applied sciences

MDPI

Article

An Integrated FMEA–HFACS–Bayesian Framework for Railway Risk Assessment: Specification, Survey-Informed Parameterisation and Demonstration

Ádám Papp [1,*] and István Lakatos [2]

[1] Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, Egyetem tér 1, 9026 Gyor, Hungary
[2] Department of Vehicle Maintenance and Diagnostics, Széchenyi István University, Egyetem tér 1, 9026 Gyor, Hungary
* Correspondence: papp.adam.sze@gmail.com

Abstract

Background: Integrating human factors into quantitative railway risk assessment remains methodologically unresolved. Failure Mode and Effects Analysis (FMEA) records human contributions as a single occurrence rating and cannot represent their organisational antecedents or their interactions. Purpose: This paper is a methodological proposal. It specifies the Integrated Human–Technical Risk Assessment (IHTRA) framework, which combines FMEA, the Human Factors Analysis and Classification System (HFACS) and Bayesian network modelling within the EN 50126 RAMS lifecycle, and demonstrates what such a specification makes analytically possible. It does not claim to validate the framework empirically. Methods: The Bayesian layer is specified in full as a ten-node network with all conditional probability tables reported. Of its twenty-two endogenous parameters, eight rest on evidence: four are derived from a survey of 89 Hungarian train drivers and four from the published fatigue literature. The remaining fourteen are declared structured assumptions awaiting expert elicitation. A conventional FMEA and the specified framework were both applied to national investigation report 2023-1152-5 (Sáp collision, 2023). Results: Human factors awareness yielded the highest domain mean (M = 4.23, SD = 1.18), with near-unanimous recognition of fatigue (M = 4.90) and workload (M = 4.87). Organisational consideration of human factors scored lowest (M = 2.66). No significant experience-group differences were observed ($p > 0.05$). The case analysis identified two HFACS levels as confirmed by the investigation findings and two further levels as plausible under the model interpretation. Inference over the specified network gives an illustrative, but not yet empirically calibrated, increase from $p = 0.00038$ to $p = 0.00059$ for the unsafe act and from $p = 0.00011$ to $p = 0.00047$ for the collision outcome. Conclusions: A specified but uncalibrated framework is a methodological contribution rather than an empirical one and is presented as such. This paper states precisely which parameters would have to be measured, and by what protocol, for the framework to become operational.

Keywords: railway safety; risk assessment; human factors; HFACS; Bayesian network; safety culture

Check for updates

Academic Editor: Suchao Xie

Received: 28 July 2026
Revised: 29 August 2026
Accepted: 7 September 2026
Published: 10 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Railway transportation plays a critical role in European mobility, and its safety record has direct implications for public welfare, regulatory compliance and operational efficiency.

Despite substantial investment in technical infrastructure and regulatory harmonisation under the European Union Agency for Railways (ERA) framework, human factors remain a persistent and insufficiently modelled contributor to railway incidents. ERA annual safety reports attribute a substantial proportion of significant accidents across European networks to human contributions [1].

The measurability of safety occupies a central position in this context. Safety is not directly observable: it must be inferred from a combination of incident rates, risk model outputs and safety culture indicators. The quantitative tools currently in use—principally FMEA and Fault Tree Analysis (FTA)—were designed for technical systems and have limited capacity to incorporate the dynamic, context-dependent nature of human behaviour [2]. This limitation is pronounced in the Central European context, where ageing infrastructure, a transitional regulatory environment and specific workforce characteristics produce risk profiles that differ from those of Western European systems. The 2023 collision at Sáp station in Hungary, investigated by the national investigation body, illustrates precisely this interaction of technical, organisational and human factors and forms the central case study of this paper.

This paper addresses the gap at two levels. Theoretically, it reviews the principal risk assessment methodologies and proposes a hybrid framework—the Integrated Human–Technical Risk Assessment (IHTRA) model—combining FMEA, HFACS and Bayesian network modelling within the EN 50126 RAMS lifecycle. Empirically, it presents original evidence from a structured survey of 89 Hungarian train drivers covering safety culture, regulatory compliance, human factor awareness and general safety perception. These findings are mapped onto the HFACS taxonomy, applied retrospectively to the Sáp case and used to identify calibration parameters for the proposed framework.

This work forms part of a broader doctoral research programme on risk-based approaches to railway safety improvement, with a particular focus on human factors in Central Europe. This study addresses the following research questions:

1. RQ1: What are the dominant human factor perceptions among Hungarian train drivers?
2. RQ2: Does professional experience moderate safety culture perceptions?
3. RQ3: How do the empirical findings map onto the HFACS taxonomy?
4. RQ4: What analytical outputs does the specified framework generate when applied to a documented railway accident, and which of them are unavailable from a conventional FMEA of the same configuration?

A note on the status of this work is necessary at the outset. This paper is a methodological proposal. It specifies a framework and demonstrates what the specification makes analytically possible; it does not claim to validate the framework empirically. Of the twenty-two endogenous parameters of the Bayesian layer, eight rest on evidence—four derived from the survey and four from the published literature—and fourteen are declared structured assumptions. At the node level, this corresponds to three nodes parameterised from the survey, one from the literature and four from assumption, as stated in the note to Table 1. Every probability reported in Sections 6.4 and 7 is therefore illustrative and conditional on those assumptions, and none should be read as an estimate of risk in Hungarian railway operations. Section 8 states precisely which parameters would have to be measured, and by what protocol, for the framework to become operational.

The remainder of this paper is organised as follows. Section 2 reviews the theoretical background, Section 3 introduces the proposed framework, Section 4 reports the empirical study, Section 5 maps the survey results onto the HFACS taxonomy, Section 6 presents the case study, Section 7 discusses the findings and Section 8 concludes.

2. Theoretical Background and Literature Review

This section establishes the theoretical foundations on which the proposed framework rests. It first examines the epistemological problem of measuring safety in sociotechnical systems (Section 2.1), then evaluates the two principal method families relevant to this study, namely FMEA (Section 2.2) and HFACS (Section 2.3). Section 2.4 considers Bayesian approaches to railway risk modelling, and Section 2.5 identifies the research gaps that motivate the present work.

2.1. Safety Measurability in Sociotechnical Systems

Measuring safety in complex sociotechnical systems presents a fundamental epistemological challenge: safety is characterised by the absence of harm rather than by its presence, which makes direct measurement inherently difficult. Hollnagel et al. distinguish between Safety-I, understood as the absence of negative outcomes, and Safety-II, understood as the presence of positive adaptive capacity, and argue that a robust measurement framework must capture both dimensions [3]. In railway systems, safety is typically operationalised through precursor indicators, accident rates per train-kilometre and risk priority numbers derived from structured assessments.

Reason's model of organisational accidents remains foundational: accidents occur when multiple defensive barriers—technical, procedural and human—are simultaneously breached [4]. This multi-barrier view has direct implications for measurability, since a meaningful safety metric must capture the integrity of all layers rather than the technical layer alone. The RAMS framework defined in EN 50126 provides the regulatory structure for technical safety quantification in railway systems, but its treatment of human factors is limited and largely qualitative [5].

2.2. FMEA and Its Limitations for Human Factor Integration

FMEA systematically identifies failure modes, their effects and their risk priority numbers (RPNs), calculated as the product of occurrence, severity and detectability. Its adoption in railway engineering is codified in EN 50126 and applied to subsystems ranging from signalling to rolling stock [2]. The method has well-documented limitations in human factor modelling: the human element is typically represented as a single failure mode entry rather than as a structured causal layer. Wang et al. observe that dynamic environmental and behavioural factors cannot be adequately captured by static RPN scores [6]. Extended variants incorporating fuzzy logic and prospect theory have been proposed [7], but none fully integrates a structured human taxonomy.

The Sáp 2023 case illustrates this limitation directly. A conventional FMEA of the signalling subsystem would assign a high RPN to the absence of Automatic Train Protection (ATP) but would record the driver's signal violation as a single undifferentiated human error entry, obscuring both the night-shift fatigue precondition and the organisational decision to defer ATP installation.

2.3. HFACS and Its Application in Railway Safety

The Human Factors Analysis and Classification System, developed by Shappell and Wiegmann from Reason's theoretical framework, provides a four-level taxonomy of human failure: Level 1, unsafe acts comprising errors and violations; Level 2, preconditions for unsafe acts including adverse mental states and physical limitations; Level 3, unsafe supervision; and Level 4, organisational influences [8].

The application of HFACS to railway safety was pioneered by Reinach and Viale, who adapted the framework for train accident investigation [9]. Zhan et al. subsequently developed a railway-specific adaptation, demonstrating the prevalence of organisational

and supervisory factors in Chinese railway incidents [10]. Gawlak applied HFACS to 148 final reports from the Polish network, examining the age and experience of the personnel involved [11]. These studies establish HFACS as the most structurally rigorous human factor tool available for railway safety analysis, while revealing a consistent gap: no study has quantitatively integrated HFACS with FMEA within a Bayesian probabilistic framework calibrated on empirical survey data.

