute time reserve is not taken into account. This criticism is particularly true if a task is investigated whose average time requirement ($t_{0.5}$) is of an order of magnitude not covered by the HCR/ORE data base. Let us assume that the $t_{0.5}$ values of the HCR/ORE data are times in the minute range. Further investigations are then required in order to clarify whether and how tasks with a time requirement in the range of seconds or hours can be quantified with these data.

The recommendations of the HCR/ORE method for the quantification of actions without predefined procedure (recovery actions) are, in principle, to be judged as positive new developments. However, their practical application still requires a clarification of further methodological details, in particular with respect to the quantification of

- dependences between the failure of an action *with* predefined procedure and the failure of an action *without* predefined procedure and
- conflicting decisions if the crew has to choose between an action with and without predefined procedure.

HCR/ORE's simplified method ($p_{NR}=p_I+p_E-p_Ip_E$) for quantifying recovery actions provides a feasible screening approach to emergencies during shutdown states of NPPs. The respected criterion (not dictated by procedures) seems to be typical of actions which are required to control emergency situations during plant shutdown, because:

- the identification (by the crew) of a correct action depends on a variety of conditions, some of which are difficult to predict in HRA, e.g. the set of systems taken out of operation for maintenance until emergency, and
- there is often a variety of correct actions which could be chosen by the crew.

HCR/ORE offers a coarse approach to deal with such inponderabilities in HRA. However, this approach is only an interim solution because some of the required PSF assessments (e.g. E3) need more detailed information about the action, and, as outlined above, actions in shutdown emergencies are difficult to predict (or identify) in HRA.

7.4 Summary evaluation

The TRC approach is usually applied for analysing the diagnosis of an abnormal event. In this context, a TRC-derived HEP is the probability pr(T>t) that the required time (T) for diagnosis is greater than the available time (t) for diagnosis. Table 29 summarises some TRCs of the methods THERP, EdF'sPHRA, HCR and HCR/ORE.

HEP	**0.5**	**0.1**	**0.01**	**0.001**	**0.0001**	**0.00001**
THERP, nominal model	2 min = $T_{0.5}$	10 min = 5 $T_{0.5}$	20 min = 10 $T_{0.5}$	30 min = 15 $T_{0.5}$	60 min = 30 $T_{0.5}$	1500 min = 750 $T_{0.5}$
THERP, initial-screening model	10 min = $T_{0.5}$	20 min = 2 $T_{0.5}$	30 min = 3 $T_{0.5}$	60 min = 6 $T_{0.5}$	1500 min = 150 $T_{0.5}$	*≈ 3000 min* (extrapolation, not tabled in [Swain 83])
EdF, curve 1	≈ 3 min = $T_{0.5}$	≈ 9 min = 3 $T_{0.5}$	≈ 21 min = 7 $T_{0.5}$	For t > 30 min both curves remain at a constant residual HEP of 0.005		
EdF, curve 1'	≈ 9 min = $T_{0.5}$	≈ 12 min = 1.33 $T_{0.5}$	≈ 21 min = 2.33 $T_{0.5}$			
EdF, curve 2	12 min = $T_{0.5}$	20 min = 1.67 $T_{0.5}$	30 min = 2.5 $T_{0.5}$	For t > 50 min (approximately) the curve remains at a constant residual HEP of 0.005		
EdF, curve 3	22 min = $T_{0.5}$	30 min = 1.36 $T_{0.5}$	For t > 300 min (approximately) the curve remains at a constant residual HEP of 0.03			
HCR, knowledge	$T_{0.5}$	2.74 $T_{0.5}$	5.84 $T_{0.5}$	9.34 $T_{0.5}$	13.19 $T_{0.5}$	17.27 $T_{0.5}$
HCR, rule	$T_{0.5}$	2.12 $T_{0.5}$	3.88 $T_{0.5}$	5.75 $T_{0.5}$	7.68 $T_{0.5}$	9.68 $T_{0.5}$
HCR, skill	$T_{0.5}$	1.52 $T_{0.5}$	2.15 $T_{0.5}$	2.74 $T_{0.5}$	3.29 $T_{0.5}$	3.82 $T_{0.5}$
HCR/ORE, response type 1	$T_{0.5}$	PWR: ≈ 2 $T_{0.5}$ BWR: ≈ 2.55 $T_{0.5}$	PWR: ≈ 3.6 $T_{0.5}$ BWR: ≈ 5.5 $T_{0.5}$	PWR: ≈ 5.5 $T_{0.5}$ BWR: ≈ 9.6 $T_{0.5}$	PWR: ≈ 8.1 $T_{0.5}$ BWR: ≈ 16 $T_{0.5}$	PWR: ≈ 9 $T_{0.5}$ BWR: ≈ 18.5 $T_{0.5}$
HCR/ORE, response type 2	$T_{0.5}$	PWR: ≈1.6 $T_{0.5}$ BWR: ≈ 2.2 $T_{0.5}$	PWR: ≈ 2.3 $T_{0.5}$ BWR: ≈ 4.1 $T_{0.5}$	PWR: ≈ 3.05 $T_{0.5}$ BWR: ≈ 6.6 $T_{0.5}$	PWR: ≈ 3.9 $T_{0.5}$ BWR: ≈ 10.2 $T_{0.5}$	PWR: ≈ 4.2 $T_{0.5}$ BWR: ≈ 11.5 $T_{0.5}$
HCR/ORE, response type 3	$T_{0.5}$	PWR: ≈ 2.65 $T_{0.5}$ BWR: ≈ 2.75 $T_{0.5}$	PWR: ≈ 5.9 $T_{0.5}$ BWR: ≈ 6.3 $T_{0.5}$	PWR: ≈ 10.55 $T_{0.5}$ BWR: ≈ 11.6 $T_{0.5}$	PWR: ≈ 18 $T_{0.5}$ BWR: ≈ 20.1 $T_{0.5}$	PWR: ≈ 20.9 $T_{0.5}$ BWR: ≈ 23.6 $T_{0.5}$

Table 29: Comparative evaluation of HRA methods according to percentiles of the time required for the diagnosis of an abnormal event. The HCR/ORE percentiles are based on imprecise readings from [Moieni 94] (Figure 6), the response types are defined in Table 28 of this report.

The percentiles (listed in Table 29) must be seen in the light of additional features of the methods. One feature concerns the limitation of the TRC approach for the extreme case that the available time (t) is very large. EdF'sPHRA covers such a case by a so-called "residual"

HEP which remains constant beyond a certain time limit. HCR considers "faulty detection or cognitive processing" as an additional (not covered by TRC), time-independent diagnosis error. In HCR/ORE such time-independent error is denoted as "failure to formulate correct response" (see Figure 12 in this report), and has to be quantified separately. THERP gives no explicit time-independent or residual HEP for diagnosis. However, de facto, the HEP of 10^{-4} at t = 60 min (THERP, nominal model) could be interpreted as residual, because beyond 60 min the HEP decreases at a negligible rate of one order of magnitude per day. Nevertheless, compared to $5 \cdot 10^{-3}$ (EdF, curve 1 or 1'), 10^{-4} is a very optimistic residual HEP.

Furthermore, Table 29 shows that, in contrast to TRC models for an absolute time scale (THERP, EdF'sPHRA), normalised time scale models need a direct numerical estimation of the median diagnosis time ($T_{0.5}$). On the one hand, this is an advantage because such an option for estimation results in high flexibility. However, on the other hand, such an option produces two elements of uncertainty:

1. the estimation itself, especially if no timing data are available;
2. inadequate consideration of the absolute time reserve ($t - T_{0.5}$).

In HRA practice, another problem of HCR is that the behaviour levels (knowledge, rule, skill) do not sufficiently correspond to real levels of cognitive behaviour (see Section 2.3.2). However, due to this point, the ORE program resulted in HCR/ORE - a modified version of HCR.

8 New developments with emphasis on expert estimates structured according to performance shaping factors

The lack of situation-specific data on human reliability leads to a situation where the analyst frequently has to estimate a particular error probability (p). There are a number of methods for estimating p not directly, but indirectly via performance shaping factors (PSFs). The advantage of these methods is an increased transparency of a subjective estimate. An essential disadvantage is the additional uncertainty of the mathematical model used for the PSF-dependent calculation of p. Even 'demathematized' models (such as the decision tree in Figure 13) exhibit this uncertainty since the mathematical operations have already been performed in advance. The different mathematical principles for the PSF-dependent derivation of an error probability p will be briefly discussed here using the example of three methods: SLIM [Embrey 84], HEART [Williams 88] and INTENT [Gertmann 92].

8.1 SLIM

BRIEF DESCRIPTION

The SLIM method was published in its preliminary version in [Embrey 83] and in its final version in [Embrey 84]. It was used, for example, in the following studies:

- Peach Bottom [Kolaczkowski 86];
- Kuosheng [Yeh 86];
- Mühleberg [MUSA 92].

The SLIM method begins with the identification of relevant performance shaping factors which are of significance for the failure or success of an operator task to be analysed; for example, for the task of diagnosing the system state in the event of a transient and intervening if a pump of the emergency feedwater system fails ([Embrey 84], Vol. I, p. 5):

- quality of information available to the operators in the control room (PSF 1);
- quality of procedures (PSF 2);
- time available for diagnosis and performance of an action (PSF 3);
- training level of the operators (PSF 4).

