Dédale

Pour

DIRECTION GENERALE SECURITÉ

Help to reflect on the strategy to reduce ESRs
Note 1

[version V3]

Suivi des versions

Date	Version	Contributions	Principaux changements
26-010-2018	V0	J Pariès	Création du document
08-12-2018	V1	J Pariès	Version provisoire ; première partie
08-01-2019	V2	J Pariès	Complément deuxième partie
11_02_2019	V3	J Pariès	Prise en compte des commentaires sur V2

Vos interlocuteurs pour cette proposition :

Suivi Technique
Jean Paries
+33 6 07 68 04 93
jparies@dedale.net

Suivi administratif et financier
Véronique Moulin
+33 6 62 07 94 92
vmoulin@dedale.net

Table des matières

1 Issues and Objectives 4

2 References in matter of human reliability 5

2.1 Metrics in human reliability [1] 5

2.2 Interactions with the issues 6

3 Application for Train Drivers (TD) and traffic Controlers (TC) SNCF 9

3.1 Datas 9

3.1.1 TD : 9

3.1.2 *TC* : 11

3.2 Discussion 11

4 Conclusion 13

5 Références 14

6 ANNEXE 1 15

1 Issues and Objectives

The SNCF Safety Directorate has undertaken a strategic reflection on ways to significantly reduce the frequency of SREs, remarkable safety events, and ES, safety events, precursors of serious accidents. In particular, it is interested in events of this type depending on the reliability of front-line operators, such as train drivers (ADC) and traffic officers (AC). For the former, these are no-urges stop signal crossings (FSA) and speed limit (DVL) crossings. For the latter, it is shipping without order or with wrong order. The indicative objective is to substantially reduce these types of SREs by 2023. Their current perimeter covers around 90 ESR and 1200 ES annually.

The traditional strategy aims to reduce the frequency of these events by analyzing the causes of those that occur, and the attempt to eradicate the factors of cause re-represented or anticipated. With regard to events related to front-line actors, this strategy includes corrective actions such as "FOH": improved work environments, procedures (e.g. simplification), ergonomics of agent/work tools interfaces, technical skills training but also non-technical; implementation of "human reliability tools" (pre-job briefings, checklists, self-checks and cross-checks, secure communication, etc.); strengthening and bringing management closer together; strengthening teamwork, etc.

This strategy works well, but only to a certain extent. Beyond this point, we stumble upon the intrinsic limits of human reliability, i.e. on an incompressible "background noise" of "errors" associated with an inevitable repetition of the same causes, or on the contrary their renewal both unpredictable and inexorable. The performance of safety improvement efforts then declines rapidly, and the traditional strategy of improving human reliability must be extended through a more systemic approach to reducing the criticality of human reliability in the safety model (i.e., reducing the dependence of system safety on human reliability). Such an approach may be based, for example, on reducing exposure to risk (or even withdrawal from exposure) processing processes, or adding automatic protections.
In defining and adjusting a safety improvement strategy, it is therefore pri-mordial to be able to position ourselves in relation to these 'aymptotic' situations, and to know how to re-create their approach.

A first way to do this is to directly monitor the evolution of the effectiveness of safety clearances, in order to detect the appearance of diminishing yields. This obviously implies a specific ability to measure so-called "secure" investments and results obtained, which is far from easy on the one hand, and it also requires being able to discriminate against the effects of the investments in question. effects of other changes over the period.

Another way to do this, or even complementary to above, is to objectify the perceived human reliability of the agents concerned, and to comparize it with scientific or empirical references, in the railway field or comparable activities, indicating what can be achieved.

This note aims to contribute to this second approach, by providing a framework for assessing the reliability of operators and then using it to assess the current situation through the data that can be collected at this stage.

2 References in matter of human reliability

2.1 Metrics in human reliability [1]

The Human Reliability Assessment (HRA) was born as a methodological approach in the United States in the early 1960s in the field of defence. The aim was to take into account the probability of human error, as well as technical failures, in the reliability analyses associated with strategic missile operations. The work focused on the development of databases concerning the generic reliability of elegy human actions such as activating a push button. They have failed because the notion of human reliability at this level of granularity makes no sense: it varies by several orders of magnitude depending on the goals and contexts of the activity in which these actions are inserted. However, research and development on the issue continued until the late 1970s.