The application of HFACS has since extended beyond accident investigation. Guo and colleagues analysed 4095 proactive safety inspection records from a large dispatching centre, a source that documents minor discrepancies which never reach accident reports, and established through path analysis a significant relationship between unsafe supervision, preconditions and unsafe acts [12]. Gawlak applied the framework to 148 final reports from the Polish network and noted that the literature concentrates on signals passed at danger while neglecting events arising from an improperly prepared route [11]—a category directly relevant to the case examined in Section 6. Comparable analyses in Australia identified adverse mental states as the most frequent precondition category [13]. These studies code observable factors recorded by investigators; the present study uses perception items, which measure a different construct, as discussed in Section 4.1.

2.4. Bayesian Networks in Railway Safety Modelling

Bayesian networks are probabilistic graphical models that represent conditional dependencies through directed acyclic graphs and conditional probability tables. Their capacity for dynamic, context-sensitive probability updating makes them well suited to modelling human–technical interaction in railway systems [14]. Shi et al. applied a data-driven Bayesian network to railway accident risk factors and identified track damage detection failures and driver skill deficits as critical nodes [15]. The principal advantage of Bayesian networks over FMEA lies in their ability to represent non-linear dependencies and to update risk probabilities as contextual variables change—a capability that static FMEA fundamentally lacks.

More recent work has extended Bayesian modelling in the railway domain in two directions. Liu and colleagues combined knowledge metatheory with dynamic Bayesian networks to trace how railway emergencies evolve after onset [16], while Li and colleagues applied weighted Bayesian networks to the structural safety of high-speed rail infrastructure, explicitly addressing the conditional independence assumption of standard models [17]. Neither models the pre-accident human and organisational causal chain within a HFACS structure.

The pairing of HFACS with Bayesian modelling is not itself novel. Bhuiyan and colleagues applied precisely this combination to railway accidents [18], and comparable hybrids exist in other safety-critical domains. The contribution claimed here is correspondingly narrower: the addition of an FMEA layer beneath the HFACS structure, the calibration of selected nodes against original survey data, and the application of the resulting three-layer model to a documented national investigation report.

2.5. Research Gaps and the Central European Context

A review by Ciani et al. covering human reliability analysis in railway engineering from 2000 to August 2021, based on 268 journal articles retrieved from Scopus, documents a marked increase in research interest while showing that the approaches applied remain fragmented across the field [19]. The Common Safety Method Regulation (EU) No 402/2013 requires structured risk assessment but prescribes no universal methodology, which has led to methodological fragmentation across European operators [20]. Most published

HFACS-railway studies originate from China, Australia and the United Kingdom; Central European railway systems remain substantially under-represented.

The claim that regional context matters is not merely assumed here. The European Union Agency for Railways reports that behind the aggregate European picture lies a considerably more diverse reality, with notably large differences in casualty rates between Member States [21], and attributes these differences in part to cultural background, the extent of line fencing and urbanisation patterns, while noting that divergent national reporting practices remain a significant confounder [22]. Comparative frameworks have been developed to benchmark safety performance across European networks on a common index [23]. Together these provide the empirical basis for treating the Central European operating context as requiring separate calibration rather than the adoption of Western European parameters.

Recent hybrid approaches demonstrate the integration potential. Liou et al. combined FMEA and HFACS for inter-city bus accident risk assessment [24], and transport safety research has combined fatigue modelling with detection methods for train drivers [25]. However, no published study applies the three-layer integration of FMEA, HFACS and Bayesian networks in a railway context calibrated with empirical survey data and illustrated against a real accident investigation report. The UIC 2024 safety report confirms that signal passed at danger (SPAD) events—the category to which the Sáp collision belongs—are driven primarily by workload, procedural and communication factors [26], precisely the HFACS Level 2 and Level 3 variables targeted by the proposed framework. This convergence of literature gaps and operational evidence constitutes the motivation for the present research.

3. The Proposed IHTRA Framework

This section introduces the Integrated Human–Technical Risk Assessment framework. Section 3.1 describes its three-layer architecture, illustrated in Figure 1, and Section 3.2 positions the framework within the EN 50126 RAMS lifecycle and the European regulatory environment.

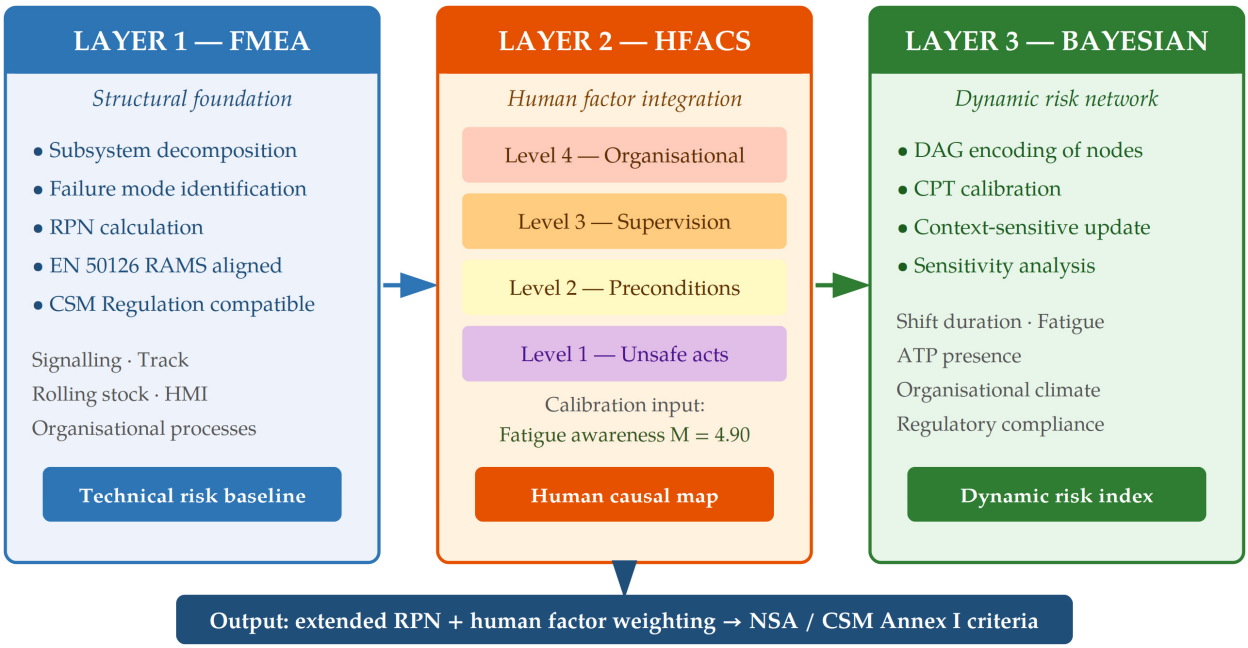


Figure 1. IHTRA framework architecture: three-layer integration within the EN 50126 RAMS lifecycle and the EU Common Safety Method regulatory structure.

3.1. Framework Architecture

The IHTRA model combines three methodological layers within the EN 50126 RAMS lifecycle. The overall architecture is presented in Figure 1.

Layer 1 provides the FMEA structural foundation. The railway system is decomposed into functional subsystems—signalling, track infrastructure, rolling stock, the human–machine interface and organisational processes. For each subsystem, failure modes are identified, their effects catalogued and initial RPN scores calculated. This establishes a technical risk baseline consistent with existing engineering practice and with the regulatory requirements of the Common Safety Method framework.

Layer 2 provides HFACS human factor integration. For each failure mode with a human causal or contributory pathway, a HFACS classification is applied across all four levels. This transforms the binary human error entry of conventional FMEA into a structured, multi-level causal map. Organisational and supervisory factors identified at Levels 3 and 4 become explicit model nodes rather than implicit background assumptions, which represents the principal departure from conventional practice.

Layer 3 provides Bayesian dynamic modelling. The HFACS-enriched FMEA entries are encoded as nodes in a Bayesian network. Conditional probability tables are calibrated using historical incident data and, as demonstrated in this paper, survey-derived empirical data on human factor perception. This layer enables context-sensitive probability estimation, adjusting human error likelihood for shift duration, fatigue level, organisational climate and regulatory compliance culture. The model output is a dynamic risk prioritisation index that extends the conventional RPN with human factor weighting and conditional contextual modifiers.

3.2. Bayesian Network Specification

The third layer is implemented as a discrete Bayesian network of ten nodes and ten directed edges. Its structure is given in Figure 2 and its node set in Table 1. The network is deliberately compact: every node corresponds to a construct for which either empirical evidence or a defensible expert judgement exists, and no node has been introduced merely to increase resolution.

Table 1. Specification of the Bayesian network nodes.

Node	Meaning	States	Parents	CPT Source
OrgHF	Organisational human factor quality	High/Low	—	Survey, n = 89
RuleEnf	Rule enforceability	High/Low	—	Survey, n = 89
SafetyInfo	Safety information flow	Adequate/Inadequate	OrgHF	Survey, n = 89
FatMgmt	Fatigue management in place	Present/Absent	OrgHF	Assumption
Fatigue	Driver fatigue	Low/High	Shift, FatMgmt	Literature
Lapse	Attention lapse	No/Yes	Fatigue, SafetyInfo	Assumption
SPAD	Signal passed at danger	No/Yes	Lapse, RuleEnf	Assumption
Collision	Collision outcome	No/Yes	SPAD, ATP	Assumption
Shift	Shift type	Day/Night > 8 h	—	Exogenous
ATP	Train protection fitted	Installed/Absent	—	Exogenous

Three of the eight endogenous nodes carry priors derived from the survey; one is parameterised from the published literature; four are structured assumptions requiring expert elicitation. Shift type and train protection are exogenous scenario variables.

