

Quantification of Occupational Safety Risk In A Boiler Area Through Integration of Quantitative Fault Tree Analysis and HIRADC: A Case Study In The Creamer Industry

Muhammad Ubaidillah^{1*}, Moh. Jufriyanto², Mega Rahayu H³

^{1,2,3}Program Studi Teknik Industri, Fakultas Teknik, Universitas Muhammadiyah Gresik,
Jawa Timur 6112, Indonesia

*Corresponden author:

Email: ubaidbatandoz@gmail.com

Abstract.

The boiler area is among the most hazardous workplaces in manufacturing because it simultaneously involves thermal energy, high pressure, and electrical energy. Applications of Fault Tree Analysis (FTA) in the Indonesian occupational safety literature commonly stop at the qualitative level, serving as a cause-mapping tool that yields neither event probabilities nor a measurable control priority ranking. This study aims to quantify occupational accident risk in the boiler area of PT Sehat Alam Segar by integrating quantitative FTA with Hazard Identification, Risk Assessment, and Determining Control (HIRADC). Fifty-three basic events across three top events were quantified by converting a frequency-anchored likelihood scale into per-exposure probabilities, then evaluated using two fault tree structures: an all-OR structure (Model A) and a defence-in-depth structure incorporating AND gates (Model B). Model A overestimated event frequency by 9.6–65.8 times relative to actual incident data for 2024–2025, whereas Model B remained within a factor of 1.49–1.86 and showed no significant difference from the actual data (Poisson test, $p = 0.061$ and $p = 1.000$). Model B top event probabilities were 1.044×10^{-1} per steam-line maintenance job, 1.221×10^{-4} per operating day (4.36×10^{-2} per year), and 1.245×10^{-2} per electrical job. Fussell–Vesely analysis showed that human and management factors contributed 51.6% of structural importance. The HIRADC control design based on the ISO 45001:2018 hierarchy reduced the mean risk score from 8.79 to 3.83 (Wilcoxon signed-rank, $Z = 6.360$; $p < 0.001$; $r = 0.874$), eliminated all Extreme-category events, and was estimated to lower top event probabilities by factors of 17–266. This reduction is a model-based estimate rather than a post-implementation measurement. These findings confirm that a fault tree structure which does not represent protective layers yields systematically erroneous risk estimates.

Keywords: quantitative fault tree analysis; HIRADC; boiler safety; importance measure and residual risk.

I. INTRODUCTION

Occupational safety and health (OHS) in the manufacturing industry relies on an organization's ability to identify hazards before they become accidents. In utility systems, boilers occupy a unique position because they combine three forms of hazardous energy: thermal energy in the form of pressurized steam with high surface temperatures, mechanical energy in the form of pressure in vessels and piping, and electrical energy in instrumentation and drive systems. This combination allows a single failure to develop catastrophic consequences quickly. These characteristics have led to boilers being repeatedly selected as the object of OHS risk assessment studies, both in the palm oil industry (Mahyunis et al., 2023) and in steam-powered power plants (Wiradara et al., 2024). Marsudi (2021) specifically examined the risk of boiler system failure using Fault Tree Analysis.

In Indonesia, the operation of steam engines is specifically regulated by the 1930 Steam Ordinance (Stoom Ordonnantie 1930) and the 1930 Steam Regulations (Government of the Dutch East Indies, 1930), which require periodic inspection and testing by labor inspectors, with Law No. 1 of 1970 concerning Occupational Safety (Republic of Indonesia, 1970) as the general umbrella. Supporting propulsion and production equipment in the same area—including feedwater pumps and blowers—are subject to Ministerial Regulation No. 38 of 2016 concerning Occupational Health and Safety of Power and Production Engines (Ministry of Manpower of the Republic of Indonesia, 2016). The risk management framework that governs the implementation of these provisions refers to ISO 31000:2018 (International Organization for Standardization, 2018a) and ISO 45001:2018 (International Organization for Standardization, 2018b); both require that risk assessments be carried out systematically and followed up with controls that follow a hierarchy.

PT Sehat Alam Segar is a manufacturer of creamer, foamer, and growing-up milk that operates a coal-fired boiler continuously 24 hours a day following a production schedule. The boiler studied had a capacity of

10,000 kg of steam per hour with a working pressure of 2.5 MPa (equivalent to 25 bar) and a furnace temperature of 686 °C. This magnitude places the research object at the upper limit of the pressure range of industrial package boilers, and is the physical basis for why the severity of the boiler explosion was assessed at the highest level in this study: the energy stored in a pressure vessel of 2.5 MPa with an evaporation rate of 10 tons per hour is large enough to cause catastrophic consequences if released uncontrolled.

Based on internal incident records for the 2024–2025 period, the boiler area recorded 17 incidents, consisting of work accidents, near misses, and work-related health problems, and accounted for approximately 35% of near miss reports compared to other utility areas. The company has a Hazard Identification and Risk Assessment (HIRA) and Job Safety Analysis (JSA) document for this area.

The problem is that both documents stop at the hazard identification level. They list hazards and their respective risk scores, but they don't specify what combination of failures must occur simultaneously for an accident to occur. A causal relationship in this study refers to the logical relationship between events that defines the necessary and sufficient conditions for an accident to occur, expressed through logic gates: an OR gate if one cause alone is sufficient to cause the event, and an AND gate if multiple causes must occur simultaneously. As an illustration, the company's HIRA document lists "exposure to hot steam during maintenance" as a single hazard with a single risk score. The document doesn't specify that exposure occurs only if two conditions occur simultaneously: thermal energy remains confined at the work point due to failed energy isolation, and the barriers protecting workers—personal protective equipment, work permits, and supervision—also fail simultaneously.