The weighted sum of these PSFs then gives a reliability index SLI ("success likelihood index") defined in the interval (0;1):

$$SLI = \sum_{i=1}^{n} (w_i x_i),$$

with

$$\sum_{i=1}^{n} w_i = 1$$

where x_i is the degree of the presence (with respect to a reliability-increasing effect) of PSF i in the action situation investigated, w_i is the relative significance (weighting factor) of PSF i and n is the number of relevant PSFs.

The values for w_i and x_i also defined in the interval (0;1) are to be determined by expert judgement. In this connection, the method description repeatedly refers to the estimation techniques known from multiattribute utility theory (MAUT); e.g. [Embrey 84] (Vol. I, page 5): »simple multiattribute rating technique (SMART)« [Edwards 77].

There are also applications of SLIM in which a failure index (FLI) instead of a reliability index (SLI) is assessed; e.g. in the Mühleberg PSA. In [Reer 93] (Section 5.1.3, English version in [Reer 94c]) a method for the quantification of decision-based errors was developed on the basis of the FLI variant of SLIM.

The following loglinear model equation is recommended in [Embrey 84] (Vol. I, page 7) for the SLI-dependent calculation of the success probability q and the failure probability p of a task examined:

$$\log(q) = \log(1-p) = a \cdot SLI + b$$

where at least two reference values are required for q and p (with associated SLI estimates) to determine the model parameters a and b. If more than two reference values are available, then a and b are to be determined by regression.

In other passages of the method description, the logarithmized error probability p is also modelled as a linear function of the reliability index SLI; cf. [Embrey 84] (Vol. I, page 14) or [Embrey 83] (page 15):

$$\log(p) = a \cdot SLI + b$$

Guidelines for uncertainty bounds estimation are given in [Embrey 84] (Vol. I, Section 1.10.6).

ROUGH EVALUATION

The use of SLIM can only be recommended up to a point. SLIM should only be applied to subtasks or errors identified in decomposing an action and for which no situation-specific error probabilities are available.

In principle, SLIM requires a carefully harmonized procedure in defining the error or task, in defining the single PSFs and in selecting suitable reference probabilities. This approach is the more problematic, the rougher is the definition of failure and the more extensive is the task. If a failure subsumes several failure types and if a task is composed of several subtasks, difficulties must be expected in the expert estimates for the PSFs, because the estimation problem is no longer uniquely definable.

For example, the following problem could arise for the SLIM user if a task to be analysed consists of the subtasks A and B: does a PSF to be evaluated relate to subtask A or B or to both?

Another problem arises if the SLIM model contains a PSF of decisive effect. The available time is such a PSF. Let us assume that there is not enough time available for practically completing the task in time; a THERP quantification would lead to a failure probability of 1.0 in this case. This logically correct consistency is not given for the SLIM model because the maximum effect of the PSF 'time' is limited by its weighting factor. The calculations performed in Table 30 illustrate this weakness of SLIM.

Task [Embrey 84] (Vol. I, Section 3.1)	Diagnosis of a transient and correct response to failure of an emergency feedwater pump	
PSFs	1) quality of information available in the control room 2) quality of procedures 3) time available for diagnosis and response 4) training level	
weighting factors (for the 4 PSFs)	$\underline{w} = (0.50;0.25;0.15;0.10)$	
PSF ratings: 0 = as unfavourable as credibly possible 1 = as favourable as credibly possible	$\underline{x} = (0.7;0.2;0.1;0.5)$ [Embrey 84] (Vol. I, page 6)	$\underline{x} = (0.7;0.2;0;0.5)$ (changed rating for PSF 3)
success likelihood index	SLI = 0.5·0.7+0.25·0.2+ 0.15·0.1+0.1·0.5 = 0.465	SLI = 0.5·0.7+0.25·0.2+ 0.15·0+0.1·0.5 = 0.450
model (decimal logarithm)	$\log(1-p) = 0.00655 \cdot SLI - 0.00567$	
failure probability	p = 0.0060	p = 0.0063

Table 30: Example calculation for the sensitivity of a SLIM result when changing the available time

Moreover, a special feature to be criticized in SLIM is the fact that the mathematical model is not uniformly represented: sometimes the logarithmized success probability (q-variant) is the

target parameter and sometimes the logarithmized error probability (p-variant). Table 31 shows that these two model variants are not equivalent to each other. For the same initial assumptions, the analysis example with the q-variant gives an error probability higher by a factor of 1.6 than with the p-variant. The methodology descriptions available for SLIM are thus inconsistent.

Task [Embrey 84] (Vol. I, Section 3.1)	Diagnosis of a transient and correct response to failure of an emergency feedwater pump	
success likelihood index	$SLI = 0.465$	
reference error probabilities with associated SLI values	$p_1 = 0.001$, $SLI_1 = 0.8$ $p_2 = 0.01$, $SLI_2 = 0.2$	
model variant	$\log(q) = \log(1-p)$ $= a \cdot SLI + b$ [Embrey 84] (Vol. I, p. 6) (q-variant of SLIM)	$\log(p) = a \cdot SLI + b$ [Embrey 83] (p. 15) (p-variant of SLIM)
model parameters	$a = \frac{\log\left(\frac{1-p_1}{1-p_2}\right)}{SLI_1 - SLI_2}$ $= 0.00655$ $b = \log(1-p_1) - a \cdot SLI_1$ $= -0.00567$	$a = \frac{\log\left(\frac{p_1}{p_2}\right)}{SLI_1 - SLI_2}$ $= -1.67$ $b = \log(p_1) - a \cdot SLI_1$ $= -1.66$
failure probability of the task (with SLI=0.465)	$p = 1 - 10^{a \cdot SLI + b}$ $= 0.0060$	$p = 10^{a \cdot SLI + b}$ $= 0.0037$

(Based on [Kosmowski 94b])

Table 31: The linear dependence of the logarithmic success probability (log (q)) is not equivalent to the linear dependence of the logarithmic error probability (log (p))

This contradiction (Table 31) has a deeper cause which was already dealt with in Sections 6.1 and 6.8. It was shown in [Reer 93] (Section 5.2.3) that the p-variant of the SLIM model is equivalent to a multiplicative model

$$p = p_0 \cdot PSF_1 \cdot PSF_2 \cdots$$

in which the wanted error probability p can be represented as the product of a nominal value (p_0) and one factor or several factors. As already discussed in Chapter 6.1, such a model is only approximately mathematically correct because a probability in the strict sense is not a multiplicative quantity. It is mathematically correct if the odds ratio, $y = (1-p)/p$, or the error

chance ratio (1/y) is modelled as a multiplicative target quantity. Table 32 shows that a consistent model is thus obtained. The same result is obtained with the logarithmized odds ratio as with the logarithmized error chance ratio.

Task [Embrey 84] (Vol. I, Section 3.1)	Diagnosis of a transient and correct response to failure of an emergency feedwater pump	
success likelihood index	$SLI = 0.465$	
reference error probabilities with associated SLI values	$p_1 = 0.001, SLI_1 = 0.8$ $p_2 = 0.01, SLI_2 = 0.2$	
reference odds ratios	$y_1 = (1-p_1)/p_1 = 999$ $y_2 = (1-p_2)/p_2 = 99$	
model variant	$\log\left(\frac{1}{y}\right) = a \cdot SLI + b$	$\log(y) = a \cdot SLI + b$
model parameters	$a = \frac{\log\left(\frac{1/y_1}{1/y_2}\right)}{SLI_1 - SLI_2}$ $= -1.67$ $b = \log\left(\frac{1}{y_1}\right) - a \cdot SLI_1$ $= -1.66$	$a = \frac{\log\left(\frac{y_1}{y_2}\right)}{SLI_1 - SLI_2}$ $= 1.67$ $b = \log(y_1) - a \cdot SLI_1$ $= 1.66$
odds ratio and failure probability of the task (with SLI=0.465)	$y = 1/10^{a \cdot SLI + b}$ $= 273$ $p = 1/(1+y)$ $= 0.0037$	$y = 10^{a \cdot SLI + b}$ $= 273$ $p = 1/(1+y)$ $= 0.0037$

Table 32: The linear dependence of the logarithmized error chance ratio is equivalent to the linear dependence of the logarithmized odds ratio

Like every other method that is using expert judgement, SLIM is exposed to additional uncertainties in the data due to the subjective judgement of the expert. Therefore, the danger of sparse consistency in the results is high. This effect may be intensified by the fact that the qualification of the experts performing the judgement is not defined (see [Berg 92], [Swain 89], [Reichart 85]).

An additional problem arises from the calibration procedure, which only needs two so-called anchor tasks. Firstly, the theoretical basis of the calibration is not validated ([Swain 89], [Bubb 92]). Secondly, uncertainties in the parameter estimation of the SLI might be high due

to the fact that no aids are given for the determination of the error probabilities of the anchor tasks or for the determination of the similarity of the two tasks which are used as anchors (see [Swain 89]). If only two tasks are taken, the variation of the calibration data cannot be calculated.

8.2 HEART

Brief Description

The HEART method [Williams 88] is like SLIM an approach based on the selection of PSFs of significance for the success or failure of an operator task investigated. It was utilised at the task-based level (e.g. 'fail to achieve a certain task'), rather than at the elemental level (e.g. 'fail to close valve') as THERP may be utilised [Kirwan 95]. The HEART method was used in various British studies [Kirwan 94].