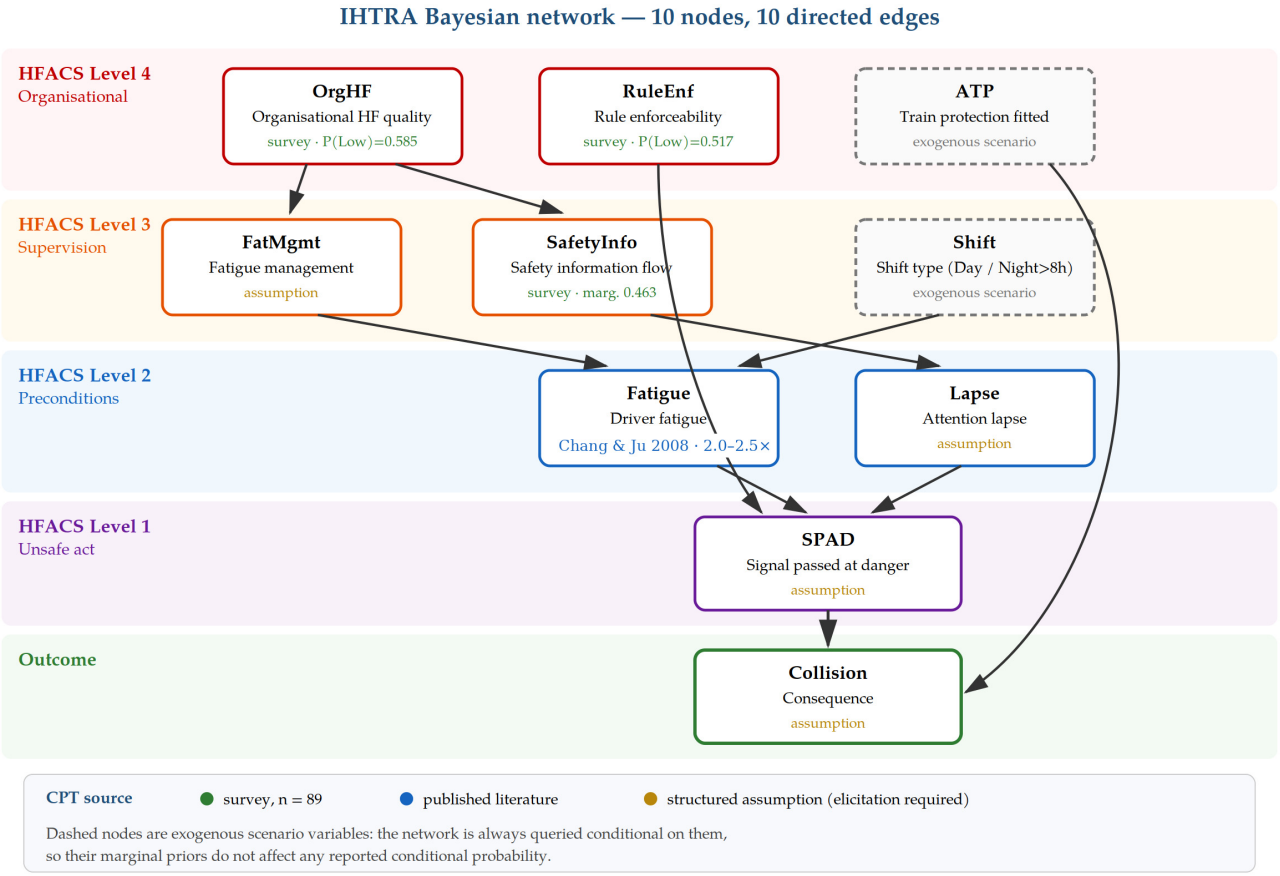


Figure 2. Specification of the Bayesian network. Ten nodes are arranged by HFACS level. Dashed nodes are exogenous scenario variables. Colour indicates the evidential source of each conditional probability table. Arrows indicate the direction of causal influence between HFACS levels Chang and Ju [27] and Härmä et al. [28] provide the fatigue multiplier shown in the figure.

Nodes are arranged by HFACS level. Organisational human factor quality and rule enforceability occupy Level 4; fatigue management and safety information flow occupy Level 3; driver fatigue and attention lapse occupy Level 2; the signal passed at danger occupies Level 1; and the collision outcome terminates the chain. Shift type and the presence of Automatic Train Protection are treated as exogenous scenario variables: the network is always queried conditional on them, so their marginal priors do not affect any conditional probability reported in this paper.

One structural decision requires comment. Automatic Train Protection does not prevent a driver from passing a signal at danger; it constrains the consequence once the signal has been passed. It is therefore specified as a parent of the collision node and not of the unsafe act. This separates error causation from consequence mitigation, and the distinction is consequential: the two quantities respond differently to the same intervention, as Section 6.4 shows.

The evidential basis of the network is heterogeneous and is therefore reported parameter by parameter in Appendix A as well as node by node in Table 1. Of the twenty-two entries in Appendix A, four derive from the survey, four from the published literature and fourteen are structured assumptions; at the node level, this corresponds to three survey nodes, one literature node and four assumption nodes. Survey responses inform exactly three nodes. Organisational human factor quality and rule enforceability carry priors obtained from the corresponding item means through the linear transformation $P(\text{unfavourable}) = (5 - M)/4$, which maps the Likert midpoint to a probability of one half,

yielding 0.585 and 0.517, respectively. The conditional table for safety information flow is set so that its marginal reproduces the survey-derived value of 0.463.

The fatigue node is parameterised from the published literature rather than from the present sample. Chang and Ju report that accident risk approximately doubles after four hours of continuous driving in freight operations [27], and Härmä and colleagues report that fatigue risk rises by approximately fifteen per cent for each additional hour of shift time [28]. The driver in the case examined in Section 6 was nine hours and twenty minutes into a night shift, so a multiplier in the range 2.0 to 2.5 is a conservative reading of both sources.

The four remaining conditional tables, comprising fourteen of the twenty-two parameters, are structured assumptions. They are reported in full in Appendix A, are identified as assumptions in Figure 2 and Table 1, and are the subject of the sensitivity analysis in Section 6.4. They are not described as calibrated, because no data are currently available against which to calibrate them.

3.3. Integration with the Regulatory Framework

The IHTRA model is positioned within the Hazard Identification and Risk Assessment phases of the EN 50126 RAMS lifecycle and is designed to be compatible with the risk evaluation process of the Common Safety Method Regulation. The probabilistic outputs of the Bayesian layer can be mapped directly onto the risk acceptance criteria specified in Annex I of that regulation. This compatibility ensures that the enhanced human factor modelling does not require operators to abandon existing compliance workflows but instead enriches them with additional causal depth. In Hungary, the National Safety Authority supervises compliance with these requirements, and the outputs of the model are structured to support authority-level reporting and audit.

4. Materials and Methods

This section reports the empirical study that provides the calibration basis for the framework. Section 4.1 describes the study design and the development of the instrument, Section 4.2 characterises the sample, and Sections 4.3–4.6 present the reliability analysis, item-level results, experience-group comparisons and inter-domain correlations, respectively.

4.1. Study Design and Instrument

A structured cross-sectional survey was administered to active train drivers in the Hungarian railway system in March 2026. The instrument comprised 26 Likert-scale items (1 = strongly disagree, 5 = strongly agree) across four thematic domains: organisational safety culture (9 items, corresponding to HFACS Levels 3 and 4); rules and regulations (5 items); human factors awareness (5 items, corresponding to HFACS Level 2); and general safety perception (7 items). Respondents additionally reported their years of professional experience.

Instrument development followed a four-step process. First, a literature review identified the established dimensions of railway safety culture measurement, drawing on the Nordic Safety Climate Questionnaire framework [29] and on HFACS-based railway research. Second, the four thematic domains were matched to key constructs from the literature in order to establish content validity. Third, items were mapped onto the HFACS four-level taxonomy so that each domain operationalises a corresponding tier. Fourth, five railway safety experts reviewed item wording for clarity and contextual fit, and three items were revised accordingly.

The survey was distributed using a snowball sampling strategy, with initial contact made through trade union channels and professional railway safety networks and participants invited to forward the questionnaire voluntarily [30]. This approach is suited to closed professional communities where a complete population register is not accessible. The exact number of active train drivers in Hungary is not publicly available, since MÁV-START, GYSEV and private operators collectively employ train drivers without a consolidated national register. The MÁV Group employed approximately 37,000 staff in 2023 [31], of whom train drivers are estimated to constitute five to eight per cent, giving an estimated sampling fraction of three to five per cent. The sample is therefore treated as exploratory rather than representative: it is suitable for hypothesis generation and for prior parameter specification, but population-level generalisation requires further validation.

Statistical analysis comprised descriptive statistics, Cronbach's alpha for internal consistency, exploratory factor analysis with the Kaiser–Meyer–Olkin measure and Bartlett's test of sphericity, Kruskal–Wallis non-parametric tests for experience-group comparison, and Pearson and Spearman correlation analysis. Participation was voluntary and anonymous, and no personally identifying information was collected.

