Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27718-cd

Human Error Prediction with Reliability Life Data Analysis Compared to Classical Human Reliability Assessment: a Case Study in Freight Train Railways

Carvalho, Alessandra Lopes

Pontifical Catholic University of Minas Gerais, Brazil. E-mail: alessandralcarvalho@pucminas.br

Figueirôa Filho, Celso Luiz Santiago

GRAMS- Inovation, Reliability Research, Brazil. E-mail: celso@g-rams.com

Silva, Joselma Ramos

Pontifical Catholic University of Minas Gerais, Brazil. E-mail: joselma.r.silva@hotmail.com

Rocha, Gabriela Ferreira Baroni

Pontifical Catholic University of Minas Gerais, Brazil. E-mail: gabibaronirocha@gmail.com

The adoption of automation technologies in freight train operations has significantly reduced operational risks; however, maintenance tasks remain critical to railway safety. Consequently, identifying, addressing, and predicting human error is essential for improving railway reliability. This study proposes the use of Reliability Life Data Analysis (LDA) as an alternative approach for predicting human failure and compares its performance with the classical Human Reliability Assessment (HRA) method, SPAR-H, in the context of a large-scale railway logistics operation. A quantitative research approach was adopted using a case study methodology. Human failure events were stratified according to task type and Performance Shaping Factors (PSFs) and distributed along a timeline. The categorized events were then transformed into error-interval data for reliability analysis. In parallel, selected human errors were assessed using the SPAR-H method. The results obtained from both approaches were qualitatively compared, and human failures were analysed in relation to the incidence of selected PSFs. The analysis was based on a historical database of human failures recorded over one year, classified according to the PSFs defined in API Guidance 770. Results were presented using reliability and probability plot, highlighting PSFs and task types with the best goodness-of-fit. The findings indicate that LDA enables predictive insights for specific task categories, while SPAR-H provides static probability estimates adaptable to different contexts. As the methods differ in nature, they are complementary and can jointly support the identification of failure patterns and the development of more effective error reduction strategies.

Keywords: Human Reliability, Human Error, Freight Train Railways, Life Data Analysis, SPAR-H, API 770.

1. Introduction

According to Dhillon (2018), human reliability is defined as the probability that humans successfully accomplish a task at any required stage of system operations within stated minimum limits.

The field of human reliability emerged in the mid-1950s, when human factors began to be systematically and seriously considered, particularly within the context of the United States Air Force.

Human reliability can be analyzed using either qualitative or quantitative approaches. Qualitative modeling enables cause-and-effect analysis and supports organizational learning and knowledge development (Kirwan 1994). However, when historical data are available, quantitative modeling can substantially support the decision-making process, as it allows for the prediction of the probability of human failure occurrence.

Human Reliability in Railway is a long term implemented research process looking for better understanding of the relevant human factors and predicting methods (Aliza et al. 2025; Ciani et al. 2022; Dhillon 2007). Applications of human system interactions (HSI) on equipment automation (Yu, Shuai and Huang, 2024) and improving nuclear HRA methods (Catelani et al. 2021) are some examples of research themes in this sector. This work presents a case study focusing on the operation of freight trains, specifically a maintenance task—equipment inspection—an area that is rarely explored and lacks significant research.

Within this context, the present study aims to propose Reliability Life Data Analysis (LDA) as a potential approach for predicting human failure and to compare it with classical Human Reliability Assessment (HRA) methods in a large-scale railway logistics operation.

2. Literature review

2.1. Reliability Modeling

Traditional life data analysis involves modeling the random variable time to failure (TTF) using a probability density function $f(t)$. The same procedure can be applied to the modeling of any random variable.

As an example, the probability density function $f(t)$ of the Weibull distribution is presented. (see Eq. (1)). The parameters λ, β, γ, and $\eta$are the quantities to be estimated. Among the most commonly used parameter estimation methods are the Least Squares method and Maximum Likelihood Estimation (O'Connor and Kleyner 2012).

$$f(t) = \frac{\beta}{\eta^{\beta}}(t-\gamma)^{\beta-1} e^{\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right]} \quad (1)$$

Where:

β is the shape or slope parameter;

γ is the minimum life (location) parameter;

η is the characteristic life (scale) parameter.