In 1979, the psychological shock of the Three Mile Island nuclear power plant accident brought the issue of human reliability into the spotlight. The result has been the development of several methods of taking human error into account, known as "1st generation. They are based on the decomposition of activities into tasks, and the assessment of the probability of failure of these tasks. This is obtained through the use of a standard error probability database, and a calculation of the influence on these basic values of factors such as stress, time pressure or equipment design. Sometimes the 'calculation' is supplemented by the use of expert judgments.

The main method is THERP (Technique for Human Error Rate Prediction), which remains in use today. THERP is based on the work of Swain and Guttman of Sandia Laboratories, published in 1983 [2], and was developed for the U.S. Nuclear Regulatory Commis-Sion (NRC). The method predicts a probability of human error (HEP: Human Error Probability) associated with a type of error. The prediction is based on a database of generic error frequencies (data based on Swain's experiments, and never published... and data from early U.S. defense studies). These generic values are then particularized to take into account the specifics of the task through the concept of Performance Shaping Factors (PSF). The possibilities of common modes of error are also taken into account in the calculation, which is important. The result is an average value of the probability of error for a given task, with an uncertainty range of 5%-95%.

THERP has been criticized for the same reasons as the attempts of the 1960s: its decomposition of the activity into tasks (even if they are less basic) and its difficulty in taking contexts into account. The very wide variability of human reliability according to goals and contexts is indeed difficult to accommodate such modelling, even by challenging a fairly generic error taxonomy and involving "Performance Shaping Factors". In a humorous way, it was said that the probability of error established under this type of method was in the form 1.47.10-4- -10-2.

This substantive criticism, considering that the activity is still located and finalized, and therefore cannot be isolated from its context, gave rise to the development in the 1990s of a "2nd generation" of HRA methods that holistically reason about modes of failure of the expected function or the overall mission in the situation, and no longer on the error of the operators considered in isolation. CREAM [3] , ATHENA [4], MERMOS [5] (developed by EDF), CAHR, are the best known of this method. They are sometimes used in industry but are qualitative, not quantitative, and remain complex.

In parallel with this development of second-generation methods, people have continued to work on simplified and improved versions of first-generation methods. In 1986 Williams published the Human Error Assessment and Reduction Tech-nic (HEART] method. The method incorporates the principles of THERP but with a much more generic (and therefore limited) error database. Its use was made simpler and more flexible. A generic task list (GTT) is associated with a database of error probabilities on those tasks. For a specific task, the GTT is chosen that best suits it, which provides a first quantification of error probability. Performance factors - here called Error Producing Conditions (EPC) - are each associated with a maximum effect (a multiplier) on the probability of basic errors. This maximum effect is then weighted by an impact factor on the actual task, appreciated by expert judgment.

The practicality of the HEART method has earned it good success in the industry, but the reverse of this simplicity was the difficulty in matching the general errors processed by the method and the errors made in practice in a industry on specific tasks. This difficulty has led to the development of party-cularized versions in certain areas, sometimes referred to as 3rd generation methods. This family is derived from HEART, narA (Nuclear Action Reliability Assessment) for nuclear, CARA (Controller Action Reliability Assessment) for air traffic control, and RARA (Railway Action Reliability) Assessment) [7] for the railway sector. Appendix 1 shows how to proceed with the RARA method.

In parallel to these methodological developments specific to certain areas, an effort has been made to develop a common database on the probability of generic errors. The CORE-DATA database was originally developed in 1995 at the University of Birmingham [7] and computerised with the support of the British Health and Safety Executive in 1999 [8] [9] [10]. This database includes data collected in the following industries: nuclear, offshore, manufacturing, railways, chemicals and aviation. She remains housed at the University of Birmingham and is not open access.