Due to the lack of logical relationships, the existing document fails to answer the three questions that ultimately determine control: what is the probability of an accident occurring per unit of work; which combination of failures provides the shortest path to an accident; and which hazards, if eliminated, provide the greatest risk reduction at the system level, not just at the hazard level itself. These three questions require a logical model, not a hazard list.

This research therefore began by developing a custom accident logic model for the boiler area. Hazard identification was conducted through field work process observations, reviewing HIRA documents, JSAs, and company incident records, and interviews with personnel directly involved in the work. This process yielded three top events and 53 basic events, each of which was assessed using a 5x5 matrix in accordance with AS/NZS 4360:2004 (Standards Australia/Standards New Zealand, 2004) following the company's established scale. These ordinal ratings are treated as a starting point rather than a final result, as the extent to which such ordinal rankings are justifiable is what this article aims to examine.

Such limitations of ordinal assessment are not unique to a single object, but rather a recurring pattern in the application of FTA in the Indonesian OHS literature. Ekoanindiyo et al. (2021) and Marsudi (2021) both used FTA as a cause mapping tool without calculating the probability of top events, while HIRARC-based studies in boiler areas by Mahyunis et al. (2023) and Wiradara et al. (2024) determined control priorities based on risk score rankings. This approach has three fundamental weaknesses. First, the matrix scores are ordinal, meaning that the multiplication of likelihood and severity has no cardinal meaning and cannot be aggregated to the system level. Cox (2008) demonstrated mathematically that risk matrices are only able to correctly distinguish a small proportion of risk pairs and, under certain conditions, actually rank quantitatively smaller risks higher. Second, without quantification, the structural contribution of a basic event to a top event—which depends on its position in the fault tree, not just on its own score—cannot be known (Vesely et al., 1981). Third, the effectiveness of recommended controls cannot be measured or tested statistically.

A more fundamental weakness lies in the fault tree structure itself. A common practice is to connect all basic events to their intermediate events through OR gates without exception, so that each basic event constitutes a minimal cut set of order one. Logically, this structure implies that a single omission—for example, a worker failing to wear heat-resistant gloves—is sufficient to cause a boiler explosion without requiring any other failures. This assertion is inconsistent with the defense-in-depth principle underlying boiler safety system design and deviates from the fault tree convention, which requires gate selection to be based on the logic of

the event's occurrence, not on the completeness of the list of causes (Vesely et al., 1981). As demonstrated in this study, this structure produces probability estimates that deviate significantly from actual event data.

The Hazard Identification, Risk Assessment, and Determining Control (HIRADC) method addresses these shortcomings on the control side. What distinguishes it from HIRA is the necessity to continue the analysis beyond the assessment stage: controls must be selected according to hierarchical levels, and the risk is reassessed after the controls have been implemented—a value known as residual risk. Henny et al. (2025) demonstrated that a structured application of HIRADC in the manufacturing industry can comprehensively map hazard–risk–control relationships and support the assessment of the effectiveness of OHS programs. However, HIRADC itself lacks a mechanism for tracing causality between causes, so prioritizing controls remains dependent on individual risk score rankings.

The method chosen in this study was not based on the assumption that integrating quantitative FTA with HIRADC is the absolute best approach, but rather on its suitability to the characteristics of the problem and the availability of data on the research object. Several other methods are commonly used for similar problems, each with advantages not offered by FTA. Hazard and Operability Study (HAZOP) excels at systematically tracking process parameter deviations using guide words, as applied by Rahmanto and Ihsan Hamdy (2022) in a coal-fired power plant unit. However, it is qualitative in nature and does not generate event probabilities. The Bowtie method combines fault trees and event trees in a single diagram, making it superior in communicating barriers to management. However, Alizadeh and Moshashaei (2015) note that this method is inherently qualitative and is generally chosen when a quantitative approach is deemed unnecessary or impractical. Failure Mode and Effect Analysis (FMEA) works inductively from components to impacts and produces a Risk Priority Number. However, the RPN is also a multiplication of ordinal scales, thus inheriting the same weaknesses as the risk matrix (Cox, 2008). HIRARC and HIRADC are strong on the control side, but assess each hazard in isolation without a causal structure.

Table 1. Comparison of K3 risk analysis methods relevant to the research problem

Method	Direction of analysis	Main output	Probability of an event	Structural importance measures	Hierarchy of control	Implementation reference
HIRARC / HIRADC	Per hazard (without causal relationship)	Risk and control scores	No	No	Yes (HIRADC)	Mahyunis et al. (2023); Wiradara et al. (2024); Henny et al. (2025)
HAZOP	Process parameter deviation	List of deviations, causes, and effects	No	No	Part	Rahmanto & Ihsan Hamdy (2022)
FMEA	Inductive (component → impact)	Risk Priority Number per failure mode	No (ordinal)	No	Part	—
Bowtie	Bidirectional (cause–barrier–impact)	Prevention and mitigation barrier diagram	Optional, generally qualitative	No	Yes (barrier based)	Alizadeh & Moshashaei (2015)
Qualitative FTA	Deductive (top event → cause)	Causal structure and cut set lists	No	No	No	Ekoanindiyo et al. (2021); Marsudi (2021)
Quantitative FTA + HIRADC (this study)	Deductive	Top event probability, minimum cut set, importance measure, residual risk	Yes	Yes	Yes	Vesely et al. (1981); ISO 45001:2018

Source: compiled by the author based on references in the last column.