The user of this method is offered point values and uncertainty intervals of nominal error probabilities (NHEPs, in HEART denoted as "nominal human unreliabilities") for nine tasks described very generally. The point values are between 0.55 (task: "totally unfamiliar, performed at speed with no idea of the likely consequences") and 0.00002 (task: "respond correctly to system command even when there is an augmented supervisory system providing accurate interpretation of systems state"); the data are based on literature studies. From among these tasks one task should be selected which is applicable to the task to be analysed. HEART offers a list of 38 factors for the situation-specific modification (by multiplication) of the corresponding error probability. Each factor is described as an error producing condition (EPC). The numerical values specified for the factors range between $K = 17$ ("unfamiliarity with a situation which is important but which only occurs infrequently or which is novel") and $K = 1.02$ ("age of personnel performing perceptual tasks"). Most of these data are also based on literature studies, but literature references are missing for some factors.

The HEART user is now requested to select those EPCs which he considers relevant for the success or failure of the task investigated. For each selected EPC a value x (rating) defined in the interval (0;1) is to be estimated which describes the degree of the presence of the EPC. The multiplicative effect of an EPC thus evaluated is then quantified using the factor

$$[1+(K-1)\cdot x]$$

Details are given in Table 33.

Error [Williams 88]	Failure to isolate a bypass route	
nominal human error probability	$p_0 = 0.003$	
error producing conditions (EPCs)	1. inexperience 2. opposite technique 3. risk misperception 4. conflict of objectives 5. low morale	
total (maximum) multiplicative effects of the EPCs: K_1 ... K_5	$\underline{K} = (3;6;4;2.5;1.2)$	
ratings (assessed proportions of the EPC effects): x_1 ... x_5	$\underline{x} = (0.4;1.0;0.8;0.8;0.6)$	
model	$p = p_0 \prod_{i=1}^{5} [1 + (K_i - 1) \cdot x_i]$ (HEART)	$\log(p) = a \cdot SLI + b$ (SLIM, p-variant)
assessed HEP	$p = 0.003 \cdot 1.8 \cdot 6.0 \cdot 3.4 \cdot 2.2 \cdot 1.12 = 0.27$	$p = 10^{a \cdot SLI + b} = 0.2$ with# $SLI = \sum_{i=1}^{5} [w_i (1 - x_i)] = 0.22$ $w_i = \log(K_i) / \sum_{i=1}^{5} \log(K_i)$ $a = -\sum_{i=1}^{5} \log(K_i) = -2.33$ $b = \log(p_0) - a = -0.19$

See [Reer 93] (Section 5.2.3.3) or [Reer 94c] for mathematical details.

Table 33: Quantitative comparison of HEART and SLIM

ROUGH EVALUATION

For detailed HRA, the use of HEART is only to be recommended with restrictions. However, elements of HEART (e.g. incorporation of certain EPCs) can be used for subtasks or errors identified in the course of an action decomposition and for which no situation-specific error probabilities are available; compare also the evaluation of SLIM in Section 8.1 of this report.

For the quantification of holistic tasks (in the sense of system functions), HEART should be considered as a screening method.

HEART is more user-friendly than SLIM because the user is provided with data on nominal error probabilities (NHEPs) and performance shaping factors (PSFs), called 'error producing conditions' (EPCs) in HEART. The user friendliness is restricted, however, by the very general description concerning the NHEPs.

A positive feature is that HEART provides the user with a list of quantitative data (most of them with sources) on error producing conditions (EPCs). This is a directive development. Expert estimates within probabilistic analyses should use available data as far as possible.

A negative feature is that HEART does not contain a calibration explicitely. It is thus assumed that an EPC selected as being relevant is not yet covered by the selected NHEP. However, the data base from which the NHEP was determined could have contained cases where this EPC was involved in the causation of errors. This would then lead to double consideration.

Table 33 thus illustrates the uncertainties which arise from the model selection alone. The input values are identical. Nevertheless, HEART produces an HEP result which is higher by a factor of 1.35 than the SLIM result. In [Reer 93] (Table 45) it is shown that

- HEART is always more pessimistic than the p-variant of SLIM , and

- the differences (between the HEP results) are negligible for small values of K.

However, further studies are required in order to clarify which model is more correct according to psychological theory.

8.3 INTENT

Brief Description

INTENT is a topical new development in the field of quantification methods based on PSF-structured expert estimates. The publication discussed here [Gertman 92] does not refer to any PSA application of INTENT.

INTENT is specialized in the quantification of decision-based errors due to errors of intention. This corresponds to the error types 3 to 6 outlined in Table 9. The INTENT user is offered a choice of 20 nominal errors. These errors were identified from a data bank with the aid of two computer programs and from North American nuclear operating experience. With respect to the causes that could lead to errors of intention, INTENT classifies these errors into four categories ([Gertman 92], Table 1):

1. Four errors due to the inclusion of possible action consequences, e.g. »tolerate an out of range situation with minor consequences«.
2. Four error attitudes leading to circumventions, e.g. »violate procedure and devise own«.
3. Six errors due to set crew response: e.g. »competing goal states lead to a wrong conclusion«
4. Six errors depending on internal (e.g. memory capacity) and external (e.g. written documents) resources (»resource dependencies«, e.g.: »crew consult inappropriate resource in emergency«.

For each error INTENT gives estimates for the lower bound (LB) and upper bound (UB) of the occurrence probability; these are expert estimates. Moreover, INTENT also includes a set of 11 (very generally and shortly described) performance shaping factors (PSFs) whose weighting factors (w_i, i = 1...11) were also determined by expert estimates.

The user of the INTENT method is now requested to select that error from among the 20 nominal errors which applies to the decision-based error for which a probability is to be estimated. Estimates for the 11 PSFs are then given on a 5-point scale. Taking the weighting factors into acccount it is then possible to calculate a reliability index SLI from these estimates, which is defined in the interval (0;1) ([Gertman 92], equation 2); this corresponds to the SLIM procedure. For the SLI-dependent calculation of the error probability p, however, a mathematical procedure deviating from SLIM is proposed. INTENT assumes that p is equal to the (1-SLI) 100 % quantile of a lognormal distribution with LB as the 5 % quantile and UB as the 95 % quantile.

$$\ln(p) = \sigma \cdot \phi^{-1}(1 - SLI) + \mu$$

with

$$\sigma = \frac{\ln(UB / LB)}{3.29}$$

$$\mu = \frac{\ln(UB\ LB)}{2}$$

where LB is the 5 % lower bound and UB the 95 % upper bound of the error probability.

ROUGH EVALUATION

INTENT is a promising approach complementing the existing methods with respect to the quantification of non-trivial decisions by the personnel.

In comparison to SLIM and HEART, it should be noted as positive that the INTENT method restricts its application range itself. INTENT does not claim to be a complete quantification

method, but is conceived for filling gaps which other quantification methods exhibit in the field of decision-based errors.

The version of INTENT described in [Gertman 92], however, is still unsuitable for practical PSA applications. There is a lack of clear definitions concerning nominal errors and performance shaping factors. Furthermore, there are no explanations concerning the error mechanisms relating to the PSFs, for example »safety culture«, which are the subject of conferences [Carnino 95] extending over several days.

Table 34 shows that the mathematical model underlying the INTENT method leads to approximately the same result as the mathematical model of the p-variant of the SLIM method. On the contrary, the result from the q-variant of SLIM is higher by about a factor of 3 – see also Table 31.

Error (example in [Gertman 92])	Nonuse of the formuli (provided inside the procedures) for technical calculations		
success likelihood index	SLI = 0.39		
calibration points (reference HEPs and respective SLI values)	$p_1 = LB = 0.0016$, $SLI_1 = 0.95$ $p_2 = UB = 0.047$, $SLI_2 = 0.05$		
model	$\ln(p) = \sigma \cdot \phi^{-1}(1-SLI) + \mu$ (INTENT)	$\ln(q) = \ln(1-p)$ $= a \cdot SLI + b$ (SLIM, q-variant)	$\ln(p) = a \cdot SLI + b$ (SLIM, p-variant)
model parameters	$\sigma = \frac{\ln(UB/LB)}{3.29}$ $= 1.03$ $\mu = \frac{\ln(UB \cdot LB)}{2}$ $= -4.75$	$a = \frac{\ln\left(\frac{1-p_1}{1-p_2}\right)}{SLI_1 - SLI_2}$ $= 0.052$ $b = \ln(1-p_1) - a \cdot SLI_1$ $= -0.051$	$a = \frac{\ln\left(\frac{p_1}{p_2}\right)}{SLI_1 - SLI_2}$ $= -3.76$ $b = \ln(p_1) - a \cdot SLI_1$ $= -2.87$
HEP (for SLI=0.39)	$p = e^{\sigma \cdot \phi^{-1}(1-SLI)+\mu}$ $= 0.011$	$p = 1 - e^{a \cdot SLI + b}$ $= 0.03$	$p = e^{a \cdot SLI + b}$ $= 0.013$

Table 34: Quantitative comparison between INTENT and SLIM

Despite all criticism, however, the attempt made in INTENT to classify decision-based errors rather according to their psychological causes and not so much according to the external accompanying circumstances should be judged as a positive development.

8.4 Summary evaluation

PSF-oriented quantifications require

- judgements (i.e. ratings, in some methods together with weightings) according to performance shaping factors (PSFs, wording used in SLIM and INTENT) or error producing conditions (EPCs, wording used in HEART), and
- calibration of the corresponding mathematical model.