2.2 Interactions with the issues

HRA methods have been, and still are, developed to support proactive risk management approaches. The objective is then to assess or even quantify the probability of human error (i.e. clearly errors of front-line operators) that could play a more or less critical role in the reliability or safety of the systems in design course. The ambition is therefore high: the predictive quantification of the "mistakes" of front-line operators.

In our case, the ambition is a little different. The aim is to compare the human reliability found in real operations in specific activities, with reference values, in order to infer the existence or absence of margins of progression. This is a more reasonable ambition, because the system is known and the measurement of current reliability should not be insurmountable. And because this is an overall reasoning, we can be satisfied with the average values: it is not necessary to determine the ranges of variation in the number of different context factors (which pose the most difficulty). We want to compare the (average) values observed with a reference of what is reasonably possible to expect from human performance. The basic values of generic error frequencies used in HEART or its domain derivatives can therefore be taken into account.

The difficulty is then to find the right matches between the activities considered and the generic "tasks" that appear in the nomenclature of the method. Establishing these corres-pondances, i.e. translating HEART's generic vocabulary into business vocabulary, was precisely the goal of rarA's development in the railway world, under the aegis of the RSSB. Unfortunately, the corresponding data is not freely available on the RSSB website, or even on the SPARK part reserved for members, for non-RARA members. However, generic values and EPC coefficients can be found in the RSSB Railway Action Reliability Assessment User Manual - A technique for the quantification of human error in the rail industry [11].

Les taux d'erreurs associés aux tâches génériques sont indiqués dans le tableau suivant :

Area	GTT	HEP	Bounds	More information
More automated and skill-based processes	R1. Respond correctly to system command even when there is an automated system providing accurate interpretation of system state.	0.00002	0.000006-0.0009	p. 35
	R2. Completely familiar, well designed, highly practiced task which is routine.	0.0004	0.00008-0.007	p. 36
	R3. Simple response to a dedicated alarm and execution of actions covered in procedures.	0.0004	0.00008-0.007	p. 36
	R4. Skill-based tasks (manual, visual or communication) when there is some opportunity for confusion.	0.003	0.002-0.004	p. 37
	R5. Fairly simple task performed rapidly or given insufficient or inadequate attention.	0.09	0.06-0.13	p. 38
More effortful and rule-based processes	R6. Restore or shift a system to original or new state, following procedures with some checking.	0.003	0.0008-0.007	p. 39
	R7. Identification of situation requiring interpretation of alarm/ indication patterns.	0.07	0.02-0.17	p. 40
Thinking outside procedures	R8. Complex task requiring a high level of understanding and skill.	0.16	0.12-0.28	p. 41

Tableau 1 : RARA- Probabilités d'erreur pour des tâches génériques

Referring to this table, without taking into account the aggravating factors below, we can therefore consider, for example, that the reference reliability of closed signal compliance, which corresponds to the R1 category of the first line, is (1- 2x10-5).

The value of (1-10-5) is in fact generally presented in the literature as the unsurpassable limit of human reliability for simple automated tasks.

The Impact of Performance Factors (EPC) is shown in the following table
:

Area	Ref	Error producing condition	Max. affect	More info.
Task design	T1	Unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel.	17	p. 46
	T2	A shortage of time available for error detection and correction.	11	p. 47
	T3	A need to unlearn a technique and apply one which requires the application of an opposing philosophy.	8	p. 48
	T4	The need to transfer specific knowledge from task to task without loss.	5.5	p. 49
	T5	An impoverished quality of information conveyed by person/person interaction.	3	p. 50
	T6	Little or no independent checking or testing of output.	3	p. 50
	T7	A conflict between immediate and long-term objectives.	2.5	p. 51
	T8	Unclear allocation of function and responsibility.	1.6	p. 51
	T9	A danger that finite physical capabilities will be exceeded.	1.4	p. 52
	T10	Prolonged inactivity or highly repetitious cycling of half hour low mental workload tasks.[1]	1.1 1.05	p. 52
Interface	In1	A low signal-noise ratio.	10	p. 53
	In2	A means of suppressing or over-riding information of features which is too easily accessible.	9	p. 54
	In3	No means of conveying spatial and functional information to operators in a form which they can readily assimilate.	8	p. 55
	In4	A mismatch between an operator's model of the world and that imagined by a designer.	8	p. 56
	In5	No obvious means of reversing an unintended action.	8	p. 57
	In6	A channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information.	6	p. 58
	In7	Poor, ambiguous or ill-matched system feedback.	4	p. 58
Competence management	C	Operator inexperience.	3	p. 59