A brief comparison of the six approaches is presented in Table 1. FTA was chosen for three reasons inherent to the problem under study. First, the research problem is deductive—moving from an undesired accident to its combined causes—which is the direction of FTA, in contrast to the inductive approach of FMEA. Second, only FTA provides structural importance measures (Fussell–Vesely, Birnbaum, Risk Reduction Worth, and Risk Achievement Worth) to rank causes based on their impact on system probability (Vesely et al., 1981). Third, FTA quantification can be performed without a complete component reliability database as long as the likelihood scale used includes frequency anchoring—a requirement met for this object. HIRADC was chosen as the counterpart because FTA lacks a hierarchical control-setting mechanism or residual risk assessment, whereas HIRADC provides both and is a well-known framework within the company.

It should also be noted that this integration has limitations. FTA requires an assumption of independence between basic events, which is not always met in practice. It does not model the time sequence of events like Event Tree Analysis, and it does not explicitly assess consequences like Bowtie. The choice of method is therefore a decision conditioned by the characteristics of the problem and the availability of data, not a claim of universal superiority over other methods.

The novelty of this research lies in three interconnected aspects. First, probability quantification is carried out by utilizing frequency anchoring inherent in the company's likelihood scale, so that quantitative FTA can be applied without a component reliability database. Second, two different fault tree structures are tested against actual incident data, rather than being simply assumed as modeling assumptions. Third, control priorities are determined from structural importance measures rather than from matrix score rankings. All three findings yield findings not obtainable in qualitative analysis: that the entire OR gate structure deviates systematically from reality and that the risk matrix rankings only explain 42.4% of the variance in structural contribution rankings. Based on the author's search of Indonesian and international OSH literature, the integration of quantitative FTA with HIRADC, accompanied by model validation against actual incident data and statistical testing of control designs, is still rarely reported in the context of food industry boiler utilities in Indonesia. This statement is presented as a limitation of the author's research, not as an absolute claim regarding the absence of similar research.

Based on the description, this study aims to achieve four objectives. First, to quantify the probability of occurrence of each top event in the boiler area and to test the suitability of the fault tree structure used to actual incident data for the 2024–2025 period. Second, to determine the basic event that provides the largest structural contribution to the probability of the top event based on the importance measure. Third, to develop a HIRADC control plan based on the ISO 45001:2018 hierarchy along with an estimate of the resulting reduction in risk levels and top event probabilities. Fourth, to test the statistical significance of this risk reduction.

The achievement of these four objectives is aimed at generating two benefits. For companies, this study provides a verifiable and replicable quantitative basis for prioritizing risk control investments in the boiler area. For scientific development, this study demonstrates the methodological consequences of selecting a fault tree structure on the validity of risk estimates, while also providing an FTA quantification procedure applicable to companies that do not yet have a component reliability database.

II. Research Methods

2.1 Research Design and Object

This research is a quantitative research with a single case study design that is analytical-predictive in nature. The object of the research is a coal-fired boiler system along with its steam lines and supporting electrical systems in the Utilities Department of PT Sehat Alam Segar, Wings Manyar Area, Gresik Regency, East Java. The boiler has a capacity of 10,000 kg of steam per hour at a working pressure of 2.5 MPa with a furnace temperature of 686 °C and a burnback temperature of 238 °C, and consumes 1.72 tons of coal and 10.5 m³ of feedwater per hour. These parameters are the physical basis for the severity assessment of Top Event 1 and Top Event 2, as well as the reason for the failure of the combustion system and feedwater to be treated as triggers for abnormal conditions. The unit of analysis is the basic event, which is a basic event that cannot be

further decomposed in the fault tree. The population and research sample are all 53 basic events identified in the three top events (saturated samples).

The research framework draws on three primary references. The fault tree structure and quantification follow the Fault Tree Handbook (Vesely et al., 1981), including logic gate design, minimum cut set derivation, and importance measure calculation. The risk management framework and hierarchical control design follow ISO 31000:2018 (International Organization for Standardization, 2018a) and ISO 45001:2018 (International Organization for Standardization, 2018b). The risk assessment scale follows AS/NZS 4360:2004 (Standards Australia/Standards New Zealand, 2004) as used by the company, ensuring that the research results remain legible within the applicable OHS documentation framework for the research object.

The shift from a qualitative to a quantitative approach in this research occurred in four aspects. First, the risk assessment, originally based on ordinal rankings, was changed to a ratio-scale probability. Second, risk aggregation from the basic event level to the top event level was performed using Boolean algebra, rather than summing or sorting scores. Third, control priorities were determined based on importance measures calculated from the tree structure, rather than individual score rankings. Fourth, the effectiveness of the control design was tested using inferential statistics, rather than narrative conclusions. The overall flow of the research stages is presented in Figure 1.

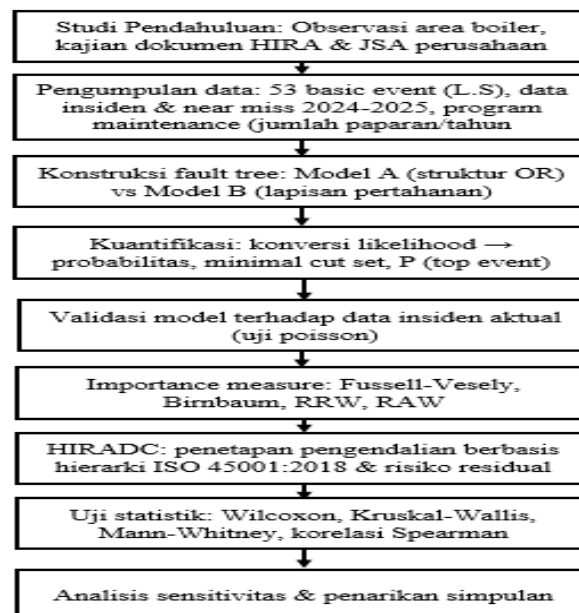


Fig. 1. Research stage flow

2.2 Types and Sources of Data

The research data consists of three groups. The first group is the likelihood (L) and severity (S) values for 53 basic events, which are primary data resulting from the expert judgment of three sources representing operational, technical, and administrative perspectives (boiler operators, mechanics, and utility staff). If there are differences in assessments between sources, the highest value is determined as the final value to ensure a conservative approach to the worst-case scenario. The assessment scale refers to AS/NZS 4360:2004 as presented in Table 2.