According to the items above, SLIM shows a high flexibility on the one hand and requires sophisticated work by the user of the method on the other hand.

The SLIM user is free in selecting or assessing

- the relevant PSFs,
- their ratings,
- their weights, and
- at least two reference HEPs for calibration.

Compared to a user of SLIM, a user of HEART or INTENT has reduced degrees of freedom on the one hand and less sophisticated work to do on the other hand.

In HEART,

- a subset of EPCs has to be selected from a predefined list of 38 EPCs,
- ratings (according to proportions of effect) are required for the selected subset of EPCs, and
- calibration is determined by the assignment of one suitable NHEP (out of a list of nine NHEPs presented for generic described tasks).

INTENT needs

- ratings for 11 preselected (and preweighted) PSFs, and for calibration,
- the selection of one error category (out of a list of 20 predefined categories).

In summary, it may be stated that relatively simple PSF methods (such as SLIM, HEART or INTENT) cannot replace standard methods (such as THERP, EdF'sPHRA or HCR/ORE), which are based on the principles of Boolean algebra (i.e. fault and event trees). Such PSF methods can at most be used as complementary methods in subareas. Boolean algebra is definitely superior to the multiattribute utility theory (MAUT) in the authors' opinion.

Another weak point which PSF methods (such as SLIM, HEART or INTENT) have in common is the neglect of interactions between PSFs. A model simple to handle for the provisional solution of this problem is proposed in [Reer 93] (Section 5.2.3, English version in [Reer 94c]) for the quantification of such interactions.

9 Requirements for a method for human reliability assessment

In several publications, the different HRA methods are discussed and often criticized according to the deficiencies of the methods. Some of these are:

- The problem of the completeness of the analysis. Incompleteness of the methods concerning organisational effects or cognitive errors may result in an erroneous estimation of the total risk, for instance ([Reer 93b], [Bley 92]).
- The problem of insufficient data for assessment and the problem of the relevance of the collected data. HRA data is currently not available for most of the situations to be assessed. Therefore, the necessary transmissions and judgments decrease the accuracy of the used data. This is only covered by high uncertainty bounds. This problem is intensified by the fact that no systematic investigation of the validity of the methods has been performed ([Sträter 94b], [ACSNI 91], [Fujita 92]).
- The problem of modelling human failures in PSA. Firstly, the modelling of human errors into FTA (fault tree analysis) and ETA (error tree analysis) is a simplified assumption treating the human error like a component error. This problem also includes the fact that the probabilistic quantification is to some extent insufficient for depicting the complexity of human behaviour ([Reichart 92], [Heslinga 93]). Secondly, concerning the treatment of PSF (Performance Shaping Factors), the methods only provide insufficient information about the coherency of different PSFs and a given situation. Due to this problem, the analysis is incomplete as is the process for optimization ([Hollnagel 92]). Thirdly, concerning the error types, cognitive errors are insufficiently considered, but are of great importance particularly for the assessment of human reliability in complex disturbance situations especially in computerized technologies and modified types of organisation of work ([Reichart 92], [Reer 93b]).

To structure the different problems and enable an assessment of the methods according to the criticism, a catalogue of requirements will be introduced. These requirements include aspects concerning the methods and the information used by the methods. The catalogue will be subdivided according to the following categories of human reliability assessment:

1. Objectivity of the method
2. Validity of the method
3. Reliability of the method

These distinctions are used in different scientific disciplines like mathematics, philosophy, psychology and can also be considered to subdivide requirements for human reliability assessment (see [Bubb 92], [ACSNI 92], [Swain 89]).

In the context of the assessment of operator actions, objectivity of a method means that the method is able to measure real human failures in the framework of PRA. If a method is less objective, it measures less human failures, but more technical failures for instance. In the theoretical scientific sense, objectivity means precision about what a human mistake is. It is to be distinguished from objectivity in the sense of 'opposite to subjectivity'. In the mathematical sense, objectivity means the correlation between the parameter measured by the method and the real object which should be measured.

Validity is defined as the power of evidence of the method. It means that a method is able to measure the parameters which are used by the method. In the mathematical sense, validity means the correlation of the measured variable with the parameter the method intends to measure.

Reliability of the method is the quality of a method to repeat gained predictions again at a later time or by different persons. In the mathematical sense, reliability means the correlation between the variable measured at different times or by different persons.

Suppose a method for human reliability assessment is optimal in the modelling of personal actions, but the underlying data are insufficient. This means that the method would be an objective one, but would be less valid. Objectivity, validity and reliability are not independent of each other and rely on each other. For an optimal HRA method a maximum of objectivity, validity and reliability is required.

9.1 Catalogue of requirements

Starting from the results of a literature review concerning actual methods for HRA, a catalogue of requirements for HRA methods is introduced according to the above-mentioned categories of requirements. The catalogue may be used for assessment and for the comparison of applications of HRA methods as well as for the evaluation of a new method. The catalogue of requirements is based on requirements mentioned in [IAEA 92], [Bubb 92], [ACSNI 91] and [Swain 89].

Because the three categories of objectivity, validity and reliability are used, this catalogue of requirements may be more balanced with respect to the different requirements than other classifications which only focus on parts of this catalogue. [Swain 89] mainly focuses on the reliability and [ACSNI 91] mainly considers the validity of the methods.

9.1.1 Objectivity of the method

RELEVANCE OF THE RESULTS

The method has to provide

- quantitative results for the operator action to be considered in PSA and
- qualitative results for improvement of the safety and availability of the investigated plant.

COMPLETENESS OF THE METHOD

The method has to be complete regarding

- all types of human actions
- the necessary degree of detail of an action
- the different depths in coarse, detailed and sensitivity analysis
- the error types to be analysed
- all factors influencing the action and
- the definition of an operator error.

9.1.2 Validity of the method

DATA (QUALITATIVE AND QUANTITATIVE)

Original data of the investigated sector of industry should be used (i.e., nuclear power plants in the nuclear area).

Data from simulators may be used, if plant experience is lacking (e.g., in the case of accident management measures). The transferability of the simulator experiments has to be proved by checking the circumstances of the experiments.

Other data (non-nuclear industry, field studies, laboratory experiments, expert judgement) may be used, if neither nuclear plant experience nor simulator experiments are available. The transferability of the data has to be approved. Expert judgement has to be performed by considering psychological constraints of human judgements.

APPROVAL OF THE VALIDITY OF THE METHOD

Regarding the validity of the assessment the following is to be demanded:

- the method must be able to describe human actions realistically
- the data on human behaviour, on which the HRA method was based, has to be transferable to the application of the HRA method in a paticular PSA study.
- indications of the validity of the results like verification by plant experience or simulator experiments
- convergence, i.e. the results have to be in accordance with the results of other methods which have to fit this catalogue of requirements as well
- compatibility of the method, the underlying assumptions and the constraints with human factor knowledge as manifested in widely recognized publications
- common acceptance of the method, criteria are, for instance: the use of the method in various safety studies and by various users, review of the method by other experts and accommodation of the methods according to qualified criticism on the method.

9.1.3 Reliability of the method

CONSISTENCY OF THE RESULTS

If the method is used for a given task

- by different users or
- by the same user at different times

the results have to agree sufficiently.

ACCESSIBILITY TO THE INFORMATION

The information in which the method is described and the needed data are documented has to be

- publicly available and
- transparently documented.

USER AIDS

Within the method, appropriate, well-defined and understandable rules for application have to be available.

If subjective judgement by user is required, the judgement process has to be supported in a qualified manner.

PROOFING

All assumptions, models and data, which are underlying the method, have to be provable and testable (cf. accessibility to the information).

A user has to be guided in a way that his results are provable.

9.2 Summary of the catalogue of requirements

The catalogue of requirements was structured according to the criteria of objectivity, validity and reliability. It allows the weak and strong points of different HRA methods to be worked out as well as a new HRA method to be evaluated. Table 35 summarises the catalogue in a short overview.

Criteria	Requirements of a method for the assessment of human reliability
Objectivity of the method	
Relevance of the results	Does the method provide quantitative and qualitative results for PSA and plant safety?
Completeness of the method	Are all types of human actions considered?
	Is the required degree of detail of an action considered?
	Are coarse, detailed, and sensivity-analysis executed?
	Are error types considered completely and is an error definition given?
	Are the effective influencing factors considered completely?
Validity of the method	
Data (qualitative and quantitative)	Is the following ranking of data sources considered? 1. Original data from the relevant production area (e.g., NPP) 2. NPP simulators (+ description of constraints) 3. Non-nuclear industry, field studies, laboratory experiments, (+ careful review of the transferability) 4. Expert judgement (+ acknowledged method for judgement)
Proof of validity	Are the human actions modelled accurately?
	Are the used data on human errors transferable to the actions to be inspected?
	Can the validity of the results be proved?
	Is the method in concordance with results of other qualified methods?
	Is the method in concordance with human factor knowledge?
	Is the method acknowledged by other experts?
Reliability of the method	
Consistency of the results	Does concordance exist between different users?
	Does concordance exist in the results by using the method at various times?
Accessibility of the information	Is the information generally available?
	Is the information transparent?
User aids	Are application rules or application instruction available?
	Will subjective ratings be supported in a qualified manner?
Proofing	Are the underlying assumptions, models and data provable and testable?

Table 35: Summary of the catalogue of requirements.