4.2. Sample

Eighty-nine train drivers completed the survey. The distribution by professional experience is presented in Table 2.

Table 2. Sample distribution by professional experience (n = 89).

Experience Group	n	Percentage
0–5 years	37	41.6
6–10 years	26	29.2
11–20 years	5	5.6
Over 20 years	21	23.6
Total	**89**	**100.0**

Bold indicates the scenario profile discussed in the accompanying text.

The sample is dominated by relatively inexperienced drivers (0–5 years: 41.6%) and mid-career drivers (6–10 years: 29.2%), with 23.6% reporting more than 20 years of experience. The 11–20 year cohort is substantially under-represented (n = 5), which limits the statistical power of comparisons involving this group and is acknowledged as a limitation in Section 7.4.

4.3. Reliability Analysis and Exploratory Factor Analysis

Internal consistency was assessed using Cronbach's alpha for each thematic domain. The results are presented in Table 3.

Table 3. Domain-level descriptive statistics and internal consistency.

Thematic Domain	Items (n)	Mean (SD)	Cronbach's α
Organisational safety culture	9	3.40 (1.21)	0.899 (excellent)
Rules and regulations	5	3.51 (1.19)	0.689 (acceptable)
Human factors awareness	5	4.23 (1.18)	0.290 (see Section 4.3)
General safety perception	7	3.37 (1.16)	0.862 (good)

Organisational safety culture ($\alpha = 0.899$) and general safety perception ($\alpha = 0.862$) demonstrate excellent and good internal consistency, respectively, confirming their validity as composite scales. Rules and regulations show acceptable reliability ($\alpha = 0.689$). The

human factors awareness domain yields a low alpha ($\alpha = 0.290$), which requires methodological interpretation.

To investigate this, an exploratory factor analysis was conducted on the five human factors items. The Kaiser–Meyer–Olkin measure of sampling adequacy was 0.627, above the 0.60 threshold for factoring, and Bartlett's test of sphericity was significant (χ^2 (10) = 56.937, $p < 0.001$), confirming that factor analysis was appropriate. The analysis showed that the item measuring organisational consideration of human factors (M = 2.66) loads only weakly on the extracted factor (loading = 0.138) and correlates negligibly with the four-item individual awareness subscale (r = −0.066). With that item removed, the remaining four items form a coherent subscale (KMO = 0.629; χ^2 (6) = 54.603, $p < 0.001$; $\alpha = 0.514$).

This pattern is consistent with a bi-dimensional reading of the domain, in which items one to four would measure individual human factor risk awareness, corresponding to HFACS Level 2, while the fifth item would measure the perceived organisational response, corresponding to HFACS Level 4. We emphasise that this interpretation was developed after inspecting the data and was not preregistered. On a five-item scale with n = 89 and a Kaiser–Meyer–Olkin value of 0.627, which is only marginally acceptable, it cannot be distinguished from a statistical artefact arising from a weak instrument. The two-factor structure is therefore treated throughout this paper as a hypothesis requiring confirmatory testing on an independent sample and not as a finding.

4.4. Item-Level Results

The mean and standard deviation for all 26 survey items are presented in Table 4, ordered by descending means.

Table 4. Item-level descriptive statistics (n = 89), ordered by descending means.

Survey Item	Mean	SD	Domain
Fatigue can affect performance	**4.90**	0.34	HF
Workload can affect safety	**4.87**	0.40	HF
Stress can affect decision-making	**4.61**	0.79	HF
Attention is difficult to maintain on long shifts	4.10	1.13	HF
Safety rules are more important than punctuality	4.03	1.24	RR
Rules contribute to safety	3.98	1.03	RR
Technical systems support safety	3.88	1.00	GP
Employees can freely report safety risks	3.87	1.21	OC
I generally feel safe at work	3.74	1.01	GP
Most employees comply with regulations	3.67	0.95	RR
I feel safe during work operations	3.61	1.15	GP
Consequences of rule violations are clear	3.60	1.28	RR
Management prioritises safety	3.56	1.07	OC
Overall railway safety level is adequate	3.36	1.38	GP
The organisation learns from past incidents	3.36	1.09	OC
Railway rules are clear	3.35	1.22	RR
The organisation supports safety observations	3.25	1.15	OC
Safety incident experience is shared	3.24	1.23	OC
Managers respond appropriately to safety problems	3.22	1.03	OC
Training improves safety	3.18	1.15	GP

Table 4. *Cont.*

Survey Item	Mean	SD	Domain
Safety information reaches drivers on time	3.16	1.04	OC
The organisation continuously improves systems	**2.94**	1.16	GP
Management encourages safety improvements	**2.93**	1.18	OC
Rules are practically enforceable	**2.93**	1.19	RR
Employees participate in safety development	**2.85**	1.14	GP
The organisation considers human factors	**2.66**	1.18	OC

Bold values indicate means $\geq$ 4.50 or $\leq$2.99. HF = human factors awareness; RR = rules and regulations; OC = organisational safety culture; GP = general safety perception.

The five highest-scoring items belong exclusively to the human factors domain, with fatigue (M = 4.90, SD = 0.34) and workload (M = 4.87, SD = 0.40) approaching ceiling values. The low standard deviations confirm the consistency of this perception across the sample. The pattern demonstrates that train drivers possess a high level of individual human factor risk awareness, which supports the HFACS Level 2 component of the proposed model.

The five lowest-scoring items are concentrated in the organisational and rules domains: organisational consideration of human factors (M = 2.66), employee participation in safety development (M = 2.85), management encouragement of safety improvements (M = 2.93), practical rule enforceability (M = 2.93) and continuous system improvement (M = 2.94). These scores reveal a consistent pattern of perceived organisational deficit corresponding to HFACS Levels 3 and 4. The gap between individual awareness and organisational consideration is the most striking finding of the survey and directly motivates the HFACS–FMEA integration proposed in Section 3.

4.5. Domain-Level Analysis and Experience Group Comparisons

Domain means by experience group are presented in Table 5. The Kruskal–Wallis test was selected as the non-parametric analogue of one-way analysis of variance, appropriate given the ordinal nature of Likert data and the unequal group sizes.

Table 5. Domain means by experience group.

Domain	0–5 yrs (n = 37)	6–10 yrs (n = 26)	11–20 yrs (n = 5)	20+ yrs (n = 21)
Organisational safety culture	3.38	3.34	3.36	3.52
Rules and regulations	3.61	3.47	2.88	3.51
Human factors awareness	4.30	4.08	3.88	4.36
General safety perception	3.45	3.38	2.80	3.33

Kruskal–Wallis test statistics: organisational safety culture H = 0.60, p = 0.896; rules and regulations H = 2.41, p = 0.492; human factors awareness H = 7.15, p = 0.067; general safety perception H = 2.75, p = 0.432.

No statistically significant differences across experience groups were found for any domain (all $p > 0.05$). The nominally lowest value for human factors awareness occurs in the 11–20 year group, but that group contains five respondents. With n = 5, a single atypical response shifts the group mean by more than one scale point, so no inference is drawn from it and no substantive interpretation is offered. The group is descriptively reported in Table 5 for completeness only, and the Kruskal–Wallis statistics should be read with this in mind.

Spearman rank correlations between domain scores and ordinal experience confirmed no significant associations (all $r < 0.10$, $p > 0.45$). The absence of significant experience-group

differences is itself informative: perceptions of organisational safety culture, regulatory compliance and general safety appear to be substantially determined by shared organisational context rather than by individual career trajectory, which elevates the importance of organisational-level variables in the proposed model.

4.6. Inter-Domain Correlation Analysis

Pearson inter-domain correlations are presented in Table 6.

Table 6. Pearson inter-domain correlation matrix.

Domain	Org. Culture	Rules	Human Factors	Perception
Organisational culture	1.000	0.500 **	0.254 *	0.739 **
Rules and regulations	0.500 **	1.000	0.209 *	0.610 **
Human factors awareness	0.254 *	0.209 *	1.000	0.163
General safety perception	0.739 **	0.610 **	0.163	1.000

** $p < 0.01$; * $p < 0.05$.

The strongest correlation is between organisational safety culture and general safety perception (r = 0.739, $p < 0.01$), indicating that drivers who perceive their organisation as safety-committed also rate their working environment as safer overall. Rules and regulations strongly correlate with general safety perception (r = 0.610, $p < 0.01$) and moderately with organisational culture (r = 0.500, $p < 0.01$). The human factors domain shows only weak correlations with all other domains (r = 0.163–0.254). This relative independence confirms that human factor risk awareness operates as a distinct cognitive dimension, largely uncoupled from organisational culture perception, and justifies treating it as an independent node in the Bayesian layer of the model.

5. Mapping Survey Results to the HFACS Taxonomy

This section links the survey findings of Section 4 to the HFACS taxonomy. The mapping serves two purposes: it supplies the three survey-derived priors used in the network specified in Section 3.2, and it locates the remaining constructs within the HFACS structure so that the parameters still requiring elicitation can be identified. It is a structuring device, not a test of the framework. The mapping is presented in Table 7.

Table 7. Mapping of survey findings onto HFACS levels.