The second group is historical data on boiler area incidents and near-misses for the 2024–2025 period, sourced from the company's incident reporting system (Table 3). This data serves two purposes: as the basis for determining likelihood values and as a basis for independent validation of the model's generated top event probabilities.

The third group is boiler maintenance program data used to determine the annual exposure (demand) for each top event (Table 4). Determining this exposure is critical because it serves as the denominator in converting frequency to probability, so its details are presented explicitly to allow for replication of the calculations.

Table 2. Likelihood and severity scales and annual frequency conversion

Mark	Likelihood category	Frequency interval	λ (times/year)	Severity category	Severity description
1	Rare	< 1 time/year	0.5	Insignificant	without injury or material loss
2	Unlikely	1–2 times/year	1.5	Minor	minor injuries, sufficient first aid treatment, workers remain on duty
3	Possible	3–5 times/year	4.0	Moderate	moderate injury, requiring medical attention, resulting in lost work days
4	Likely	6–10 times/year	8.0	Major	serious injury to permanent disability, major material loss
5	Almost Certain	> 10 times/year	12.0	Catastrophic	loss of life, total equipment damage, widespread environmental impact

Source: AS/NZS 4360:2004 (Standards Australia/Standards New Zealand, 2004), modified. The λ value is the midpoint of the frequency interval; for open-boundary levels 1 and 5, 0.5 and 12.0 times/year are used.

Table 3. Incident and near miss data for boiler areas for the period 2024–2025

No	Type of incident	Category	Amount	Relevance	Information
1	Exposure to hot steam during maintenance of sensors, valves and pressure gauges	Work accident	3	TE1 Realization	Steam pressure is not zero when the sensor is dismantled
2	Hand contact with the hot pipe surface	Work accident	4	TE1 Realization	Pipes not insulated / PPE incomplete
3	Exposure to hot steam when taking water samples	Work accident	2	TE1 Realization	Not following work procedures
4	Small leak in steam valve	Near Miss	5	TE2 precursor	Seal/valve is worn out
5	Electrocuted while crossing high voltage cables	Work accident	1	TE3 Realization	There are cables that are peeled due to friction
6	Work fatigue due to high temperatures	K3 disorders (heat stress)	2	Outside the top event	Duration of work without adequate rest rotation
	Total		17		

Source: PT Sehat Alam Segar incident reporting system, processed. TE1 realization amounted to 9 incidents (4.50 per year) and TE3 amounted to 1 incident (0.50 per year); TE2 was never realized during the observation period.

Table 4. Basis for determining the annual amount of exposure (demand)

Top Event	Peak event	Definition of one exposure	N	Calculation details
TE1	Exposure to hot steam by workers during maintenance	1 pressure steam line maintenance job	80	52 weekly PM (safety valve test, burner check) + 12 monthly PM (strainer, gauge calibration) + 1 annual PM (overhaul) + 15 corrective/breakdown = 80
TE2	Boiler explosion	1 day of boiler operation	365	Continuous operation 24 hours $\times$ 365 days = 365 operating days (8,760 hours)
TE3	Electrocuted during boiler maintenance	1 electrical work in the boiler area	60	12 monthly panel inspections + 24 electrical correctives + 24 mechanical work assistance = 60

Source: PT Sehat Alam Segar's preventive, corrective, and breakdown maintenance programs, processed.

2.3 Fault Tree Construction and Restructuring

A fault tree was constructed from the hazard identification results for the research object, as described in Section 2.2. Three top events were identified from recorded and potential accident types in the boiler area, then decomposed into intermediate and basic events until an event that could not be further decomposed was reached. The tree structure was then structured into two models, which were explicitly compared. Model A replicates the structure commonly used in FTA applications in practice and in some OHS literature (Ekoanindiyo et al., 2021; Marsudi, 2021), where all basic events are connected to their intermediate events through OR gates, and all intermediate events are connected to the top event also through OR gates. Consequently, all minimal cut sets are of order one.

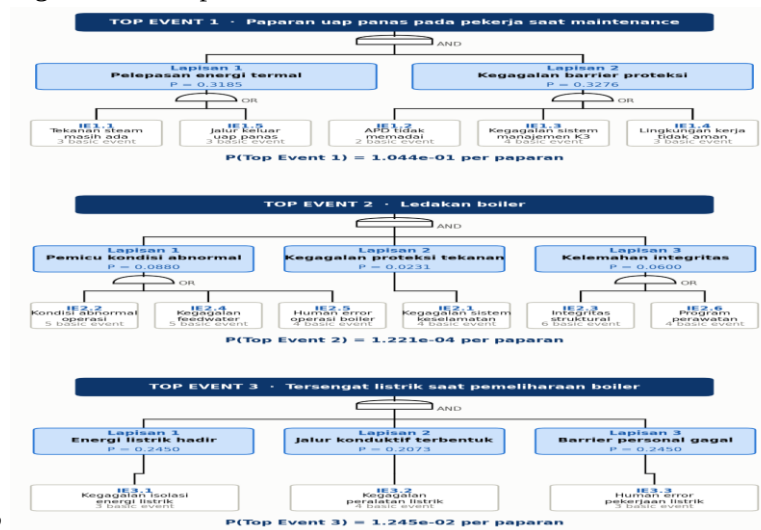
Model B is a restructuring based on the principle of defense in depth, where the top event occurs only if all relevant defense layers fail simultaneously; the connections between layers therefore use AND gates, while connections within a layer still use OR gates. The determination of layers is based on the physical mechanism of each accident, not on mathematical considerations.