10 Conclusions

The evaluations carried out here on the general quantifiability of safety-relevant types of human interactions and on the concrete application of quantification by the example of THERP and other trend-setting methods show the present high quality standard in the field of predictive human reliability analysis (HRA). However, the great significance attributed to safety in our society makes new demands on safety standards and evaluation capability. This also increases the demands made on the quality level of such analyses. Their further development is therefore an important component of publicly sponsored research projects.

There is still a useable potential for industrial HRA applications, especially for the quantitative component of an HRA, because an HRA-produced number and the respective uncertainty bounds provide an aid for decision-making on diffferent levels (e.g. design of displays, preparation of written instructions, or calculation of a premium for the insurance of a certain industrial hazard). For this purpose (producing a number), the current version of THERP (published 1983) is known as an established and widely accepted tool. However, THERP has some weak points, especially according to guidelines and data for the quantification of cognitive error mechanisms in decision-making. Some new developments (published 1984-1994) offer approaches to diminish these weak points:

- data-based time-reliability correlations (TRCs) (EdF'sPHRA, HCR/ORE),
- extended guidelines and data for the incorporation of PSFs (SLIM, HEART, INTENT).

On the whole, THERP is still a standard framework for HRA. The new developments (outlined above) make use of essential basic concepts and data from THERP, e.g.:

- breaking down an accident response into diagnosis and post-diagnosis actions,
- using TRCs for the quantification of diagnosis failures,
- dependence model,
- HEP adjustment according to PSFs, or
- some specific HEPs from the THERP data tables (see Figure 13 in this report).

In this context, it is not advisible to readily adapt guidelines and data from the new developments (1983-1994) because of the sparse published information.

In spite of the advanced state of the art in HRA, there still is a need for further investigations. For example, in [Reer 93b] 21 weak points of the current HRA approach are identified. In addition to data acquisition, there is still a need for basic investigation, especially on the subjects of DEPENDENCE (see Section 6.6) and TIME (Section 7.4).

11 Literature

[ACSNI 91] Advisory Committee on the Safety of Nuclear Installations, *Human Reliability Assessment – a Critical Overview*. Study Group on Human Factors. Second Report. HMSO. London (GB), 1991

[Barriere 94] M. Barriere, W. Luckas, D. Whitehead, A. Ramey-Smith, *An Analysis of Operational Experience During Low Power and Shutdown and a Plan for Addressing Human Reliability Assessment Issues*. NUREG/CR-6093, Washington, DC (USA), 1994

[Beare 84] A. N. Beare, R. S. Dorris, C. R. Bovell, D. S. Crowe, E. J. Kozinsky, *A Simulator-Based Study of Human Errors in Nuclear Power Plant Control Room Tasks*. NUREG/CR-3309, Washington, DC (USA), 1984

[Berg 92] H. P. Berg, H. Schott, *Stand von Wissenschaft und Technik auf dem Gebiet der Quantifizierung der menschlichen Zuverlässigkeit – Dezember 1991 –*. Bundesamt für Strahlenschutz, Fachbereich Kerntechnische Sicherheit, KT-2/92, Salzgitter (D), 1992

[Bello 80] G. C. Bello, V. Colombari, The Human Factors in Risk Analysis of Process Plants: The Control Room Model, TESEO. *6th Advances in Reliability Technology Symposium*. NCSR-R23, United Kingdom Atomic Energy Authority, Warrington (GB), 1980

[Bley 92] D. Bley, S. Kaplan, D. Johnson, The Strengs and Limitations of PSA: Where We Stand, *Reliability Engineering and System Safety* **38** (1992) pp. 3–26

[BMI 77] *Besondere Vorfälle in Kernkraftwerken der Bundesrepublik Deutschland. Berichtszeitraum: 1965-1976*. Der Bundesminister des Innern, Bonn (D), 1977

[Bubb 92] H. Bubb (Ed.), *Menschliche Zuverlässigkeit*. EcoMed, Landsberg (D), 1992.

[Bubb 93] H. Bubb, H. Schmidtke, Systemergonomie. In: H. Schmidtke (Ed.), *Ergonomie*. Hanser, München (D), 1993

[Carnino 95] A. Carnino, G. Weimann (Eds.), *Proceedings of the International Topical Meeting on Safety Culture in Nuclear Installations, Vienna, 24.-25. April 1995*. American Nuclear Society, Austria Local Section, Vienna (A), 1995

[Coombs 75] C. H. Coombs, R. M. Dawes, A. Tversky, *Mathematische Psychologie*. Beltz, Weinheim (D), Basel (CH), 1975

[Degen 94] G. Degen, J. Mertens, B. Reer, ESAP, an Easy-To-Use Expert System for the Systematic Analysis of Operator Actions within the Scope of Probabilistic Safety Assessment. *Proceedings of Seminar/Workshop on Methodological Issues of PSA/HRA and Supporting Software Tools, Gdansk, June 13-17, 1994*, Gdansk (PL), 1994

[Dougherty 93] E. M. Dougherty jr., Context and Human Reliability Analysis. *Reliability Engineering & System Safety* **41** (1993), pp. 25-47

[Dougherty 94] E. M. Dougherty jr., A Note on HRA at PSAM-II. *Reliability Engineering & System Safety* **46** (1994), pp. 291-294

[Drouin 87] M. T. Drouin, J. L. La Chance, B. J. Shapiro, F. T. Harper, T. A. Wheeler, *Analysis of Core Damage Frequency from Internal Events: Grand Gulf, Unit 1*. NUREG/CR-4550 Vol. 6, Washington, DC (USA), 1987

[DRS-B] *Deutsche Risikostudie Kernkraftwerke, Phase B*. Gesellschaft für Reaktorsicherheit, Köln und Garching. Verlag TÜV Rheinland, Köln (D), 1990.

[Edwards 77] W. Edwards,
How to Use Multi-Attribute Utility Measurement for Social Decision Making. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-7, 5, 1977.

[Embrey 83] D. E. Embrey,
The Use of Performance Shaping Factors and Quantified Expert Judgement in the Evaluation of Human Reliability: An Initial Appraisal.
NUREG/CR-2986, Washington, DC (USA), 1983

[Embrey 84] D. E. Embrey, P. Humphreys, E. A. Rosa, B. Kirwan, K. Rea,
SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgement, Vol. I: Overview of SLIM-MAUD, Vol. II: Detailed Analyses of the Technical Issues. NUREG/CR-3518, Washington, DC (USA), 1984

[Embrey 86] D. E. Embrey, J. T. Reason,
The Application of Cognitive Models to the Evaluation and Prediction of Human Reliability. *Proceedings of the International Meeting on Advances in Human Factors in Nuclear Power Systems, Knoxville, 21-24 April 1986.* American Nuclear Society, LaGrange Park (USA), 1986

[ESCIS 86] Einführung in die Risikoanalyse. Systematik und Methoden. Chemische Rundschau, Separatdruck, *Schriftenreihe der Expertenkommission für Sicherheit in der chemischen Industrie der Schweiz (ESCIS)*, Heft 4, 2. unveränderte Auflage, 1986

[EPS 900] IPSN, Framatome,
EPS 900. A Probabilistic Safety Assessment of the Standard French 900 MWe Pressurized Water Reactor. Main Report. Raven International France, Clamart (F), 1990

[EPS 1300] *EPS 1300. Probabilistic Safety Assessment of Reactor Unit 3 in the Paluel Nuclear Power Centre (1300 MWE). Overall Report.* Electricité de France, 1990

[ESI 92] ESI VIII, Risk Management in Complex Production and Transportation System. *Final Report, Eighth European Summer Institute (ESI VIII)*, Ed.: Swedish OR Association, c/o FOA, Sundbyberg (S), 1992

[Flanagan 54] J. C. Flanagan,
The Critical Incident Technique. *Psychol. Bull.* 51 (1954), pp. 327-358

[Fleming 75] K. N. Fleming, P. H. Raabe, G. W. Hannaman, W. J. Houghton, R. D. Pfremmer, F. S. Dombek,
HTGR Accident Initiation and Progression Analysis Status Report, Volume II, AIPA Risk Assessment Methodology. GA/A13617 Vol. II UC-77, General Atomic Co., San Diego (USA), 1975

[Fujita 92] Y. Fujita,
Human Reliability Analysis - A Human Point of View.
Reliability Engineering and System Safety 38 (1992), pp. 71-79

[Gautschi 89] K. Gautschi,
Ausbildung und Übungen des Notfallstabs der Pikettingenieure und der Notfallequipen. In: [SVA 89]

[Gerdes 93] V. G. J. Gerdes,
HRA Techniques; A Selection Matrix. KEMA, Risk and Reliability Analysis group, the Netherlands. *Proceedings of the SRE symposium*, Arnhem, The Netherlands, October 1993: 12 pp

[Gertman 92] D. I. Gertman, H. S. Blackmann, L. N. Haney, K. S. Seidler, H. A. Hahn,
INTENT: A Method for Estimating Human Error Probabilities for Decision Based Errors. *Reliability Engineering & System Safety* 35 (1992) pp.127-136

[Ghertman 85] F. Ghertman, P. Dietz,
Human Error Data Collection Analysis Program Undertaken since 1982 by Electricité de France with INPO. *Proceedings of the ANS/ENS Topical Meeting on Probabilistic Safety Methods and Applications*, Vol. 2, Paper 89. San Francisco (USA), 1985