HFACS Level	HFACS Category	Corresponding Survey Item	Mean
Level 4—Organisational	Resource management/climate	Management encourages safety improvements	**2.93**
Level 4—Organisational	Organisational process	The organisation considers human factors	**2.66**
Level 3—Supervision	Inadequate supervision	Safety information reaches drivers on time	3.16
Level 3—Supervision	Failure to correct known problem	Managers respond appropriately to safety problems	3.22
Level 2—Preconditions	Adverse mental states	Fatigue can affect performance	**4.90**
Level 2—Preconditions	Physical/mental limitations	Stress can affect decision-making	**4.61**
Level 1—Unsafe acts	Errors/violations	Rules are practically enforceable	**2.93**

Bold values indicate critical gaps (mean $\leq$ 2.99) or domain strengths (mean $\geq$ 4.50).

The mapping reveals a bifurcated risk profile. At Level 2, drivers demonstrate high awareness of fatigue, workload and stress as safety-critical variables, which are cognitively salient to the workforce. At Levels 3 and 4, by contrast, multiple critical gaps are identified: management does not adequately encourage safety improvements (M = 2.93), the organisa-

tion does not systematically consider human factors (M = 2.66), and rules are not perceived as practically enforceable (M = 2.93).

This profile has direct implications for model calibration. The high Level 2 awareness scores suggest that precondition variables should carry high prior probabilities in the Bayesian nodes corresponding to human error events. The low Level 3 and Level 4 scores indicate that organisational and supervisory barriers are weak, substantially reducing the capacity of the system to detect and correct developing unsafe conditions before they manifest in accidents. In Bayesian terms, the conditional probability of an unsafe act, given a poor organisational climate, should be substantially elevated above baseline, as quantified in Section 6.4.

6. Case Study: Retrospective Application to the Sáp 2023 Railway Collision

This section demonstrates the framework through retrospective application to a documented Hungarian railway accident. Section 6.1 summarises the case, Section 6.2 presents the three-layer analysis, Section 6.3 compares the output with that of conventional FMEA, and Section 6.4 illustrates the Bayesian layer through a conditional probability table.

6.1. Case Overview

On 15 November 2023 at 04:20, freight train 47487-2 passed the stop signal at Sáp station without authority at approximately 60 km/h and collided with passenger train 6419, which was standing at the platform. Both locomotives were severely damaged; the driver of the passenger locomotive sustained life-threatening injuries, and four further persons received serious or minor injuries. The national investigation body conducted a formal safety investigation under case number 2023-1152-5 [32].

Key factual findings from the investigation report include the following: The Sáp station signalling system was operating normally, controlled remotely from Püspökladány. No Automatic Train Protection was installed on the line section. The freight train driver commenced his shift at 19:00 on 14 November, placing the collision 9 h and 20 min into a night shift. The assigned route for the freight train was pre-set to Platform III, but the train proceeded to Platform II, where the passenger train was standing. The investigation was conducted as a safety investigation and did not attribute individual culpability, consistent with the mandate of national investigation bodies under Directive (EU) 2016/798.

6.2. IHTRA Three-Layer Analysis

The retrospective application of the model across all three layers is presented in Table 8, and the corresponding HFACS causal chain, shown alongside the corresponding survey perceptions, is illustrated in Figure 3.

Table 8. IHTRA retrospective analysis of investigation case 2023-1152-5 (Sáp collision, November 2023).

IHTRA Layer	Component	Finding in the Sáp Collision	HFACS Level
Layer 1—FMEA	Technical failure mode	No Automatic Train Protection installed on the line section; the signalling system operated normally but without train-stop capability	—
Layer 1—FMEA	RPN elevation factor	Absence of ATP removes the technical barrier between the unsafe act and the collision; it does not affect the probability of the act itself	—
Layer 2—HFACS	Unsafe act	Freight driver proceeded past the stop signal without authority (procedural violation)	1
Layer 2—HFACS	Precondition	Shift commenced 19:00, collision 04:20—9 h 20 min night shift; high-probability fatigue precondition	2

Table 8. *Cont.*

IHTRA Layer	Component	Finding in the Sáp Collision	HFACS Level
Layer 2—HFACS	Supervision	No fatigue-monitoring or shift-handover protocol reported (IHTRA interpretation)	3
Layer 2—HFACS	Organisational influence	ATP deferred despite infrastructure upgrade; resource-allocation decision (IHTRA interpretation)	4
Layer 3—Bayesian	Probability update	P(human error | night shift > 9 h, no ATP, low organisational quality) substantially elevated above baseline	All

ATP = Automatic Train Protection; RPN = risk priority number.

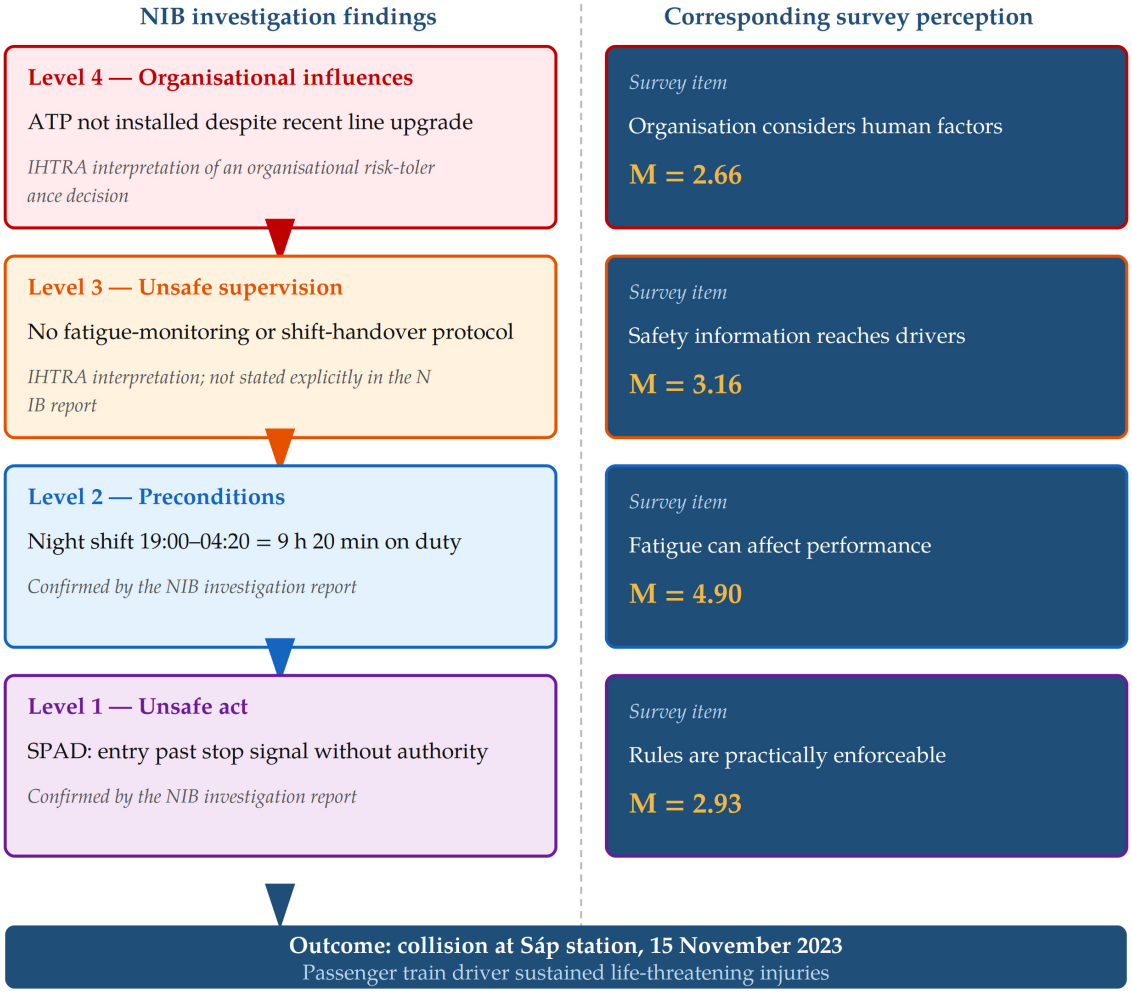


Figure 3. HFACS causal chain of the Sáp 2023 collision, shown alongside the corresponding survey perceptions. Italic annotations indicate interpretations derived from the IHTRA model that are not stated explicitly in the investigation report.

As Figure 3 shows, the HFACS layer structures the human causal chain across all four levels. At Level 1, the unauthorised entry past the stop signal constitutes a procedural violation. At Level 2, a night shift of 9 h and 20 min represents a high-probability fatigue precondition, consistent with the survey finding that fatigue was almost unanimously recognised as affecting performance (M = 4.90, SD = 0.34). At Level 3, the absence of any documented fatigue monitoring or shift handover protocol in the investigation report is interpreted, from a HFACS perspective, as indicative of supervisory failure to correct a known systemic risk; this constitutes an interpretation derived from the model rather than an explicit finding of the investigation. At Level 4, the decision not to install ATP

on a line that had recently received infrastructure upgrades reflects a resource management and risk tolerance decision at the organisational level, which likewise constitutes an interpretive inference.