Area	Ref	Error producing condition	Max. affect	More info.
Procedures	PR1	Ambiguity in the required performance standard.	5	p. 60
	PR2	An impoverished quality of information conveyed by procedures.	3	p. 60
Person	P1	A mismatch between perceived and real risk.	4	p. 61
	P2	Fatigue from shift and work patterns.[2]	2.6	p. 61
	P3	High level emotional stress.	2	p. 61
	P4	Little opportunity to exercise mind and body outside the immediate confines of a job.	1.8	p. 62
	P5	Little or no intrinsic meaning in a task.	1.4	p. 62
	P6	Low workforce morale.	1.2	p. 62
Environment	E	A poor or hostile environment.	8	p. 63

1: 1.1 for first half hour. x 1.05 for each hour thereafter.

2: The user of this EPC is encouraged to consider the HSE fatigue risk index as a means of reviewing the maximum effect (HSE, 2006, http://www.hse.gov.uk/research/rrhtm/rr446.htm).

Tableau 2 : RARA- Effets des conditions de production d'erreur

On trouve également quelques exemples de calcul fait pour certaines opérations spécifiques :

Ouverture des portes par un conducteur de train	Détection d'un signal d'anomalie dans une tâche de contrôle en maintenance
Rate of human error (RSSB) • Driver opens doors 88 times per day • Incident review: – Action slips in door release – Underlying factors – equipment design and personal factors • 42 reported door release incidents at stations – Oct 2010-Mar 2013 • Doors were released 12,599,599 times at stations (based on timetable data) • 1 in 299, 990 times • Human error probability 3.33 x 10^{-6} • Plan to publish EHF 2015 • Helped company to evaluate the task	**Current task** (RSSB) • Task: Identify an 'intrusive' signal when presented on the display • Carried out throughout a shift. Automated manual and visual activities – waiting for a 'blip' which means that there is a crack in the axle – GTT R3 'Simple response to a dedicated alarm and execution of actions covered in procedures' - HEP = 0.0004 (1 in 2500) • Signals are rare – EPC – Unfamiliarity – Affect = 17 times worse • Signals are masked by noise – EPC - Low signal-noise ratio – Affect = 10 times worse • Environment is challenging – EPC – Environment – Affect = 4.5 times worse

As part of its Bow Tie-based Safety Risk Model (SRM), the RSSB also maintains a very comprehensive database of the relative frequency of accident precursors, including SPADS (Signal Passed at Danger). But (at least in the part of SPARK accessible to ordinary members without special rights) the frequency values are related to the activity (ex trains-km) and not to the number of exhibitions.

The MRS therefore does not allow, as it stands, the access conditions to specify the values of the previous tables, as it does not allow the denominators of the 'number of failures/number of exposures' fractions to be determined.

3 Application for Train Drivers (TD) and traffic Controlers (TC) SNCF

3.1 Datas

3.1.1 TD :

Pour les besoins d'étude, nous avons convenu des approximations suivantes pour les ADC, pour chaque thème :

1. **FSA + FSE** (Franchissement de signal d'arrêt – FSA, et franchissement de signal évité par un automatisme - FSE)

   12000 ADC, réalisant chacun 170 JS/An (journée de service par an)
   Nous faisons l'hypothèse qu'un ADC rencontre 2 signaux fermés/JS (en mode nominal, chaque train étant dans son sillon, tous les signaux sont normalement "ouverts", ce chiffre est donc « à dire d'expert »).
   Ceci conduit à 4 080 000 signaux fermés rencontrés.
   140 FSA/FSE en 2017 ==> taux de franchissement de **$3{,}5.\ 10^{-5}$ /SF**