In Top Event 1 (exposure to hot steam during maintenance), an accident requires two conditions to occur simultaneously: thermal energy is released at the work site (failure of energy isolation or the presence of a leak path), and the barrier protecting the worker fails (inadequate PPE, failure of the safety management system, or unsafe environmental conditions). If the energy is completely isolated, PPE failure does not result in exposure; conversely, if all barriers are functional, the energy release does not result in injury.

In Top Event 2 (boiler explosion), the explosion requires three conditions to occur simultaneously: the occurrence of an abnormal trigger (abnormal operating conditions, feedwater system failure, or human error in operation), the failure of the pressure protection system that should release excess energy (safety valve, pressure gauge, pressure controller, alarm), and a structural integrity weakness that makes the material unable to withstand the pressure (wall thinning, weld cracks, or negligence in the maintenance program). The overpressure condition that was successfully released by the safety valve did not develop into an explosion; nor did overpressure in a vessel whose integrity was still adequate.

In Top Event 3 (electrical shock during maintenance), the shock requires three conditions to occur simultaneously: electrical energy is still present due to failure of energy insulation, the formation of a conductive path due to electrical equipment failure (stripped cable, damaged insulation, malfunctioning grounding), and the failure of a personal barrier in the form of voltage-free verification and the use of electrical PPE. A live panel with intact cable insulation and functioning grounding does not cause shock to workers wearing electrical PPE.

The minimal cut set is computed through a top-down Boolean expansion: the OR gate produces the union of its children's cut sets, while the AND gate produces the Cartesian product, and then each cut set is tested for minimality by discarding the entire superset. The structure of Model B for the three top events is



presented in Figure 2.

Fig. 2. Fault tree structure of defense layer-based Model B for the three top events

2.4 Probabilistic Quantification

Quantification was performed in three stages. The first stage converted the ordinal likelihood values to annual frequency of occurrence (λ) using the midpoints of the frequency intervals attached to each scale level (Table 2). The company's AS/NZS 4360:2004 likelihood scale already includes explicit frequency anchors—for example, level 3 (Possible) is defined as 3–5 times per year—so this conversion did not introduce new assumptions but rather made explicit the information already contained in the scale.

The second stage converts the annual frequency into a probability per exposure by dividing λ by the number of annual exposures (N) to the relevant top event, as in Equation (1). This form is a relative frequency estimator, namely the proportion of exposures in which the relevant basic event occurs.

$$P_i = \lambda_i / N_j \quad (1)$$

where P_i is the probability of the i -th basic event per exposure, λ_i the annual frequency of the i -th basic event, and N_j the number of annual exposures to the j -th top event.

The third stage aggregates the basic event probabilities to the top event level using Boolean algebra assuming independence between basic events, following Equation (2) for the OR gate and Equation (3) for the AND gate. The annual occurrence frequency of the top event is then obtained from Equation (4), and the annual probability from Equation (5).

$$P(OR) = 1 - \prod (1 - P_i) \quad (2)$$

$$P(AND) = \prod P_i \quad (3)$$

$$f(Te) = P(Te) \times N \quad (4)$$

$$P(annual, Te) = 1 - [1 - P(Te)]^N \quad (5)$$

The resulting probability values are triangulated against generic reference values from industry reliability databases to ensure they are within a reasonable range. For basic human error events, the benchmark is the nominal value of the Generic Task Type in the Human Error Assessment and Reduction Technique (HEART) method as summarized by Bell and Williams (2018), which ranges from 0.0004 for routine tasks performed by highly trained personnel to 0.26 for tasks performed without procedures or supervision. For basic hardware failure events, the benchmark is the component failure rate range in OREDA (2015), for example, failure of a safety relief valve to open on demand on the order of 10^{-3} – 10^{-2} per demand and internal valve leakage on the order of 3×10^{-6} per hour. It should be emphasized that these generic values are used as reasonable limits, not as direct input into the model, because the operating conditions, maintenance quality, and safety culture of the research object differ from the population underlying the database.

Model validation was performed by comparing the model's annual event frequencies against the actual event frequencies for the 2024–2025 period. The model output at the top event level, namely the expected annual event frequencies, was validated, not the probabilities of each basic event, which lacks empirical comparison at the individual level. Goodness-of-fit was tested using a two-tailed Poisson test: the number of observed events over two years was compared against the expected number according to the model. For Top Event 2, which never occurred, goodness-of-fit was assessed based on the probability of observing zero events during the observation period according to each model. The consequences of a model not representing field conditions are explicitly stated in Section 3.2.

2.5 Importance Measure

Importance measures are calculated to determine the structural contribution of each basic event to the probability of the top event, which depends not only on the probability of the basic event itself but also on its position in the tree structure. Four measures are used following Vesely et al. (1981). Fussell–Vesely (FV) measures the proportion of the probability of the top event that is lost if the i -th basic event is confirmed not to occur, thus answering the question of how much risk can be eliminated. Birnbaum (B) measures the rate of change of the probability of the top event with respect to the change of the probability of the basic event, thus indicating the structural sensitivity regardless of the current probability of the basic event. Risk Reduction Worth (RRW) represents the factor by which the probability of the top event decreases if the basic event is completely eliminated, and Risk Achievement Worth (RAW) represents the factor by which the probability of

the top event increases if the basic event is confirmed to occur—this last measure identifies the basic event whose failure would be most detrimental to the protection system. All four measures are calculated according to Equations (6) to (9).

$$FVi = [P(TE) - P(TE | Pi = 0)] / P(TE) \quad (6)$$

$$Bi = P(TE | Pi = 1) - P(TE | Pi = 0) \quad (7)$$

$$RRWi = P(TE) / P(TE | Pi = 0) \quad (8)$$

$$RAWi = P(TE | Pi = 1) / P(TE) \quad (9)$$

Basic events are then ranked according to their FV value and classified using the Pareto principle: the vital few are basic events that cumulatively contribute 80% of the total FV of the top event in question.