Through goodness-of-fit tests, it is possible to select the most appropriate probability distribution to represent a given dataset. Once the probability density function $f(t)$is defined, the cumulative probability of failure function $F(t)$is given by Eq. (2). The reliability function $R(t)$is defined as the complement of the failure probability function $F(t)$.

$$F(t) = \int_0^t f(t)dt = 1 - R(t) \quad (2)$$

Where:

$F(t)$is the cumulative probability of failure function;

$f(t)$is the probability density function;

$R(t)$is the reliability function.

2.2. Human Reliability Assessment

Human reliability assessment involves the use of qualitative and quantitative methods aimed at characterizing the contribution of human error within a generic risk scenario. There are many different methods available, some of which were developed in a customized manner to meet the specific needs of particular industrial sectors. The human error probability (HEP) is simply defined by Eq. 3 (Kirwan 1994):

$$HEP = \frac{Number\ of\ errors\ occurred}{Number\ of\ opportunities\ for\ error} \quad (3)$$

Qualitative analysis is essential for identifying cause-and-effect relationships and, consequently, for supporting the development of effective action plans. Quantitative analysis enables the objective and structured estimation of the probability of human failure occurrence. Given the considerable difficulty in obtaining reliable quantitative data, the qualitative approach can contribute to understanding human interactions and cannot be omitted. Therefore, the two processes are complementary and jointly contribute to understanding sociotechnical interactions and predicting failures.

A significant evolution of methodologies for human reliability analysis has been observed, particularly from 1979 onward, motivated by the major nuclear accident at Three Mile Island. This accident caused profound changes in the nuclear industry in the United States, also impacting several other countries (Spurgin 2010). Following this event, significant changes occurred in the operation of nuclear power plants, especially in control room instrumentation, training, simulations, and emergency operating procedures.

The evolution of HRA has transitioned over the last several decades from a discipline rooted in engineering mathematics to one deeply integrated with cognitive psychology and organizational behaviour (Kirwan 1994). These methods are generally categorized into three "generations" based on their theoretical foundations and intended outcomes. Dougherty (1990) proposed this classification based on the level of model complexity, the manner in which human behavior is represented, and the degree of integration with the operational context.

In this work, the SPAR H (Standardized Plant Analysis Risk-Human Reliability Analysis). method was chosen for implementation (NUREG/CR-6883 2005), considering wide application in different economic sectors, and validated works (Elidolu et al. 2023).

2.3. The SPAR-H Method

The SPAR-H method is a framework developed by the U.S. Nuclear Regulatory Commission (NRC) in collaboration with Idaho National Laboratory (INL).

The method enables the quantification of HEP by distinguishing between two primary types of tasks: diagnosis and action. Action tasks are defined as tangible, physical activities involving the execution of a sequence of steps or manual operations. In contrast, diagnosis tasks are cognitive activities that rely primarily on the operator's knowledge, experience, and information processing capabilities rather than on physical execution. In the SPAR-H framework, nominal human error probabilities (NHEPs) of 0.01 and 0.001 are assigned to diagnosis and action tasks, respectively.

The nominal human error probability, whether associated with diagnosis or action, is adjusted through the application of Performance Shaping Factors (PSFs). PSFs are variables that directly influence human performance and may either increase or decrease the likelihood of human error (Boring, Blackman, 2007).

The SPAR-H method defines eight primary PSFs: (1) available time, (2) stress and stressors, (3) task complexity, (4) experience and training, (5) procedures, (6) ergonomics and human–machine interface, (7) fitness for duty, and (8) work processes.

The available time PSF refers to the amount of time available for the operator to complete a task. Insufficient time significantly increases the likelihood of errors and is therefore considered one of the most critical PSFs. The stress and stressors PSF evaluate physical and psychological conditions that may impair decision-making and task execution, including fatigue, excessive noise, time pressure, and unexpected events.

Task complexity reflects the intrinsic difficulty of the task. Tasks with high complexity require greater cognitive effort, which may degrade performance, particularly in the presence of ambiguity or multiple interdependent steps.

The experience and training PSF accounts for the operator's familiarity and proficiency with the task. Operators with adequate training and experience generally exhibit lower error probabilities, even under demanding conditions.
The quality of procedures is another key factor. Clear, accessible, and well-structured procedures significantly reduce the probability of human failure. Similarly, the ergonomics and human–machine interface PSF assesses how effectively systems and equipment support operator interaction. Poorly designed or unintuitive interfaces increase the likelihood of misinterpretation and erroneous actions.