2. **Omissions d'arrêt**
   13000 trains /jour (JOB), avec 5 arrêts en moyenne (à dire d'expert)
   Soit 65000 arrêts par jour et 16 900 000 arrêts annuels (260 jours)
   1000 omissions d'arrêt en 2017 ==> taux d'omission d'arrêt de **6.10^{-5}**/arrêt demandé

3. **Portes ouvertes TGV**
   2000 ADC Tgv avec 170 JS/an, avec 3 séquences ouv./ferm. portes par JS (à dire d'expert)
   Soit 255 000 manœuvres de portes/an.
   4 ESR en 2017 ==> taux d'erreur de **$1{,}5\ 10^{-5}$**/séquence d'ouv/ferm

   Nota 1 : Si on voulait faire une estimation un peu plus étayée, il faudrait trouver le nombre d'arrêts auprès de SNCF Réseau.

   Nota 2 : une étude du RSSB sur le sujet pour la période Oct 2010-Mars 2013 fait état de fréquences d'ouverture beaucoup plus élevées (88 manœuvres par jour et par conducteur en moyenne), et d'une fréquence d'erreur constatée nettement plus faible de 3.10^{-6}).

4. **DVL >15 (Dépassement de vitesse limite supérieur à 15km/h)**

13000 trains /jour pendant 260 jours soit 3 380 000 trains.
5 à 10 transitions de vitesse par train et par jour (à dire d'expert) soit entre 16 900 000 et 33 800 000 transitions de vitesse par an.
1000 (590 + 350) DVL>15 en 2017 ==> taux de DVL>15 **entre 3 et 6 10^{-5}** par transition.

Note: If we wanted to make a slightly more substantiated estimate, we would have to study some trains such as Paris Amiens, Paris Clermont, Paris Brive, TGV Strasbourg Bordeaux, count the number of transitions speeds decreasing -15, reduce this number to 100 km and then estimate that each of the 13,000 daily trains are 100 km long. This would result in a number of speed reductions that would be reduced to the number of DVL-15s

3.1.2 *TC* :

Les événements de défaut d'ordre concernent soit des ordres erronés soit des trains expédiés sans ordre. On dispose des chiffres suivants pour les défauts d'ordres en IdF :

Types d'Ordres	nbre d'ordres	**Risques causés par l'ESOP**			**Conséquence Ultime**
AVEN (Enrayages)	542	heurt nez à nez	heurt prise en écharpe	heurt Rattrapage	atteinte aux personnes (Décès)
BAPO (Baissé Panto)	65	arrachement de caténaire			Infrastructure endommagée - Mobil endommagé - Voie Impraticable
DERA (Reprise de Gardiennage PN)	156	Heurt voiture	heurt piéton		Atteinte aux personnes (Décès)
DIPU (Dérangement installation destinées aux Public)	11	heurt Voyageurs			Atteinte aux personnes (Décès)
OCAR (Ordre de Circuler avec Restriction)	**3646**	déraillement	heurt circulation	heurt voyageur ou tiers	Atteinte aux personnes (Décès)
Passe Partout	367	déraillement	heurt circulation	heurt piéton	Atteinte aux personnes (Décès)
RATO (Raté d'ouverture PN)	1694	Heurt voiture	heurt piéton		Atteinte aux personnes (Décès)
VAIG (Vérification Aiguille)	117	déraillement			Atteinte aux personnes (Décès)
VECA (Vérification Caténaire)	33	arrachement de caténaire			Infrastructure endommagée - Mobil endommagé - Voie Impraticable
VEFE (Vérification Feux routiers PN)	121	Heurt voiture	heurt piéton		Atteinte aux personnes (Décès)
VEVO (Vérification Voie)	50	déraillement			Atteinte aux personnes (Décès)
Total **ordres ci dessus** délivrés par EIC Parisiens en 2017	**6802**	ayant donné lieu à : **7 ESOP EN 2017 / 13 EN 2016 / 5 EN 2016 / 5 EN 2014 / 8 EN 2013**			
Total **tous types** d'ordres délivrés par 5 EIC IDF en 2017	**8237**				

En 2017 on a donc 6802 ordres délivrés par les EIC Parisiens et 7 ESOP, dont 2 ESR, soit un **taux d'erreur conduisant à l'ESR d'environ 3. 10^{-4} / ordre délivré**.