2.6 HIRADC and Residual Risk Estimation

The control structure follows the HIRADC framework with a five-level hierarchy as stipulated in ISO 45001:2018, moving from the elimination of the hazard source, its replacement with a safer alternative, engineering interventions on hardware, administrative work procedures, and personal protective equipment as the final layer. Controls are defined for all 53 basic events, with formulation priorities for basic events in the Extreme and High categories as well as basic events with the highest FV ranking. Therefore, the HIRADC in this study does not stand as a separate table, but rather receives direct input from the FTA results: the FV ranking determines the order of control formulation, and the structure of the defense layers determines at which layers controls need to be strengthened. The technical and economic feasibility of the company's actual conditions are taken into consideration in selecting the applied hierarchy layers.

Residual risk is estimated using the hierarchical effectiveness convention explicitly set out in Table 5. This convention embodies the principle that higher layers of the hierarchy act on the source of the hazard, reducing the likelihood of occurrence, while personal protective equipment acts on the recipient of exposure, reducing only the severity. When multiple layers are applied simultaneously, the reductions are cumulative, with a maximum of three levels for likelihood and two levels for severity, and a minimum of one for each dimension. This restriction prevents the assumption that the stacking of controls can completely eliminate the risk. The residual likelihood values are then converted back to probabilities using the same procedure as for the initial conditions, so that the probability of the top event after the controls can be calculated using the same tree structure.

It should be emphasized from the methods section that the residual values generated by this procedure are design estimates, that is, estimates of the level of risk if the proposed controls were implemented as designed. These values are not field measurements, as the proposed controls were not implemented during the study period. This distinction is maintained consistently throughout the presentation of the results.

Table 5. Convention of the effectiveness of the hierarchy of controls on reducing risk scores

Hierarchical layers	L Reduction	S Reduction	Working mechanism
Elimination	-3	-1	Eliminate sources of danger from work systems
Substitution	-2	-1	Replacing components/methods with safer alternatives
Engineering	-2	0	Technical modifications, interlocks, redundancy, automatic protection
Administrative	-1	0	Procedures, work permits, training, supervision, PM schedule
Personal protective equipment	0	-1	The last layer of defense; reducing the severity

Source: ISO 45001:2018 control hierarchy, operationalized by the author. Cumulative reduction is limited to a maximum of -3 for L and -2 for S, with a minimum value of 1. The amount of reduction per layer is the author's convention and has not been calibrated against empirical data.

2.7 Statistical Analysis and Sensitivity

Statistical testing was performed at a 5% significance level. Test selection was based on data type and assumption checking, not on prevalence alone. Risk scores are the product of two ordinal scales and therefore do not meet the requirements for interval scales; tests comparing scores therefore use rank-based

nonparametric statistics. The normality of the difference between the initial and residual risk scores was first tested using Shapiro–Wilk as the basis for choosing between parametric and nonparametric tests.

Because the differences were not normally distributed, the significance of the risk reduction was tested with a one-sided Wilcoxon signed-rank test—a paired test, since each basic event was assessed twice (before and after control)—with an effect size of $r = Z/\sqrt{n}$. Differences in risk scores between three independent top events were tested with Kruskal–Wallis with eta-squared effect sizes. Differences in risk scores between basic events originating from human and management sources and basic events originating from technical and environmental sources were tested with a one-sided Mann–Whitney U with rank-biserial correlation. The relationships between likelihood and severity, as well as between 5×5 matrix scores and Fussell–Vesely values, were tested with Spearman correlation because both are ranked data. Associations between failure sources and top events were tested with chi-square with Cramér's V; all cells of the contingency table were checked for expected frequency. All calculations were performed in Python 3 (NumPy, pandas, and SciPy libraries), and all calculation files are included as appendices for verification of the results.

A sensitivity analysis was performed by varying all λ values for the basic event by $\pm 10\%$ and $\pm 25\%$ relative to the base case, then observing the deviation in the probability of the top event. This analysis tested the robustness of the conclusions to the uncertainty in determining the midpoint of the frequency interval.

III. RESULTS AND DISCUSSION

3.1 Initial Risk Profile of Basic Event

The initial risk assessment of 53 basic events resulted in a distribution of 5 basic events (9.4%) in the Extreme category, 32 basic events (60.4%) in the High category, 16 basic events (30.2%) in the Medium category, and none in the Low category (Table 6). Thus, 69.8% of basic events were in the category requiring high to immediate priority control actions. The absence of the Low category indicates that all identified failure mechanisms in the boiler area have the potential for significant consequences, consistent with the characteristics of the work area which combines thermal energy, pressure, and electricity.

on the source of failure, 26 basic events (49.1%) were technically induced, 14 basic events (26.4%) were human error-induced, 10 basic events (18.9%) were management weaknesses, and 3 basic events (5.7%) were work environment conditions. The chi-square test showed a significant association between the source of failure and the top event ($\chi^2 = 17.955$; $df = 6$; $p = 0.006$; Cramér's V = 0.412), indicating that the three top events had different causal profiles and therefore required different control strategies. Top Event 2 was dominated by technical failures (18 of 28 basic events), while Top Event 3 was evenly divided between technical failures and human errors (5 of 10 basic events each) with no contribution from environmental factors.