6.3. Comparison with Conventional FMEA

A conventional FMEA of the Sáp station configuration was carried out following IEC 60812, covering eight failure modes across signalling, interlocking, route setting, the driver–signal interface, train protection, radio communication, braking and crew rostering. Severity, occurrence and detection ratings were assigned by the first author on the basis of professional railway safety experience. The results are given in Table 9.

Table 9. Conventional FMEA of the Sáp station configuration. RPN = S × O × D.

Item/Function	Failure Mode	S	O	D	RPN	Rank
Lineside signalling	Permissive aspect displayed when route not set	10	1	2	20	7
Interlocking	Conflicting route released	10	1	2	20	7
Route setting (remote)	Route set to a platform other than that expected	8	3	8	192	4
Driver–signal interface	Signal passed at danger	10	3	9	270	3
Train protection	SPAD not automatically prevented (no ATP fitted)	10	7	1	70	5
Radio communication	Warning not issued or received in time	8	4	9	288	2
Rolling stock braking	Degraded braking performance	8	2	3	48	6
Train crew rostering	**Driver on duty beyond fatigue-critical shift length**	**9**	**5**	**10**	**450**	**1**

S = severity, O = occurrence, D = detection, each on a 1–10 scale. The highest-ranked row is shown in bold. Detection is rated as the probability that the deficiency is identified before it produces its effect.

Two results deserve comment. First, a properly conducted FMEA does capture fatigue: crew rostering carries the highest risk priority number in the analysis. It would be too strong to claim that conventional FMEA fails to model fatigue at all. The defensible claim, which the present analysis supports, is that FMEA registers fatigue as an occurrence rating on a single row, but cannot represent the organisational conditions that produce it, the interaction between contributors, or diagnostic updating after an event.

Second, the detection rating exposes an internal inconsistency in the risk priority number. The absence of Automatic Train Protection is perfectly detectable in the FMEA sense: it is documented, known to the operator and to the National Safety Authority, and recorded in the infrastructure register. Rated on that reading, its detection score is 1, and its risk priority number is 70, placing it fifth of eight—even though it is the most consequential deficiency in the case. Had detection instead been rated as the inability to intervene in real time, the same row would score 10 and yield 700, placing it first. The metric conflates inability to detect with inability to act and therefore systematically under-ranks hazards that are known but unmitigated.

A Bayesian barrier representation avoids this. In the network specified in Section 3.2, train protection enters through its conditional effect on the collision node rather than through a detection score, so a known but unmitigated barrier cannot be ranked away. This is examined quantitatively in Section 6.4.

As Table 10 indicates, the model produces four additional actionable outputs: a Level 2 recommendation for shift-length monitoring and fatigue-risk-assessment protocols; a Level 3 recommendation for supervisory procedures requiring pre-shift fatigue assessment for night shifts exceeding eight hours; a Level 4 recommendation for a formal organisational review of the ATP deferral decision against the risk acceptance criteria of the Common Safety Method; and a Bayesian node update propagating the elevated night-shift error probability across comparable routes without train protection.

Table 10. Conventional FMEA compared with the IHTRA model on the same case.

Capability Element	Conventional FMEA	IHTRA Model	Source
Technical failure mode identification	Full	Full	Table 9
Risk priority scoring	Static RPN	Dynamic, context-conditioned	Table 11
Fatigue as a contributor	Occurrence rating only	Explicit Level 2 node	Figure 2
Causal antecedents of fatigue	Not representable	Level 3 and Level 4 nodes	Figure 2
Interaction between contributors	Row-by-row only	Joint conditional probability	Table 11
Diagnostic updating after an event	Not available	Posterior inference	Section 6.4
Empirical calibration of parameters	Not available	Three nodes from survey	Table 1

Table 11. Probability of the unsafe act by scenario, with plausible ranges.

Shift Type	Organisational HF Quality	P(SPAD)	Ratio to Baseline	Plausible Range
Day shift	High (M > 3.5)	0.00038	1.00	0.00019–0.00076
Day shift	Low (M < 3.0)	0.00046	1.22	0.00023–0.00092
Night shift > 8 h	High (M > 3.5)	0.00046	1.21	0.00023–0.00092
Night shift > 8 h	**Low (M < 3.0)**	**0.00059**	**1.57**	**0.00030–0.00118**

The Sáp case profile is shown in bold. Ranges are obtained by propagating each assumption-based parameter across the interval given in the sensitivity analysis in Section 6.4.

6.4. Inference over the Bayesian Network

Layer 3 is operationalised by inference over the network specified in Section 3.2. Two quantities are reported: the probability of the unsafe act and the probability of the collision outcome. They are reported separately because Automatic Train Protection is a parent of the second and not of the first, so the two respond differently to the same scenario.

Table 11 gives the probability of a signal passed at danger. Train protection does not appear as a conditioning variable, because it does not influence whether the driver passes the signal. The plausible range in the final column propagates the uncertainty in the three assumption-based conditional tables reported in Appendix A.

Table 12 gives the probability of the collision outcome across the full scenario set. Here, train protection is decisive: with protection fitted, the collision probability is close to the baseline irrespective of shift type or organisational quality, because the barrier intervenes after the unsafe act. With protection absent, the same organisational and fatigue conditions raise the outcome probability by a factor of approximately four.

Table 12. Probability of the collision outcome by scenario.

Shift Type	ATP	Organisational HF Quality	P(collision)	Ratio	Notes
Day shift	Fitted	High	0.000108	1.0	Baseline
Day shift	Fitted	Low	0.000109	1.0	ATP absorbs
Day shift	Absent	High	0.000370	3.4	Barrier removed
Day shift	Absent	Low	0.000408	3.8	—
Night shift > 8 h	Fitted	High	0.000109	1.0	ATP absorbs
Night shift > 8 h	Fitted	Low	0.000112	1.0	ATP absorbs
Night shift > 8 h	Absent	High	0.000406	3.8	—
Night shift > 8 h	**Absent**	**Low**	**0.000466**	**4.3**	**Sáp profile**

Absolute values are conditional on the assumed baseline and should be read as relative rather than actuarial quantities. Bold indicates the scenario profile discussed in the accompanying text.

These figures differ substantially from what a multiplicative treatment of the same modifiers would produce, and the difference is instructive. Multiplying the modifiers directly gives a 15-fold increase for the Sáp profile, whereas propagation through the network yields a 1.6-fold increase in the probability of the unsafe act and a 4.3-fold increase in the probability of the collision. Multiplicative modification overstates compound risk

because most of the probability mass lies on the path in which the driver is neither fatigued nor inattentive, and that mass is not redistributed by multiplying modifiers. We report this as a finding: naive multiplicative risk modification, as used in several extended FMEA variants, can overstate compound risk by close to an order of magnitude. It is an argument for probabilistic propagation rather than for weighted RPN adjustment.

Because the network supports inference in both directions, it also answers a question that FMEA cannot address. Conditioning on the occurrence of a collision, the posterior probability that an attention lapse was present rises from 0.044 to 0.325, a factor of 7.3, while the posterior for organisational human factor quality rises only marginally. The model therefore identifies the proximate contributor most strongly implicated by the observed outcome, while indicating that the distal organisational variables are less sharply diagnostic on the present parameterisation. Diagnostic inference of this kind is not available from a static risk priority number.

Table 13 reports the sensitivity of the results to the parameters that are not empirically grounded. One parameter dominates: the probability of a signal passed at danger in the absence of an attention lapse, for which no data exist. Varying it across three orders of magnitude moves the compound ratio for the unsafe act between 1.1 and 1.9 and the corresponding ratio for the collision outcome between 3.8 and 7.8. We state this explicitly rather than presenting a single figure, because the width of that interval is itself the most important result of the sensitivity analysis: the compound risk ratio cannot be pinned down without incident data, and no amount of structural refinement substitutes for that.

Table 13. Sensitivity of the reported probabilities to the assumption-based parameters.

Parameter	Range Examined	Effect on P(SPAD)	Source
P(fatigue | night shift, no fatigue management)	0.25–0.40	−7% to +8%	Literature
P(lapse | high fatigue, poor information flow)	0.10–0.22	−12% to +17%	Assumption
P(SPAD | lapse, low rule enforceability)	0.006–0.016	−21% to +31%	Assumption
P(SPAD | no lapse)	**1×10^{-6}–1×10^{-3}**	**ratio 1.1 to 1.9**	**Assumption**

Each parameter was varied independently across the stated range, with all others held at their central value. The final row, shown in bold, identifies the parameter that dominates the compound ratio and for which no empirical data are available.

7. Discussion

This section interprets the findings in relation to the research questions and the wider literature. Sections 7.1–7.3 discuss the empirical support for the framework, the regulatory compliance gap and the implications for Bayesian calibration, while Section 7.4 sets out the limitations of this study.

7.1. What the Survey Constrains, and What It Does Not

The survey supplies four of the eight evidence-based parameters of the network, distributed across three nodes, and is consistent with three structural premises of the framework, addressing RQ1 and RQ3. Consistency is not confirmation, and the following observations should be read as motivating the specification rather than as testing it. First, human factors are not peripheral to railway risk but are pervasively recognised by the operational workforce as dominant safety determinants; the near-ceiling awareness scores (M = 4.87–4.90) are consistent with the premise that the static treatment of human error in FMEA is inadequate, although a perception survey cannot establish that premise on its own. Second, the organisational layer is the weakest link in the current safety system, as evidenced by the lowest domain means. Third, human factor awareness correlates only weakly with organisational culture perception (r = 0.163). The exploratory analysis reported in Section 4.3 is consistent with a separation between the two but is not confirmatory; the

modelling decision to treat them as distinct nodes rests on the observed correlation and on the HFACS structure, not on the factor solution.