3.2 Discussion

A family of national-based figures is available for TDs that cover errors associated with different actions (respecting a stop signal, closing doors, modulating speed, etc.) and which are nevertheless remarkably similar. Values range from 1.5 to 6.10-5 "failures" per share, depending on the nature of the action. While they have been established by fairly crude approximations, and will need to be refined by more precise and robust methods, it is unlikely that they will be able to obtain scales that cause them to change their order of magnitude.

Subject to confirmation of current estimates, it can therefore be said that the reliability of TDs, on these actions considered critical to safety, revolves around (1-5.10-5), which constitutes, according to the literature review summarized in Section 2, a value that is in the order of magnitude of the attainable maximum, but with a potential margin of progression of about a factor 2.

However, it is unlikely that such progress will be achieved on a constant basis, by the sole work on the "human reliability" of TDs (or in conventional terms by working on their "behaviour"). The most significant areas of progress are rather linked to changes in the socio-technical system concerned, for example
:

- Targeted correction of specific cognitive failure modes related to operational "traps" (such as the existence of exceptions to high regularities, such as the interruption of the warning-signal signal sequence closed by a station stop).
- Re-design of HMIs to improve their ergonomics by making the presentation of in-trainingmores more intuitive and/or incorporating decision aids
- The addition of automatic catch-up loops

For CAs, the figures available for the number of orders issued seem to be more accurate. But they cover a smaller base (the IDF) and are therefore likely to be affected by possible regional variability. Most importantly, given the low numbers, the logic of selecting "failures" associated with an SRA can cause the result to vary significantly. However, again, it is highly unlikely that an out-of-court assessment will change the order of magnitude.

If we take the above calculated value of 3.10-4 as a failure rate in order of delivré, then we are on reliability levels of a lesser order of magnitude than that of the TDs. This in no way prejudges the relative position of TDs and ACs vis-à-vis their attainable optimum, because the trades are different, and the difference in numbers illustrates above all the great variability, already mentioned, of human reliability according to activity and operational and human contexts.

The CAs figures point to good levels of reliability for this type of activity, but it appears that significant progress is still possible. To go further in this slide, one would need to strengthen confidence in the value of reliability obtained. In particular, it is envisaged to systematically record the order ings by processing BS over 1 month, which will allow this sampling to be an annual assessment. Another review should focus on the characterization of the selected events to ensure consistency with the SRS/TDs.

Once the current reliability assessment has been consolidated and clarified, it will also be necessary to specify and make the reference (the realistic target) to which this value will have to be compared. Referring to Table 1 and thus the HEART/RARA method, we are in the case of R2 tasks, associated with an average error probability of 4.10-4 per share. We can see that the current estimated value (3.10-4) is slightly better. But the range of probability values indicated by the table is very broad (from 7.10-3 to 8.10-5), which at the same time confirms a theoretical potential for significant progress (up to 2 orders of magnitude), and introduces a high uncertainty about realistic expectations. In this case only a benchmark would really make it necessary to decide. The question was asked of the DSNA as part of this study with regard to air traffic controllers (and an answer is still expected), but the ideal would be to have a comparison on the same trade for example via the RSSB, or other rail players via the network of The UIC, OTIF, ERA, etc.

For this category of staff, however, it can be considered "expert" that there is a margin of progress of a factor of 2 or 3, which can be achieved by introducing in a volontarist way traditional techniques of "reliability of human activity" (self-control, cross control, communication 3 times, raising doubt, etc.), adapted to the activity

.