The Kruskal–Wallis test showed a significant difference in risk scores between top events ($H = 9.803$; $df = 2$; $p = 0.007$; $\eta^2 = 0.156$). The highest median risk score was found in Top Event 3 (10.5), followed by Top Event 1 (9.0) and Top Event 2 (8.0). This finding is interesting because it contradicts common intuition: electrical work in the boiler area, which is often perceived as support work, actually has the highest risk profile per basic event.

The Mann–Whitney U test showed that basic events originating from human and management factors had a significantly higher risk score than basic events originating from technical and environmental factors ($U = 506$; $p = 0.002$; mean 10.29 versus 7.55; rank-biserial correlation = 0.454). This confirms that in the research object, weaknesses in the behavioral aspects and safety management systems contributed a greater level of risk than hardware reliability.

One noteworthy finding at this stage is the lack of correlation between likelihood and severity values ($\rho = 0.017$; $p = 0.906$). The independence of these two dimensions is methodologically important. Cox (2008) showed that risk matrices can produce worse decisions than random selection precisely when likelihood and severity are negatively correlated, because under these conditions the matrix rankings can reverse the quantitative risk rankings. The results of the test in this case did not find such a correlation, so the use of the

5×5 matrix as an initial assessment tool can still be maintained. However, the limitations of the ordinal scale resolution that Cox also noted remain, and this is precisely why further quantification is necessary in this study.

Table 6. Distribution of risk categories of 53 basic events before and after control

Risk categories	TE1	TE2	TE3	Initial total	% beginning	Total residual	% residual
Extreme	2	1	2	5	9.4%	0	0.0%
High	10	15	7	32	60.4%	1	1.9%
Medium	3	12	1	16	30.2%	29	54.7%
Low	0	0	0	0	0.0%	23	43.4%
Total	15	28	10	53	100.0%	53	100.0%

Description: Extreme = 15–25; High = 8–14; Medium = 4–7; Low = 1–3. The residual column is the result of the control design estimation, not the post-implementation measurement.

3.2 Top Event Probability and Model Validation

The top event probabilities were calculated for both models and compared against the actual event frequencies for the 2024–2025 period (Table 7, Figure 3). Model A yielded an expected frequency of 43.34 steam exposure events per year, 59.34 boiler explosions per year, and 32.89 electrocution events per year. These figures exceeded the actual occurrences by 9.6 times for Top Event 1 and 65.8 times for Top Event 3. For Top Event 2, Model A predicted 118.68 explosions over the two years of observation, so the probability of observing zero explosions according to the model was only 2.9×10^{-52} —practically impossible. Since there were no boiler explosions in the two years, Model A was decisively rejected by the data.

Model B produces estimates that are much closer to reality. The expected frequency of steam exposure is 8.35 events per year versus 4.50 actual events (ratio 1.86), and electric shock is 0.747 events per year versus 0.50 actual events (ratio 1.49). A two-tailed Poisson test does not reject model fit at the 5% level for either top event (Top Event 1: expected 16.70 events in two years versus 9 observed events, $p = 0.061$; Top Event 3: expected 1.49 events versus 1 observed event, $p = 1.000$). For Top Event 2, Model B predicts an expected 0.089 explosions over two years, giving a 91.5% probability of observing zero events, consistent with field observations.

These results need to be interpreted with caution. Model B is not proven correct; what can be stated is that Model B's output is consistent with the actual incident data and does not show statistically significant differences from it, while Model A shows differences that cannot be explained by random variation. This validation also has three limitations that need to be stated openly. First, the observation period is only two years with a small number of events, so the Poisson test has low power; the p -value of 0.061 for Top Event 1 is close to the significance level and indicates the model tends to be somewhat conservative. Second, the validation was performed at the aggregate top event level, so the accuracy of the probabilities of each basic event was not validated. Third, for Top Event 2, the validation consisted only of a test of consistency with respect to the absence of an event, which logically could be met by many other low probability values besides the one generated by the model.

The consequences of models that do not adequately represent field conditions are immediate and asymmetrical. Models that overestimate probabilities, such as Model A, encourage companies to spread resources across all causes indiscriminately and erase the distinction between hazardous conditions and actual accidents. Conversely, models that underestimate probabilities delay necessary control measures. Therefore, the probability figures in this study are used to establish priorities and compare conditions before and after control measures, not as predictions of the number of future accidents.

It should be emphasized that both models use identical basic event probability values; the only difference lies in the logical structure of the fault tree. Therefore, the deviation in Model A stems not from data assignment errors but solely from structural modeling errors. The all-OR gate structure implicitly implies that there is no effective defense layer in the boiler system, so any deviation, no matter how small, immediately leads to an accident.

The final probability of the top event according to Model B is 1.044×10^{-1} per pressurized steam line maintenance job, 1.221×10^{-4} per operating day (equivalent to 4.36×10^{-2} per year), and 1.245×10^{-2} per boiler area electrical job. Practical interpretation: approximately one in every ten steam line maintenance jobs has the potential to cause exposure to hot steam, and approximately one in every eighty electrical jobs has the potential to cause electric shock. For boiler explosions, the frequency of 4.36×10^{-2} per year is above the individual worker risk acceptance limit of 1×10^{-3} per year set by the Health and Safety Executive (2001) as the upper limit of the tolerable area, so quantitatively it does require additional control even though the event has never been realized.

The decomposition into layers of defense (Table 8) reveals which layer is weakest for each top event. In Top Event 1, both layers have nearly equal failure probabilities (0.3185 and 0.3276), so no single layer can be relied upon as the primary barrier. In Top Event 2, the pressure protection layer is the strongest (0.0231), while the abnormal condition trigger is the weakest (0.0880), meaning the boiler instrumentation system is relatively reliable, but abnormal conditions occur too frequently. In Top Event 3, all three layers are equally weak (0.2073–0.2450), which explains why this top event has the highest median risk score.