The case analysis identified two HFACS levels as confirmed by the investigation findings—the shift pattern and the signal passed at danger—and two further levels, supervision and organisational influence, as plausible under the model interpretation but not stated in the investigation report. This distinction is maintained throughout, and no claim is made that all four levels were established empirically. The Level 2 finding of night-shift fatigue maps directly onto the highest-scoring survey item. The two are not equivalent: one is a causal factor in a single documented event, the other a general perception reported by a different population of drivers. The correspondence is therefore suggestive rather than confirmatory and is presented as such.

7.2. The Regulatory Compliance Gap

A particularly significant finding is the gap between practical rule enforceability (M = 2.93) and perceived rule importance (M = 3.98). This compliance–reality gap between normative acceptance and operational compliance capacity has been documented in other transport domains [33]. Within the HFACS framework, it corresponds to the intersection of Level 1 violations and the Level 4 organisational processes that create the conditions for violation. The finding suggests that regulatory frameworks, including the Common Safety Method Regulation, may create formal compliance obligations that are not operationally achievable under current working conditions. This has direct implications for the National Safety Authority: RPN scores derived from conventional FMEA that assume full regulatory compliance will systematically underestimate actual risk in contexts where enforceability is perceived as low.

7.3. Implications for Bayesian Network Calibration

The empirical data provide concrete inputs for Bayesian parameterisation. The near-unanimous awareness scores establish high prior probabilities for fatigue, stress and workload as active risk factors in Hungarian railway operations. The low organisational consideration scores (M = 2.66) suggest that organisational barrier effectiveness should be modelled with substantially lower conditional probabilities than would be assumed in Western European contexts with higher safety culture ratings. This Central European calibration, distinct from the British and Chinese data used in published Bayesian railway models, constitutes an original contribution to the regional applicability and context-specificity of hybrid risk assessment frameworks.

7.4. Limitations

Several limitations must be acknowledged. The 11–20-year experience cohort comprises five respondents. This group has been excluded from the inferential comparisons in Section 4.5 on the ground that $n < 10$ does not support inference, and no substantive interpretation is offered for it. Each HFACS subcategory in Table 7 is operationalised by a single survey item; single-item measures do not permit reliability assessment and are susceptible to wording effects, which contrasts with the multi-item treatment of the four thematic domains. The survey relies on self-reported perception, which introduces potential social desirability bias, particularly for items relating to organisational performance. The cross-sectional design precludes causal inference. The sample is drawn from the Hungarian context, so generalisation to other national railway systems requires caution.

The questionnaire is a custom-developed instrument that has not undergone formal psychometric validation by confirmatory factor analysis. An exploratory factor analysis was conducted after inspection of the data and was not preregistered. It suggested a two-factor structure, but on a five-item scale with n = 89 and KMO = 0.627, this cannot be distinguished

from a statistical artefact. The two-factor interpretation is therefore treated as a hypothesis requiring confirmatory testing on an independent sample, not as a finding. Comparability with industry-validated instruments such as the Nordic Safety Climate Questionnaire is correspondingly limited [29]. The snowball sampling strategy introduces a bias whose direction cannot be determined. Recruitment through trade union channels might over-represent safety-committed drivers, but it might equally over-represent respondents inclined to report dissatisfaction. Without population-level comparison data, which are not available, no directional inference is warranted, and none is drawn. Finally, the case analysis is retrospective and based on a single case; it therefore constitutes a proof-of-concept demonstration rather than empirical validation, and prospective validation across multiple incidents is required to establish predictive validity.

8. Conclusions

This paper has specified an integrated FMEA–HFACS–Bayesian framework for railway risk assessment and demonstrated what the specification makes analytically possible. The contribution is methodological. A framework has been specified in full, with its evidential basis reported parameter by parameter; a survey of 89 Hungarian train drivers supplies four of its eight evidence-based parameters, distributed across three nodes; and a documented investigation report has been used to show what the specified framework generates that a conventional FMEA of the same configuration does not. The principal findings are as follows:

- Human factor risk awareness is near-universal among Hungarian train drivers, with fatigue (M = 4.90) and workload (M = 4.87) approaching ceiling values.
- Organisational consideration of human factors is critically low (M = 2.66), identifying HFACS Levels 3 and 4 as the primary systemic weakness.
- Exploratory analysis of the human factors domain is consistent with a separation between individual awareness and perceived organisational response, but this post hoc interpretation requires confirmatory testing before it can be treated as established.
- A regulatory compliance gap exists between rule acceptance (M = 3.98) and practical enforceability (M = 2.93).
- Professional experience does not significantly moderate safety culture perception (all $p > 0.05$), indicating the primacy of organisational over individual variables.
- The case analysis identified two HFACS levels as confirmed by the investigation findings and two further levels as plausible under the model interpretation; inference over the specified network gives an illustrative increase from $p = 0.00038$ to $p = 0.00059$ for the unsafe act and from $p = 0.00011$ to $p = 0.00047$ for the collision outcome.

Future research will pursue the following directions:

- Prospective validation against the corpus of published final investigation reports of the Hungarian national investigation body, which are freely available; access to operator-held rostering data has not been secured and is not assumed by this plan.
- Formal calibration of the four assumption-based conditional probability tables by structured expert elicitation following Cooke's Classical Model, with calibration questions drawn from published national safety indicators, a target panel of ten to fifteen railway safety experts, and performance-based weighting.
- Confirmatory factor analysis of the survey instrument, applied in parallel with the Nordic Safety Climate Questionnaire.
- Comparative application across Central European railway operators in order to assess cross-national generalisability.

Taken together, these findings demonstrate that safety in Central European railway systems cannot be adequately measured without a framework that captures both technical failure modes and the organisational conditions that shape human reliability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app16188976/s1, S1_anonymised_survey_data.csv: anonymised survey responses (n = 89); S2_bayesian_network_model.py: executable implementation of the Bayesian network, written in Python 3 using the open-source pgmpy library version 1.1.2, reproducing the values reported in Tables 11–13 and the diagnostic posteriors in Section 6.4; S3_codebook.txt: codebook mapping each survey item to its thematic domain.

Author Contributions: Conceptualisation, Á.P. and I.L.; methodology, Á.P.; formal analysis, Á.P.; investigation, Á.P.; data curation, Á.P.; writing—original draft preparation, Á.P.; writing—review and editing, Á.P. and I.L.; visualisation, Á.P.; supervision, I.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study does not require national-level ethical review and approval from ETT-TUKEB. Under Hungarian law, mandatory ethical approval applies only to medical/biomedical human subject studies as defined in the Healthcare Act of 1997 (CLIV) (Article 157), Ministerial Decree No. 23/2002 of 2002 (V.9.), and Government Decree No. 235/2009 of 2009 (X.20.). This study is a non-interventional, anonymous survey of opinions from professionals in engineering disciplines and therefore does not fall under the scope of the aforementioned legal documents. Széchenyi István University has no institutional ethics committee empowered to review or grant exemptions for non-medical questionnaire research, so no institutional ethical clearance was obtained for this work.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study. Participants were informed of the purpose of the research and of the anonymous nature of data collection before completing the questionnaire.

Data Availability Statement: The anonymised survey dataset supporting the reported results is provided as Supplementary Material and deposited in a public repository as a fully anonymised file containing the 26 Likert responses and the broad experience category, with no free-text fields and no information permitting re-identification. The investigation report analysed in Section 6 is publicly available from the Hungarian national investigation body.

Acknowledgments: The authors thank the train drivers who participated in the survey and the railway safety experts who reviewed the questionnaire items. During the preparation of this manuscript, the authors used AI tools for the following purposes: statistical analysis of the raw questionnaire data collected by the authors; construction and parameterisation of the Bayesian network structure; preparation of the figures; identification and verification of literature sources; and preparation of the manuscript text, including the English wording. The research question, the study design, the data collection, the domain expertise applied and all substantive decisions are the authors' own. The authors reviewed all output, verified the statistical results against the raw data, and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
ATP	Automatic Train Protection
CPT	Conditional Probability Table
CSM	Common Safety Method
EFA	Exploratory Factor Analysis

ERA	European Union Agency for Railways
FMEA	Failure Mode and Effects Analysis
FTA	Fault Tree Analysis
HFACS	Human Factors Analysis and Classification System
IHTRA	Integrated Human–Technical Risk Assessment
KMO	Kaiser–Meyer–Olkin measure of sampling adequacy
NIB	National Investigation Body
NSA	National Safety Authority
RAMS	Reliability, Availability, Maintainability and Safety
RPN	Risk Priority Number
SPAD	Signal Passed at Danger

Appendix A

Table A1 reports every conditional probability table of the network specified in Section 3.2. Each entry gives the probability of the unfavourable state of the node, conditional on the stated parent configuration; the complementary probability is implied. The source column identifies the evidential basis of each entry. Entries marked as assumptions have not been calibrated against data and are the subject of the sensitivity analysis reported in Table 13.