[Grote 93] G. Grote, C. Künzler,
Sicherheit in soziotechnischen Systemen. Zwischenbericht. ETH Zürich, Polyprojekt „Risiko und Sicherheit technischer Systeme", Polyprojekt-Bericht 05/93, Zürich (CH), 1993

[Haas 82] P. M. Haas, T. F. Bott,
Criterion for Safety Related Plant Operator Actions: A Preliminary Assessment of Available Data. *Reliability Engineering* **3** (1982) pp.59-72

[Hacker 80] W. Hacker,
Allgemeine Arbeits- und Ingenieurpsychologie. VEB Deutscher Verlag der Wissenschaften, Berlin (D), 1980

[Hacker 86] W. Hacker,
Arbeitspsychologie. Huber, Bern (CH), 1986

[Hall 82] R. E. Hall, J. Fragola, J. Wreathall,
Post Event Human Decision Errors: Operator Action Tree / Time Reliability Correlation. NUREG/CR-3010, Washington, DC (USA), 1982

[Hannaman 84] G. W. Hannaman, A. J. Spurgin, Y. Lukic,
Human Cognitive Reliability Model for PRA Analysis.
Draft report NUS-4531, EPRI Project RP2170-3, San Diego (USA), 1984

[Hannaman 85] G. W. Hannaman, A. J. Spurgin, Y. Lukic,
A Model for Assessing Human Cognitive Reliability in PRA studies. *IEEE Third Conference on Human Factors in Nuclear Power Plants, Monterey, California, June 23-27, 1985.*
Institute of Electronic and Electrical Engineers, New York (USA), 1985

[Hannaman 86] G. W. Hannaman, P. Moieni, J. P. Spurgin,
La Salle Human Reliability Measurement Program Data Analysis Report. Draft NUS-4965, NUS Corporation, San Diego, Nov. 1986, (Proprietary)

[Hannaman 88] G. W. Hannaman, D. H. Worledge,
Some Development in Human Reliability Analysis Approaches and Tools. *Reliability Engineering & System Safety* **22** (1988) pp.235-257

[Hansmann 89] W. Hansmann,
Notfallübungen, Empfehlungen und Mitarbeit der Aufsichtsbehörden.
In: [SVA 89]

[Hennings 95] W. Hennings, J. Mertens, B. Reer,
Methodik der Risikoanalyse für Kernkraftwerke. Eine bewertende Bestandsaufnahme mit Blick auf regionale Sicherheitsplanung. vdf Verlag der Fachvereine, Zürich (CH), 1995

[von Herrmann 83a] J. L. von Herrmann,
Methods for Review and Evaluation of Emergency Procedure Guidelines, Volume I: Methodologies. NUREG/CR-3177, Washington, DC (USA), 1983

[von Herrmann 83b] J. L. von Herrmann, W. A. Brinsfield, R. E. Brown,
Methods for Review and Evaluation of Emergency Procedure Guidelines, Volume II: Applications to Westinghouse Plants. NUREG/CR-3177, Washington, DC (USA), 1983

[Heslinga 93] G. Heslinga, H. Arnold,
Human Reliability: To what extent can we consider Humans as System Components.
Proceedings, ENS TOPNUX 1993. Hague, 25-28 April 1993, The Netherlands

[Hirschberg 90] S. Hirschberg (Ed.),
Dependencies, Human Interactions and Uncertainties in Probabilistic Safety Assessment.
Final Report of the NKA Project RAS 470, Västerås, Sweden. ABB Atom AB Library, prepared by ABB Atom, Sweden, Risø National Laboratory, Denmark, Studsvik AB, Sweden, Technical Research Centre of Finland, 1990

[Hoffmann 89] E. Hoffmann,
Vorbereitung am Simulator zur Handhabung auch schwerer Störfälle.
In: [SVA 89]

[Hoffmeister 74] N. Hoffmeister,
Analyse einiger wichtiger Vorkommnisse in Kernkraftwerken. In: SVA (Ed.),
Informationstagung über die Sicherheit von Kernkraftwerken, 25./26. November 1974.
Tagungsreferate. SVA, Zürich (CH), Bern (CH), 1975

[Hollnagel 92] E. Hollnagel,
The Reliability of Man-Machine Interaction. *Reliability Engineering and System Safety* 38 (1992) p.81 ff.

[IAEA 89] *Models and Data Requirements for Human Reliability Analysis.*
IAEA TECDOC 499. IAEA, Vienna (A), 1989

[IAEA 90] *Human Error Classification and Data Collection.*
IAEA TECDOC 538. IAEA, Vienna (A), 1990

[IAEA 92] *Procedures for Conducting Probabilistic Safety Assessments of Nuclear Power Plants (Level 1).* Safety Series No. 50 P, IAEA, Vienna (A), 1992, p.51ff

[Indian Point 82] *Indian Point Probabilistic Study.* Power Authority of the State New York and Consolidated Edison Company of New York, Inc., New York (USA), 1982

[IRS 72] *IRS Kurzinformationen. Kernkraftwerk Obrigheim, Bersten eines Entwässerungsbehälters durch Kühlmittelverlust über das Entwässerungssystem.* Institut für Reaktorsicherheit der Technischen Überwachungsvereine (IRS), Köln (D), 1972

[Kemeny 79] S. G. Kemeny,
Report of the President's Commission on the Accident at Three Mile Island. U.S. Government Printing Office, Washington, DC (USA), 1979

[KFA 81] KFA, GRS,
Sicherheitsstudie für HTR-Konzepte unter deutschen Standortbedingungen, Hauptband zur Phase I B. *Spezielle Berichte der Kernforschungsanlage Jülich - Nr. 136*, KFA, Jül-Spez-136 Bd. 1, Jülich (D), 1981

[KFA 84] KFA,
Sicherheitstechnische Untersuchungen zum Störfallverhalten des HTR-500. *Spezielle Berichte der Kernforschungsanlage Jülich - Nr. 240*, KFA, Jül-Spez-240, Jülich (D), 1984

[Kirwan 94] B. Kirwan, private note, 1994

[Kirwan 95] B. Kirwan, private note, 1995

[Kolaczkowski 86] A. M. Kolaczkowski, J. A. Lambright, W. L. Ferell, N. G. Cathey, B. Najafi, F. T. Harper, *Analysis of Core Damage Frequency from Internal Events: Peach Bottom, Unit 2.*
NUREG/CR-4550 Vol. 4, Washington, DC (USA), 1986

[Kolb 82] G. J. Kolb, D. M. Kunsman, B. J. Bell, N. L. Brisbin, D. D. Carlson, S. W. Hatch, D. P. Miller, B. J. Roscoe, D. W. Stack, R. B. Worell, J. Robertson, R. O. Wooton, S. H. McAhren, W. L. Ferrell, W. J. Galyean, J. A. Murphy,
Interim Reliability Program: Analysis of the Arkansas Nuclear One - Unit 1 Nuclear Power Plant. NUREG/CR-2787 Vols. 1 and 2, Washington, DC (USA), 1982

[Kosmowski 94] K. T. Kosmowski, G. Degen, J. Mertens, B. Reer,
Development of Advanced Methods and Related Software for Human Reliability Evaluation within Probabilistic Safety Analyses. *Berichte des Forschungszentrums Jülich; 2928*, KFA, Jül-2928, Jülich (D), 1994

[Kosmowski 94b] K. T. Kosmowski, private note, 1994

[Kozinski 84] E. J. Kozinski, L. H. Grey, A. N. Beare, D. B. Burks, F. E. Gomer,
Safety-Related Operator Actions: Methodology for Developing Criteria. NUREG/CR-3515, Washington, DC (USA), 1984

[Lanore 87] J. M. Lanore, J. L. Caron, A. Ellia-Hervy, J. L'Henoret,
Interaction Between Thermal/Hydraulics, Human Factors and System Analysis for Assessing Feed and Bleed Risk Benefits. In: *Probabilistic Safety Assessment and Risk Managment/PSA'87*, Volume 1, pp 116-119. Europ. Nuclear Soc. / Swiss Nuclear Soc., Verlag TÜV Rheinland, Köln (D), 1987

[Madjar 93] M. Madjar,
Überblick über Methoden und Einflußgrößen bei der Risikoermittlung chemischer Anlagen.
ETH Zürich, Polyprojekt „Risiko und Sicherheit technischer Systeme", Polyprojekt-Bericht 04/93, Zürich (CH), 1993

[Miller 87] D. P. Miller, A. D. Swain,
Human Error and Human Reliability. In: G. Salvendy (Ed.), *Handbook of Human Factors/Ergonomics*, Chapter 2.8. Wiley & Sons, New York (USA), 1987

[Moieni 86] P. Moieni, G. W. Hannaman,
Uncertainty Analysis in Time-Dependent Human Reliability Models - Application to the Human Cognitive Reliability (HCR) Model. NUS-4915, NUS Corporation, Gaithersburg, MD (USA), 1986

[Moieni 94] P. Moieni, J. Spurgin, A. Singh,
Advances in Human Reliability Analysis Methodology. Part I: Frameworks, Models and Data. *Reliability Engineering and System Safety* 44 (1994), pp. 27–55

[Mosneron 89] F. Mosneron Dupin,
Méthode pratique de prise en compte du Facteur Humain dans les études de séquences accidentelles. EPS FH 006 c, Electricité de France, 1989, (restricted circulation)