4 Conclusion

Ce premier survol rapide d'une approche quantitative de la fiabilité des actions critiques pour la sécurité effectuées par les ADC et les AC suggère des premières hypothèses:

- ADCs would be at a very efficient level, close to the maximum attainable in a constant system. Additional reliability efforts should therefore focus on an evolution of the sociotechnical system: improvement of HMIs, implementation of complementary automatic protections, and reduction of "risk exposure".
- CAs, on the other hand, would be at an 'average' level in the category of activity under consideration, and therefore likely to make more significant progress, including the implementation of so-called 'reliable human activity' (self-monitoring, cross-checking, 3-stroke communication, doubt ingeration, etc.) adapted to their business.

These assumptions need to be validated or invalidated by an improvement in the accuracy and reliability of the estimates made at this stage on the reliability found, estimates that are made here either at a large stroke or on an overly localized basis. Benchmarking is also required to refine benchmarks (reasonable targets).

In any case, the "measurement of exposure to these risks" will have to be consolidated, reliably, in order to be able to be permanently included in the tools of safety control, in a long time.

In addition, the dissemination of the items presented should allow for the establishment of other types of SREs on which to establish a constant/regular measure of risk exposure, for example on THE fields of ST, and on other trades.

5 References

[1] Kirwan, B. (1994) A guide to practical human reliability assessment. London: Taylor & Francis.

[2] Swain, A.D. and Guttmann, H.E. (1983) - A handbook of human reliability analysis with emphasis on nuclear power plant applications USNRC-Nureg/CR-1278, Washington DC-20555.

[3] Hollnagel, E. (1998) CREAM: Cognitive Reliability Error Analysis Method. Elsevier Science Ltd, Oxford

[4] Cooper, S.E., Ramey-Smith, A.M., Wreathall, J., Parry, G.W., Bley, D.C., Luckas, W.J., Taylor, J.H., and Barriere, M.T. (1996) A technique for human error analysis (ATHEANA) - Technical Basis and Method Description. Nureg/CR-6350. US Nuclear Regulatory Commission, Washington D.C. 20555.

[5] Le Bot, P. (2003) 'Methodological validation of MERMOS by 160 analyses' in Proceedings of the International Workshop Building the New HRA: Errors of Commission from Research to Application. NEA/CSNI/R(2002)3. OECD: Issy-les-Moulineaux.

[6] Williams, J.C., "HEART - A Proposed Method for Assessing and Reducing Human Error", Proceedings of the 9th "Advances in Reliability Technology" Symposium, University of Bradford, 1986.

[7] Gilroy, J. and Grimes, E. (2005) The development and application of a rail human reliability assessment tool. In Proceedings of the Second European Conference on Rail Human Factors, London, 21-23 November 2005.