Table A1. Conditional probability tables of the Bayesian network, with evidential sources.

Node	Condition	P(unfavourable)	Source
OrgHF	—	0.585	Survey, M = 2.66
RuleEnf	—	0.517	Survey, M = 2.93
SafetyInfo	OrgHF = High	0.20	Survey (marginal 0.463)
SafetyInfo	OrgHF = Low	0.65	Survey (marginal 0.463)
FatMgmt	OrgHF = High	0.25	Assumption
FatMgmt	OrgHF = Low	0.80	Assumption
Fatigue	Day, management present	0.08	Literature [27,28]
Fatigue	Day, management absent	0.13	Literature [27,28]
Fatigue	Night > 8 h, management present	0.20	Literature [27,28]
Fatigue	Night > 8 h, management absent	0.32	Literature [27,28]
Lapse	Low fatigue, adequate information	0.020	Assumption
Lapse	Low fatigue, inadequate information	0.035	Assumption
Lapse	High fatigue, adequate information	0.090	Assumption
Lapse	High fatigue, inadequate information	0.150	Assumption
SPAD	No lapse, high enforceability	0.0001	Assumption
SPAD	No lapse, low enforceability	0.0002	Assumption
SPAD	Lapse, high enforceability	0.005	Assumption
SPAD	Lapse, low enforceability	0.010	Assumption
Collision	No SPAD, ATP fitted	0.0001	Assumption
Collision	No SPAD, ATP absent	0.0002	Assumption
Collision	SPAD, ATP fitted	0.020	Assumption
Collision	SPAD, ATP absent	0.450	Assumption

Survey-derived priors use the transformation P(unfavourable) = (5 − M)/4 applied to the corresponding item mean. Literature-derived entries follow the fatigue multiplier range of 2.0 to 2.5 for extended night shifts reported in [27,28]. The model implementation reproducing every value in this table is provided as Supplementary Material.

References

1. European Union Agency for Railways. *Annual Safety Report: European Railway Safety*; Publications Office of the European Union: Luxembourg, 2023.
2. Wang, W.; Liu, X.; Qin, Y. A modified FMEA model to improve reliability and safety evaluation for dynamic systems: A case study of railway turnout systems. *Qual. Reliab. Eng. Int.* **2018**, *34*, 1536–1551.
3. Hollnagel, E.; Woods, D.D.; Leveson, N. *Resilience Engineering: Concepts and Precepts*; Ashgate: Aldershot, UK, 2006.
4. Reason, J. *Human Error*; Cambridge University Press: Cambridge, UK, 1990.
5. *EN 50126*; Railway Applications—The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS). European Committee for Electrotechnical Standardization: Brussels, Belgium, 2017.
6. Wang, W.; Liu, X.; Qin, Y. Dynamic FMEA with fuzzy cognitive maps and prospect theory for railway systems. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108163.
7. Fu, S.; Yan, X.; Zhang, D.; Zhang, M. Risk assessment for railway transportation using an extended FMEA with prospect theory under fuzzy environments. *Entropy* **2020**, *22*, 1424.
8. Shappell, S.A.; Wiegmann, D.A. *The Human Factors Analysis and Classification System (HFACS)*; FAA Civil Aeromedical Institute: Oklahoma City, OK, USA, 2000.
9. Reinach, S.; Viale, A. Application of a human error framework to conduct train accident/incident investigations. *Accid. Anal. Prev.* **2006**, *38*, 396–406. [CrossRef] [PubMed]
10. Zhan, Q.; Zheng, W.; Zhao, B. A hybrid human and organizational analysis method for railway accidents based on HFACS-Railway Accidents (HFACS-RAs). *Saf. Sci.* **2017**, *91*, 232–250. [CrossRef]
11. Gawlak, K. Analysis and assessment of the human factor as a cause of occurrence of selected railway accidents and incidents. *Open Eng.* **2023**, *13*, 20220398. [CrossRef]
12. Guo, Z.; Pang, H.; Zhang, J.; Zhang, J.; Wang, J.; He, C.; Li, C. Using HFACS to understand human error in railway dispatcher performance: A study of proactive safety inspection records. *Ergonomics* **2025**, *68*, 37–50. [CrossRef] [PubMed]
13. Baysari, M.T.; McIntosh, A.S.; Wilson, J.R. Understanding the human factors contribution to railway accidents and incidents in Australia. *Accid. Anal. Prev.* **2008**, *40*, 1750–1757. [CrossRef] [PubMed]
14. Ale, B.J.M.; Bellamy, L.J.; Cooke, R.M. Further development of a Causal model for Air Transport Safety (CATS): Building the mathematical heart. *Reliab. Eng. Syst. Safe* **2009**, *94*, 1433–1441. [CrossRef]
15. Shi, L.; Liu, J.; Zhang, Y.; Liang, W. Data-driven Bayesian network analysis of railway accident risk. *IEEE Access* **2024**, *12*, 38631–38645. [CrossRef]
16. Li, Y.; Ding, S.; Wang, S.; Sun, Y.; Ge, D. Multiscenario deduction analysis for railway emergencies using knowledge metatheory and dynamic Bayesian networks. *Reliab. Eng. Syst. Saf.* **2025**, *255*, 110675. [CrossRef]
17. Li, Y.; Ding, S.; Wang, S.; Zhang, J.; Liu, H. Safety assessment method of high-speed rail interval structure based on weighted Bayesian network. *KSCE J. Civ. Eng.* **2024**, *28*, 3286–3300. [CrossRef]
18. Bhuiyan, M.R.; Ibtihal, S.A.; Nokshi, K.N. Involvement of human and organizational factors in railway accidents: Application of the human factors analysis and classification system and Bayesian network. *Transp. Res. Rec.* **2023**, *2677*, 496–508. [CrossRef]
19. Ciani, L.; Guidi, G.; Patrizi, G. Human reliability in railway engineering: Literature review and bibliometric analysis of the last two decades. *Saf. Sci.* **2022**, *151*, 105755. [CrossRef]
20. European Commission. *Commission Implementing Regulation (EU) No 402/2013 on the Common Safety Method for Risk Evaluation and Assessment*; Official Journal of the European Union: Brussels, Belgium, 2013.
21. European Union Agency for Railways. *Report on Railway Safety and Interoperability in the EU 2020*; Publications Office of the European Union: Luxembourg, 2020.
22. European Union Agency for Railways. *Railway Safety Performance in the European Union 2014*; Publications Office of the European Union: Luxembourg, 2014.
23. Sangiorgio, V.; Mangini, A.M.; Precchiazzi, I. A new index to evaluate the safety performance level of railway transportation systems. *Saf. Sci.* **2020**, *131*, 104921. [CrossRef]
24. Liou, J.J.H.; Hsu, C.-C.; Liu, L. A hybrid model integrating FMEA and HFACS to assess the risk of inter-city bus accidents. *Complex Intell. Syst.* **2022**, *8*, 3893–3907. [CrossRef]
25. Fan, C.; Huang, S.; Lin, S.; Xu, D.; Peng, Y.; Yi, S. Types, risk factors, consequences, and detection methods of train driver fatigue and distraction. *Comput. Intell. Neurosci.* **2022**, *2022*, 8328077. [CrossRef] [PubMed]
26. International Union of Railways. *Railway Safety Indicators for 2023: Focus on Signals Passed at Danger (SPADs)*; UIC Communications: Paris, France, 2024.
27. Chang, H.-L.; Ju, L.-S. Effect of consecutive driving on accident risk: A comparison between passenger and freight train driving. *Accid. Anal. Prev.* **2008**, *40*, 1844–1849. [CrossRef] [PubMed]
28. Härmä, M.; Sallinen, M.; Ranta, R.; Mutanen, P.; Müller, K. The effect of an irregular shift system on sleepiness at work in train drivers and railway traffic controllers. *J. Sleep Res.* **2002**, *11*, 141–151. [CrossRef] [PubMed]

29. Kines, P.; Lappalainen, J.; Mikkelsen, K.L.; Olsen, E.; Pousette, A.; Tharaldsen, J.; Tómasson, K.; Törner, M. Nordic Safety Climate Questionnaire (NOSACQ-50): A new tool for diagnosing occupational safety climate. *Int. J. Ind. Ergon.* **2011**, *41*, 634–646. [CrossRef]
30. Atkinson, R.; Flint, J. Accessing hidden and hard-to-reach populations: Snowball research strategies. *Soc. Res. Update* **2001**, *33*, 1–4.
31. MÁV Group. *MÁV Group Annual Report 2023*; Magyar Államvasutak Zrt.: Budapest, Hungary, 2023.
32. Hungarian Transport Safety Bureau. *Final Investigation Report 2023-1152-5: Collision of Trains 47487-2 and 6419 at Sáp Station, 15 November 2023*; Közlekedésbiztonsági Szervezet: Budapest, Hungary, 2025. Available online: https://kbsz.hu/dokumentumok/2023-1152-5_zj_alairt.pdf (accessed on 17 August 2026).
33. Reason, J.; Parker, D.; Lawton, R. Organizational controls and safety: The varieties of rule-related behaviour. *J. Occup. Organ. Psychol.* **1998**, *71*, 289–304. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.