Finally, the work processes PSF encompasses organizational aspects, including work organization, clarity of roles and responsibilities, coordination among teams, safety culture, and the broader organizational climate. Each PSF is evaluated qualitatively using a linguistic scale such as *nominal*, *inadequate*, or *better than nominal*, which are then translated into multipliers that modify the nominal human error probability. Levels such as *better than nominal*, *high*, or *good* (or equivalent descriptors) reduce the probability of error, whereas levels such as *inadequate*, *poor*, or *very poor* (or equivalent descriptors) increase the probability of error.

The PSF multipliers for Diagnostic Tasks (DT) and Action Tasks (AT) at various levels of influence are available at NUREG (2005). The estimation of the HEP for activities involving action (HEP_{AT}) or diagnosis (HEP_{DT}) is performed by multiplying the nominal value (NHEP) by the product of all PSFs, as shown in Eq. (4) and (5). The total human error probability (HEP_{total}) is calculated according to Eq. (6).

$$HEP_{AT} = NHEP_{AT} \times \prod PSF_{AT} \quad (4)$$

$$HEP_{DT} = NHEP_{DT} \times \prod PSF_{DT} \quad (5)$$

$$HEP_{total} = HEP_{AT} + HEP_{DT} \quad (6)$$

3. Methodology

This study was developed based on a quantitative research approach, using the case study method. The execution of the project was carried out in two main phases: a literature review and a case study. The process of conducting the case study was divided into four critical stages.

Stage 1: Data collection and preliminary calculations

The database structure was organized in accordance with API 770 (2001). With respect to the professional field of activity of the personnel involved, human errors were classified according to technical specialty, namely electrical, mechanical, and pneumatic. In this categorization process, the data is curated, identifying and removing inconsistencies and outliers.

PSFs were recorded considering three categories: internal factors, which act within the individual; external factors, which act upon the individual (environment, tasks, equipment, procedures, etc.); and psychological and physiological stress factors. The results are presented for better visualization in a dashboard screen

Stage 2: Reliability modeling

Once the failure modes of interest had been identified, the random variable *time to failure* was calculated.

A specific task must be selected for the analysis, and an initial frequency basis estimation for HEP using the events data.

Stage 3: HRA - Implementation of the SPAR-H method

At this stage, the SPAR-H method was implemented based on the analysis of the database and after the selection of a task.

Stage 4: Final analysis

In the final stage, the results obtained were analyzed and the corresponding conclusions were developed.

4. Case Study

Data related to human errors that occurred during maintenance tasks were collected from a freight train railroad company over a complete one-year period. Initially, the database was properly processed using Excel™, allowing the identification and removal of inconsistencies and outliers. Subsequently, Power BI™ was used to develop interactive charts to facilitate the visualization of the most relevant aspects of the dataset.

The company under study carries out annual, semiannual, and quarterly preventive maintenance plans, performed across a fleet of 299 locomotives. It was observed that most failures occur after the quarterly maintenance activities.

Across all analyzed categories a predominance of human errors associated with the electrical specialty was observed (Table 1).

Table 1. Categorization of human errors

Classification	Number of occurrences
Electrical specialty	79
Mechanical specialty	14
Pneumatic specialty	7

A specific task, "Electrical inspection", was selected. The resulting time values were used to model the reliability function, applying the concepts previously described. Python was employed as the computational tool during this stage.

5. Results

As shown in Table 1, the electrical specialty following quarterly preventive maintenance is the predominant human error, therefore, the analysis focused on human errors involving electrical tasks. For the case study, the database returned 39 occurrences, although all electric inspectors that perform this task have large experience.

Each quarterly maintenance plan comprises 33 tasks, performed across a fleet of 299 locomotives. Thus, the preliminary calculation of the human error probability was carried out using Eq. (3), resulting in a value of 0.0988%.

After obtaining the preliminary value of human error probability, the random variable representing the time between human failures was modeled based on the collected database. The maximum likelihood estimation (MLE) method was adopted due to its greater robustness. The selection of the probability distribution that best represents the data set was performed using the Kolmogorov–Smirnov (K–S) goodness-of-fit test and the calculation of the coefficient of determination (R^2). Python software was used throughout all stages of the modeling process.

The time between failures was calculated considering the date of the failure and the date of the most recent preventive maintenance (in days). Several probability distributions were tested, and the Weibull distribution provided the best fit, as indicated by the K–S test results and the R^2 coefficient. The estimated parameters were β (beta) = 0,8877 and η (eta) = 50301. Figures 1 and 2 present the probability plot and the reliability function, respectively. In both graphs, time was considered in days. The coefficient of determination (R^2) obtained was 0.9567.