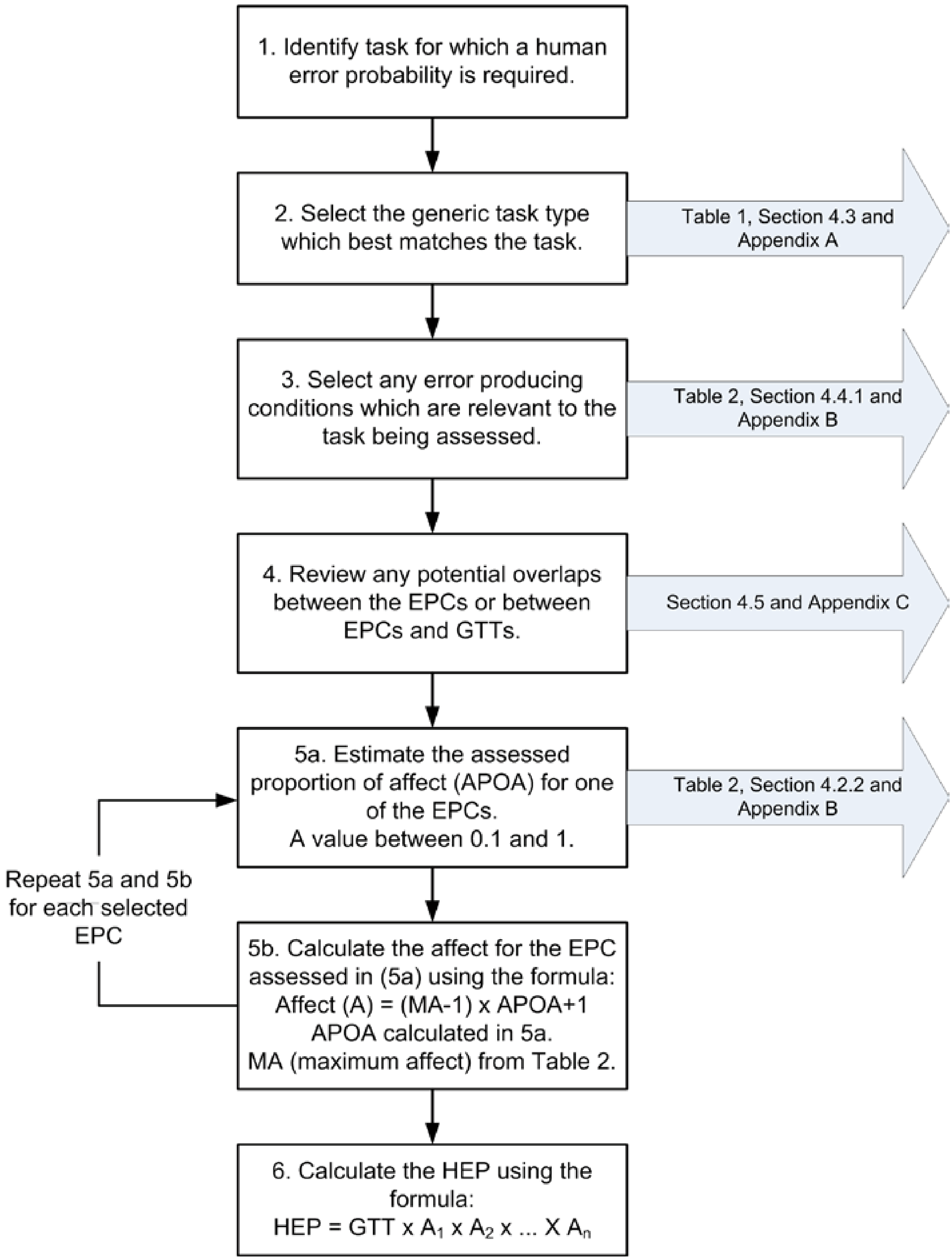
[8] Taylor-Adams, S.E., and Kirwan, B. (1995) 'Human Reliability Data Requirements' in International Journal of Quality and Reliability Management, 12, 1, pp 24 - 46.

[9] Gibson, H., Basra, G., and Kirwan, B. Development of the CORE-DATA database. Safety & Reliability Journal, Safety and Reliability Society, Manchester, 19, 1, 6 - 20, 1999.

[10] Gibson, W.H., and Megaw, E.D. (1999) 'The Implementation of CORE-DATA, a Computerised Human Error Probability Database' HSE contract research report 245/199, Suffolk: Health and Safety Executive.

[11] RSSB (2012) Development of a human error quantification technique for use in the railway industry Research. RSSB Brief T270 - July 2012

6 Appendix 1


1. Identify task for which a human error probability is required.
2. Select the generic task type which best matches the task.
Table 1, Section 4.3 and Appendix A
3. Select any error producing conditions which are relevant to the task being assessed.
Table 2, Section 4.4.1 and Appendix B
4. Review any potential overlaps between the EPCs or between EPCs and GTTs.
Section 4.5 and Appendix C
5a. Estimate the assessed proportion of affect (APOA) for one of the EPCs. A value between 0.1 and 1.
Table 2, Section 4.2.2 and Appendix B
Repeat 5a and 5b for each selected EPC
5b. Calculate the affect for the EPC assessed in (5a) using the formula: Affect (A) = (MA-1) x APOA+1 APOA calculated in 5a. MA (maximum affect) from Table 2.
6. Calculate the HEP using the formula: HEP = GTT x A_1 x A_2 x ... X A_n