[Mosneron 90] F. Mosneron Dupin, A. Villemeur, J. M. Moroni,
Paluel Nuclear Power Plant PSA: Methodology for Assessing Human Reliability. *7th International Conference on Reliability and Maintainability, Brest, 1990*, Brest (F), 1990

[Mosneron 92] F. Mosneron Dupin, G. Saliou, F. Lars,
Probabilistic Human Reliability Analysis: The Lessons Derived from Plant Operation at Electricité de France. In: IAEA (Ed.), *Use of Probabilistic Safety Assessment for Operational Safety. PSA '91. Proceedings of an International Symposium, Vienna, June 3-7, 1991*. IAEA-SM-321/57, Vienna (A), 1992

[Mosneron 93] F. Mosneron Dupin,
Characteristics of the EdF Probabilistic Human Reliability Analysis Methodology. Note Interne Technique, T54/93-20, EDF/DER/RNE, Electricité de France, 1993

[Mosneron 94] F. Mosneron Dupin,
Is Probabilistic Human Reliability Analysis Possible? Presented at the *EdF International Seminar on PSA and HRA, Paris, November 21-23, 1994*, Paris (F), 1994

[Mosneron 95] F. Mosneron Dupin, private note, 1995

[Mosneron 96] F. Mosneron Dupin, private note, 1996

[MUSA] D. Haschke,
Die probabilistische Sicherheitsanalyse des Kernkraftswerks Mühleberg. In: *SVA-Vertiefungskurs „Fortgeschrittene Sicherheitsanalyse"*. Schweizerische Vereinigung für Atomenergie, Bern (CH), 1991.

[NUREG-1154] *Loss of Main and Auxiliary Feed Water Event at the Davis-Besse Plant on June 9, 1985*. NUREG-1154, Washington, DC (USA), 1985

[Oswald 82] A. J. Oswald, C. D. Gentillon, S. D. Matthers, T. R. Mitchum,
Generic Data Base for Data and Models Chapter of the National Reliability Evaluation Program (NREP) Guide. EGG-EA-5887
Informal Report, EG&G Idaho, Inc., Idaho Falls (USA), 1982

[Parry 91] G. W. Parry, A. Singh, A. Spurgin, P. Moeini, A. Beare,
An Approach to the Analysis of Operating Crew Responses With Simulator Exercises. In: U. Hauptmanns (Ed.), *Proceedings of the OECD/BMU Workshop on Special Issues of Level 1 PSA*. GRS-86, Köln (D), 1991

[Parry 92] G. W. Parry et al.,
An Approach on the Analysis of Operator Actions in Probabilistic Risk Assessment. EPRI TR-100259, Electric Power Research Institute, Palo Alto (USA), 1992

[Pew 81] R. W. Pew, D. C. Miller, C. E. Feeher,
Evaluation of Proposed Control Room Improvements Through Analysis of Critical Operator Decisions. EPRI-NP-1082, Electric Power Research Institute, Palo Alto (USA), 1981

[Phillips 85] L. D. Phillips, P. Humphreys, D. E. Embrey, D. L. Selby,
A Socio-Technical Approach to Assessing Human Reliability (STAHR). In: *Pressurized Thermal Shock Evaluation of the Calvert Cliffs Unit 1 Nuclear Power Plant*, Appendix D. NUREG/CR-4022, Washington, DC (USA), 1985

[Potash 81] L. M. Potash, M. Stewart, P. E. Dietz, C. M. Lewis, E. M. Dougherty Jr.,
Experience in Integrating the Operator Contributions in the PRA in Actual Operating Plants.
In: *Proceedings of the ANS/ENS Topical Meeting on Probabilistic Risk Assessment, Port Chester, NY, September 1981*, Vol. II, pp. 1054-1063. American Nuclear Society, LaGrange Park (USA), 1981

[Poucet 88] A. Poucet,
Survey of Methods Used to Assess Human Reliability in the Human Factors Reliability Benchmark Exercise. *Reliability Engineering & System Safety* **22** (1988) pp. 257-268

[PRA-PG] *PRA Procedures Guide – A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants. Final Report.*
NUREG/CR-2300, Washington, DC (USA), Jan. 1983

[Rasmussen 79] J. Rasmussen,
On the Structure of Knowledge - A Morphology of Mental Models in a Man-Machine Context. Risø-M-2192, Risø National Laboratory, Roskilde (DK), 1979

[Reason 90] J. Reason,
Human Error. Cambridge University Press, Cambridge (GB), 1990

[Reason 94] J. Reason,
Errors, Outcomes and Circumventions: a Reply to Dougherty, *Reliability Engineering & System Safety* **46** (1994), pp. 297-298

[Reer 88] B. Reer,
Deutsche Übersetzungen und Erläuterungen zu den Daten aus dem Swain-Handbuch über menschliche Zuverlässigkeit in Kernkraftwerken. Interner Bericht, KFA-ISF-IB-6/88, Forschungszentrum Jülich GmbH, Jülich (D), 1988

[Reer 93] B. Reer,
Entscheidungsfehler des Betriebspersonals von Kernkraftwerken als Objekt probabilistischer Risikoanalysen. *Berichte des Forschungszentrums Jülich; 2811*, KFA, Jül-2811, Jülich (D), 1993

[Reer 93b] B. Reer, J. Mertens,
Zukünftige Forschungsthemen zur systematischen Erweiterung der Methodik probabilistischer Mensch-Maschine Systemanalysen. Interner Bericht, KFA-ISR-IB-9/93. Institut für Sicherheitsforschung und Reaktortechnik (ISR), Forschungszentrum Jülich GmbH (KFA), Jülich (D), 1993

[Reer 94a] B. Reer,
Dynamische Analysemethode für Operateurhandlungen (DAO): Ein Verfahren zur Quantifizierung der zeitabhängigen Zuverlässigkeit von Notfallmaßnahmen. In: Deutsches Atomforum e.V. und Kerntechnische Gesellschaft e.V. (Ed.), *Jahrestagung Kerntechnik '94. Tagungsbericht.* INFORUM Verlagsgesellschaft, Bonn (D), 1994

[Reer 94b] B. Reer,
A Probabilistic Method for Analyzing the Reliability Effect of Time and Organizational Factors. *European Journal of Operational Research* **75** (1994), pp. 521-539. (Revised version of a paper presented in [ESI 92])

[Reer 94c] B. Reer,
Incorporation of Risk-Taking Behaviour into Human Reliability Analysis. In: A. Stritar (Ed.), *International Meeting: PSA/PRA and Severe Accidents '94.* Nuclear Society of Slowenia, Ljubljana (SLO), 1994

[Reer 94d] B. Reer, F. Przybylski, W. Ullwer,
An Extended View of Plant Specific Data for Human Reliability Analysis: The Impact of Failure Data and Modelling Details.
In: H.-P. Balfanz (Ed.), *4th TÜV Workshop on Living PSA Application in Hamburg, 2-3 May 1994.* TÜV Nord e.V., Hamburg (D), 1994

[Reichart 85] G. Reichart, in:
W. Frey, H. Hörtner, J. von Linden, G. Rappl, G. Reichart,
Deutsche Precursor Studie - Auswertung anlagenspezifischer Betriebserfahrung im Hinblick auf Vorläufer zu möglichen schweren Kernschäden. Report GRS-A-1149, GRS, Köln (D), Garching (D), 1985

[Reichart 92] G. Reichart, in: H. Bubb (Ed.),
Menschliche Zuverlässigkeit. EcoMed, Landsberg (D), 1992, pp. 106 ff.

[Roth-Seefried 89] H. Roth-Seefried,
Betriebsanweisungen zur Störfallbeherrschung bei Siemens/KWU-Druckwasser-Reaktoren.
In: [SVA 89]

[Samanta 85] P. K. Samanta, J. N. O'Brien, H. W. Morrison,
Multiple Sequential Failure Model: Evaluation of and Procedure for Human Error Dependency. NUREG/CR-3837, Washington, DC (USA), 1985

[Seaver 83] D. A. Seaver, W. G. Stillwell,
Procedures for Using Expert Judgement to Estimate Human Error Probabilities in Nuclear Power Plant Operations. NUREG/CR-2743, Washington, DC (USA), 1983

[Semmer 94] N. Semmer,
Der menschliche Faktor in der Arbeitssicherheit: Mechanismen, Verhütung und Korrektur von menschlichen Fehlhandlungen. In: SVA (Ed.), *SVA-Ausbildungsseminar (Workshop), Ursachenanalyse von Störfällen in Kernkraftwerken*. Schweizerische Vereinigung für Atomenergie (SVA), Bern (CH), 1994

[SHARP] G. W. Hannaman, A. J. Spurgin,
Systematic Human Action Reliability Procedure (SHARP). EPRI-NP-3583, Electric Power Research Institute, Palo Alto (USA), 1984

[Siegel 84] A. I. Siegel, W. D. Bartter, J. J. Wolf, H. E. Knee,
Maintenance Personnel Performance Simulation (MAPPS) Model: Description of Model Content, Structure, and Sensitivity Testing. NUREG/CR-3626, Washington, DC (USA), 1984

[Smidt 79] D. Smidt,
Reaktorsicherheitstechnik. Springer, Berlin (D), 1979

[Spurgin 90] A. J. Spurgin et al.,
Operator Reliability Experiments Using Power Plant Simulators, Vols. 1 and 2: Executive Summary and Technical Report. EPRI-NP-6937, Electric Power Research Institute, Palo Alto (USA), 1990