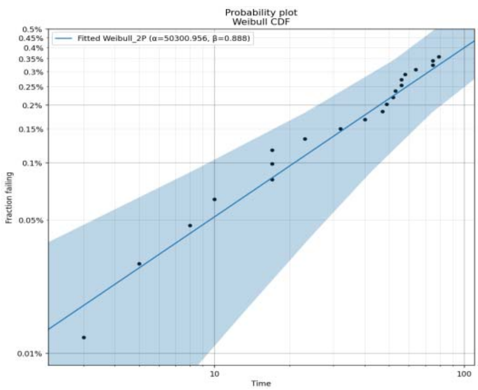

Fig. 1. Probabitity Plot for Weibull distribution

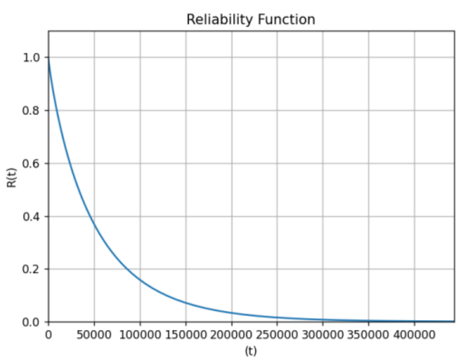

Fig. 2. Reliability Function (time in days)

Table 2 presents human reliability values calculated from the estimated function for the electrical inspection task.

Table 2. LDA Results

Time (days)	R(t)
7	0.9996
30	0.9986
90	0.9964
120	0.9953
150	0.9943

The quarterly maintenance plan consists of 33 tasks of varying complexity. These tasks were divided into two groups. Group 1 consists of 22 simple tasks involving only direct actions. Group 2 consists of 11 more complex tasks involving actions and diagnostic activities. Table 3 presents the application of the SPAR-H method to Group 1.

Table 3. PSF for Action Tasks (Group 1)

PSF	PSF Level	AC
Available Time	Nominal time	1
Stress/ Stressors	Nominal	1
Complexity	Obvious diagnosis	0.1
Experience/ Training	Nominal	1
Procedures	Nominal	1
Ergonomics/ HMI	Nominal	1
Fitness for Duty	Nominal	1
Work Processes	Poor	5

According to Table 3, the value obtained for the human error probability for Group 1 tasks is 0.0005.

Tables 4 and 5 present the results of the implementation of the SPAR-H method for Group 2, considering action tasks and diagnostic tasks, respectively.

Table 4. PSF for Action Tasks (Group 2)

PSF	PSF Level	AC
Available Time	Nominal time	1
Stress/ Stressors	Nominal	1
Complexity	Moderately complex	2
Experience/ Training	Nominal	1
Procedures	Available but poor	5
Ergonomics/ HMI	Nominal	1
Fitness for Duty	Nominal	1
Work Processes	Poor	5

Table 5. PSF for Diagnostic Tasks (Group 2)

PSF	PSF Level	DM
Available Time	Nominal time	1
Stress/ Stressors	Nominal	1
Complexity	Moderately complex	2
Experience/ Training	Nominal	1
Procedures	Available but poor	5
Ergonomics/ HMI	Nominal	1
Fitness for Duty	Nominal	1
Work Processes	Poor	2

According to Tables 4 and 5, the probability of human error for Group 2 tasks is 0.05 for action tasks and 0.2 for tasks involving diagnosis. The human error probability (total HEP) for the tasks in Group 2 is 0.25.

Thus, the total value of the human error probability, considering Group 1 (22 tasks) and Group 2 (11 tasks), is estimated as follows:

$$HEP_{total} = 0{,}0005\ x\ 0{,}66 + 0{,}25\ x\ 0{,}33$$

$$HEP_{total} = 0{,}08282 \approx 8\%$$

Table 6 presents a resume of the results for HEP estimation.

Table 6. Humam Error Probability results for quarterly maintenance locomotive electrical inspection .

Time (days)	HEP (%)
HEP by Eq. (3)	0.099
LDA (90 days)	0.360
SPAR-H	8.282

Table 6 indicates that the obtained HEP values have a wide range, more than 80 times between them. SPAR-H yields the highest value, as it does not account for potential error recovery, which represents a limitation of the technique. The LDA approach was calculated assuming a maintenance interval of three months and is strongly dependent on the quality of the root cause analyses in the original database.