[Sträter 94a] O. Sträter,
The Role of Plant Experience to Consider the Human Factor in Living PSA.
In: H.-P. Balfanz (Ed.), *4th TÜV-Workshop on Living PSA Application in Hamburg, 2.-3. May 1994*. TÜV Nord, Hamburg (D), 1994

[Sträter 94b] O. Sträter, W. Preischl, A. Berning,
Vergleichende Auswertung probabilistischer Sicherheitsanalysen im Bereich Personalhandlungen: Untersuchung spezieller Methodenfragen zu Personalhandlungen.
GRS-A-2151, Gesellschaft für Reaktorsicherheit (GRS) mbH, Köln (D), 1994

[Sträter 95a] O. Sträter,
Eine Methode zur Erfassung und Auswertung von Betriebserfahrung im Hinblick auf menschliche Fehler. In: *Technische Zuverlässigkeit. Tagung Fulda. 26/27.9.1995*, VDI-GSP, Düsseldorf (D), 1995

[Sträter 95b] O. Sträter,
Ergonomic Principles for Accident Management Support (AMS). Final Report for the EU project on AMS, Gesellschaft für Reaktorsicherheit (GRS) mbH, Köln (D), 1995

[SVA 89] *SVA-Vertiefungskurs „Störfallmanagement in Kernkraftwerken"*. Schweizerische Vereinigung für Atomenergie, Bern (CH), 1989

[Swain 80] A. D. Swain, H. E. Guttmann,
Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Draft Report. NUREG/CR-1278, Washington, DC (USA), 1980

[Swain 83] A. D. Swain, H. E. Guttmann,
Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. Final Report. NUREG/CR-1278, Washington, DC (USA), 1983

[Swain 87] A. D. Swain,
Accident Sequence Evaluation Program Human Reliability Analysis Procedure. NUREG/CR-4772, Washington, DC (USA), 1987

[Swain 88] A. D. Swain, L. M. Weston,
An Approach to the Diagnosis and Misdiagnosis of Abnormal Conditions in Post-Accident Sequences in Complex Man-Machine Systems. In: L. P. Goodstein, H. Anderson, S. Olson (Ed.), *Task, Errors and Mental Models.* Taylor & Francis, London (GB), New York (USA), 1988

[Swain 89] A. D. Swain,
Comparative Evaluation of Methods for Human Reliability Analysis. GRS-71, Gesellschaft für Reaktorsicherheit (GRS) mbH, Köln (D), Garching (D), 1989

[Villemeur 86] A. Villemeur, J. M. Moroni, F. Mosneron Dupin, T. Meslin,
A Simulator-Based Evaluation of Operator Behaviour by Electricité de France. *Proceedings of the International Meeting on Advances in Human Factors in Nuclear Power Systems, Knoxville, 21-24 April 1986.* American Nuclear Society, LaGrange Park (USA), 1986

[Vuorio 87] U. M. Vuorio, J. K. Vaurio,
Advanced Human Reliability Analysis Methodology and Applications. In: *Probabilistic Safety Assessment and Risk Management / PSA'87*, Volume 1, pp. 104-109. Europ. Nuclear Soc. / Swiss Nuclear Soc., Verlag TÜV Rheinland, Köln (D), 1987

[Wakefield 87] D. J. Wakefield, C. D. Adams,
Quantification of Dynamic Human Errors in the TMI-1 PRA. In: *Probabilistic Safety Assessment and Risk Management / PSA'87*, Volume 1, pp. 110-115. Europ. Nuclear Soc. / Swiss Nuclear Soc., Verlag TÜV Rheinland, Köln (D), 1987

[Wakefield 92] D. I. Wakefield, G. W. Parry, G. W. Hannaman, A. J. Spurgin,
SHARP1 - Revised Systematic Human Action Reliability Procedure, Final Report, TR-101711, Research Project 3206-01, Electric Power Research Institute, Palo Alto (USA), 1992

[Wechsler 52] D. Wechsler,
Range of Human Capacities. Williams & Wilkins, Baltimore (USA), 1952

[Whalley 89] S. P. Whalley, B. Kirwan,
An Evaluation of Five Human Error Identification Techniques. *5th International Loss Prevention Symposium, Oslo, 1989*, Oslo (N), 1989

[Wickens 84] C. D. Wickens,
Engineering Psychology and Human Performance. C. E. Merrill Publishing Company, A Bell & Howell Company. Columbus, Toronto (CDN), 1984.

[Williams 88] J. C. Williams,
A Data-Based Method for Assessing and Reducing Human Error to Improve Operational Performance. *Proceedings of the IEEE 4th Conference on Human Factors in Power Plants, Monterey, California, June 6–9, 1988.* Institute of Electronic and Electrical Engineers, New York (USA), 1988

[Woods 82] D. D. Woods, J. A. Wise, L. F. Hanes,
Evaluation of Safety Parameter Display Concepts, Vol. 1. EPRI-NP-2239, Electric Power Research Institute, Palo Alto (USA), 1982

[Woods 84] D. D. Woods,
Some Results on Operator Performance in Emergency Events. *Institute of Chemical Engineers Symposium Series,* No. 90: 21, 1984, (quoted from: [Embrey 86])

[Woods 88] D. D. Woods, E. Roth, H. Pople jr.,
Modelling Human Intention Formation for Human Reliability Assessment. *Reliability Engineering & System Safety* 22 (1988), pp. 169–200

[Woods 90] D. D. Woods, E. Roth, H. Pople jr.
The Cognitive Environment Simulation as a Tool for Modeling Human Performance, Main Report. NUREG/CR-5213, Vol. 2, Washington, 1990

[Wortman 78] D. B. Wortman, S. D. Duket, D. J. Seifert, R. L. Hann, G. P. Chubb,
The SAINT User's Manual. AMRL-TR-77-62, Wright-Patterson Air Force Base, OH (USA), 1978

[Yeh 86] Y.-C. Yeh, M. G. K. Evans,
Model for Human Reliability and its Incorporation in a PRA. In: *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems, Knoxville, 21-24 April 1986.* American Nuclear Society, LaGrange Park (USA), 1986

[Zimolong 90] B. Zimolong,
Fehler und Zuverlässigkeit. In: C. F. Graumann et al. (Eds.), *Enzyklopädie der Psychologie, Themenbereich D, Serie III, Bd. 2.* Verlag für Psychologie Hogrefe, Göttingen (D), 1990

[Zimolong 91] B. Zimolong,
Empirical Evaluation of THERP, SLIM and Ranking to Estimate HEPs. *Reliability Engineering and System Safety* 35 (1991), pp.1-11

[Zion 81] *Zion Probabilistic Safety Study.* Commonwealth Edison Co., Chicago (USA), 1981

12 Abbreviations

AIPA	Accident Initiation and Progression Analysis
ASEP	Accident Sequence Evaluation Program
BHEP	Basic Human Error Probability
BMI	Federal Minister of the Interior
BWR	Boiling Water Reactor
CD	Complete Dependence
CES	Cognitive Environment Simulation
CM	Confusion Matrix
DNE	Direct Numerical Estimation
DRS-B	Deutsche Risikostudie Kernkraftwerke (German Risk Study on Nuclear Power Stations), Phase B
ECOM	Error of Commission
EdF	Electricité de France
Ed.	Editor
EOM	Error of Omission
EPRI	Electric Power Research Institute
ESAP	Expert System for the Systematic Analysis of Personnel Actions
ESCIS	Expertenkommission für Sicherheit in der chemischen Industrie der Schweiz
HD	High Dependence
HCR	Human Cognitive Reliability model
HCR/ORE	Human Cognitive Reliability/Operator Reliability Experiment
HEART	Human Error Assessment and Reduction Technique
HEP	Human Error Probability
HPI	High Pressure Injection
HRA	Human Reliability Analysis
INTENT	Method for estimating HEPs for decision-based errors of intention
IRS	Incident Reporting System
KFA	Kernforschungsanlage (Research Centre) Jülich (legal name since 1990: Forschungszentrum Jülich GmbH)
LB	Lower Bound
LD	Low Dependence
MAPPS	Maintenance Personnel Performance Simulation model
MAUT	Multiattribute Utility Theory
MD	Moderate Dependence
MDN	Median
MSFM	Multiple-Sequential Failure Model
NHEP	Nominal Human Error Probability
NPP	Nuclear Power Plant
OAT	Operator Action Tree

ORE	Operator Reliability Experiment
PC	Paired Comparison
PHRA	Probabilistic Human Reliability Analysis
PRA	Probabilistic Risk Analysis
PRA-PG	PRA Procedures Guide
PSA	Probabilistic Safety Analysis
PSF	Performance Shaping Factor
PWR	Pressurized Water Reactor
RO	Reactor Operator
SAINT	System Analysis of Integrated Networks of Tasks
SLI	Success Likelihood Index
SLIM	Success Likelihood Index Methodology
SHARP	Systematic Human Action Reliability Procedure
SMART	Simple Multiattribute Rating Technique
STAHR	Socio-Technical Approach to Assessing Human Reliability
TESEO	Tecnica Empirica Stima Errori Operatori Model
THERP	Technique of Human Error Rate Prediction
TMI	Three Mile Island
UB	Upper Bound
UF	Uncertainty Factor (Error Factor)
ZD	Zero Dependence

Jül-3222
May 1996
ISSN 0944-2952