The findings highlight that LDA analysis allows predictions for some task types. The classical approach using SPAR-H estimates static probability; however, it can simulate different contexts by modifying the PSFs. As the probabilities are different in nature, the methods can be complementary and have the potential to identify failure patterns.

The combination of the techniques can support better error reduction strategies.

6. Conclusions

This paper presented the potential of Reliability Life Data Analysis (LDA) approach to predict human failure. The results for a specific maintenance task in the context of a large-scale railway logistics operation shows that LDA can be useful to standard and routine task, and it can predict the human error probability for a maintenance plan.

First-generation human reliability analysis methods (e.g., THERP, HEART, SPAR-H) and second-generation methods (e.g., CREAM) provide a meaningful cause–effect analysis. However, these methods rely on generic values for human error probability across different contexts, weighted by performance shaping factors (PSFs). In contrast, Life Data Analysis (LDA) requires the use of context-specific data collected locally at the site where the study is conducted.

Classical HRA, as SPAR-H technique established a static HEP, and compared in a qualitative way to the LDA process. The findings highlight that LDA analysis allows predictions for some task types. The classical approach using SPAR-H estimates static probability; however, it can simulate different contexts by modifying the PSFs.

The quantitative approach adopted, using the case study method, results in different values. As the probabilities are different in nature, the methods can be complementary and have the potential to identify failure patterns. The combination can support better error reduction strategies.

The SPAR-H method enables a clear understanding of the task being performed, and the analysis of Performance Shaping Factors (PSFs) provides a basis for defining an action plan. Life Data Analysis (LDA), in turn, allows the visualization of how the probability of error evolves over time through a probabilistic distribution, enabling prediction based on historical data.

The process can be applied in others transport and industrial sectors for the kind of task, electrical equipment inspection tasks, which is a common job. Therefore, the process requires further applications to reach full maturity. However, the approach is sufficiently general to be applicable across distinct economic sectors.

References

Aliza, M. E. M. et al. (2025). Optimizing Railway Safety by Analyzing Human Reliability Techniques - A review. *Journal of Physics: Conference Series*, *2933*(1). https://doi.org/10.1088/1742-6596/2933/1/012014

American Petroleum Institute. *API 77 (2001). A Manager's Guide to Reducing Human Errors: Improving Human Performance in the Process Industries*. Washington, DC: American Petroleum Institute.

Boring, R. L.; H. S. Blackman. (2007). The origins of the SPAR-H method's performance shaping factor multipliers. *IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting*, 177–184. 2007. https://doi.org/10.1109/HFPP.2007.4413202

Catelani, M. et al . (2021). An enhanced SHERPA (E-SHERPA) method for human reliability analysis in railway engineering. *Reliability Engineering & System Safety*, *215*, 107866. https://doi.org/10.1016/J.RESS.2021.107866

Ciani, L., G. Guidi, G. Patrizi..(2022). Human reliability in railway engineering: Literature review and bibliometric analysis of the last two decades. *Safety Science*, *151.* https://doi.org/10.1016/j.ssci.2022.105755

Dhillon, B. S. (2007) *Human Reliability and Error in Transportation Systems.* Springer.

Dhillon, B. S. (2018). *Safety, Reliability, Human Factors, and Human Error in Nuclear Power Plants*. Boca Raton, FL: CRC Press.

Dougherty, E. M. (1990). Human reliability analysis – where shouldst thou turn?. *Reliability Engineering and System Safety*, 29(3), 283–299

Elidolu, G. et al. (2023). Applying evidential reasoning extended SPAR-H modelling to analyze human reliability on crude oil tanker cargo operation. *Safety Science*, *164.* https://doi.org/10.1016/j.ssci.2023.106169

Kirwan, Barry (1994). *A Guide to Practical Human Reliability Assessment*. London: Taylor & Francis.

O'Connor, P. D. T., A. Kleyner (2012). *Practical Reliability Engineering*. 5th ed. Chichester, England: John Wiley & Sons.

Spurgin, A. (2010). *Human Reliability Assessment: Theory and Practice*. New York: CRC Press.

United States Nuclear Regulatory Commission (U.S. NRC) (2005). *The SPAR-H Human Reliability Analysis Method*. NUREG/CR-6883. Washington, DC: U.S. Nuclear Regulatory Commission.

Yu, Y., B. Shuai, W. Huang (2024). Resilience evaluation of train control on-board system considering common cause failure: Based on a beta-factor and continuous-time bayesian network model. *Reliability Engineering and System Safety*, *246*. https://doi.org/10.1016/j.ress.2024.110088