The tree restructuring also fundamentally changes the character of the minimal cut sets. Model A produces 53 minimal cut sets, all of which are of order one, so that each basic event is a single point of failure. Model B produces 650 minimal cut sets: 54 of order two in Top Event 1, and 560 and 36 of order three in Top Events 2 and 3, respectively. The absence of order one cut sets in Model B means that no single failure directly leads to an accident—a characteristic much more in line with the design of a real boiler safety system. Conversely, the conclusion that all basic events are single points of failure—which follows automatically from the structure of Model A—is untenable.

Table 7. Probability of top events for Model A and Model B and validation against actual incident data

Top Event	Total BE	N/year	P(TE) Model A	P(TE) Model B	f Model A (/th)	f Model B (/th)	actual f (/th)	Ratio A	Ratio B
TE1	15	80	0.5418	0.1044	43.34	8,349	4.50	9.6×	1.86×
TE2	28	365	0.1626	1.221×10^{-4}	59.34	0.045	0.00	—	—
TE3	10	60	0.5482	0.0124	32.89	0.747	0.50	65.8×	1.49×

Description: P(TE) TE1 and TE3 per job, TE2 per operating day. Ratio = model frequency divided by actual frequency; TE2 cannot be ratioed because it has never occurred (see Poisson test in the text).

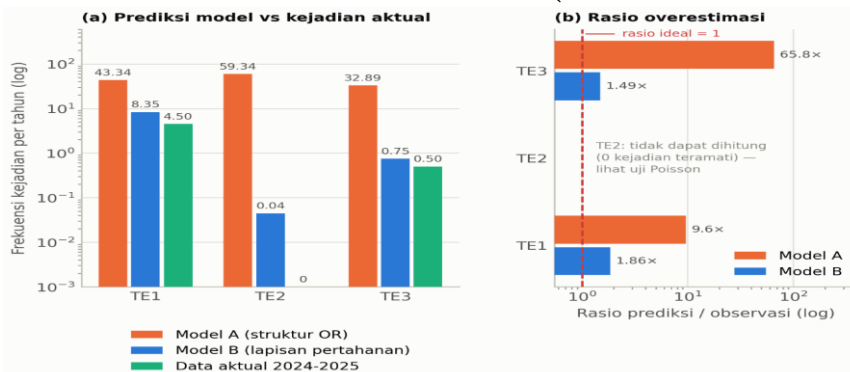


Fig. 3. Model validation against actual incident data for the period 2024–2025

Table 8. Probability of failure of each defense layer (Model B)

TE	Layer	Layer description	Total BE	P start	P residual
TE1	Layer 1	Thermal energy release (IE1.1 + IE1.5)	6	0.3185	0.0490
TE1	Layer 2	Protection barrier failure (IE1.2 + IE1.3 + IE1.4)	9	0.3276	0.1240
TE2	Layer 1	Abnormal condition triggers (IE2.2 + IE2.4 + IE2.5)	14	0.0880	0.0244

TE	Layer	Layer description	Total BE	P start	P residual
TE2	Layer 2	Failure of pressure protection system (IE2.1)	4	0.0231	0.0055
TE2	Layer 3	Structural integrity & maintenance weaknesses (IE2.3 + IE2.6)	10	0.0600	0.0190
TE3	Layer 1	Non-isolated electrical energy (IE3.1)	3	0.2450	0.0248
TE3	Layer 2	Conductive path is formed (IE3.2)	4	0.2073	0.0329
TE3	Layer 3	Personal barrier failure (IE3.3)	3	0.2450	0.0573

Description: P residual is an estimate of the control design. The probability of the top event is obtained from multiplying the probabilities of all layers (AND gate).

3.3 Importance-Based Control Priorities

Importance measures were calculated in Model B to determine control priorities (Table 9, Figure 4). The three basic events with the highest Fussell–Vesely values were failure to implement Lock Out Tag Out on electrical work (FV = 0.4740), failure to check voltage before work (FV = 0.4740), and inaccurate pressure gauge readings (FV = 0.4685). The Risk Reduction Worth values for these three were 1.901, 1.901, and 1.882, respectively, meaning that eliminating just one of them reduced the probability of the associated top event by approximately half.

The most important finding at this stage is the discrepancy between the rankings based on matrix scores and those based on structural contributions. The Spearman correlation between the 5×5 risk score and the Fussell–Vesely value is significant ($\rho = 0.651$; $p < 0.001$), but only explains 42.4% of the variance in the rankings. This discrepancy is particularly evident in Top Event 2: the basic event 'operator fails to respond to high pressure alarm' received the highest matrix score (15, Extreme category) but was ranked only eighth based on Fussell–Vesely (FV = 0.1148), while 'pressure gauge shows inaccurate reading' received only a score of 9 (High category) but was ranked first (FV = 0.4685). The reasons are structural: the pressure gauge is located in the pressure protection layer, which has only four basic events, so its failure accounts for a large portion of the overall failure layer; conversely, the operator's response failure is located in the trigger layer, which has fourteen basic events, so its contribution is divided. The basic event pressure gauge also had the highest Risk Achievement Worth across the entire study (RAW = 43.29), meaning that if this component were confirmed to fail, the probability of a boiler explosion would increase more than fortyfold.

The practical implication is clear: prioritizing control based solely on matrix score rankings—as is common in HIRA/HIRARC practice—risks allocating resources to the most frightening basic events, but not those that most determine the probability of an accident. This is precisely the key added value of integrating quantitative FTA into the HIRADC framework.

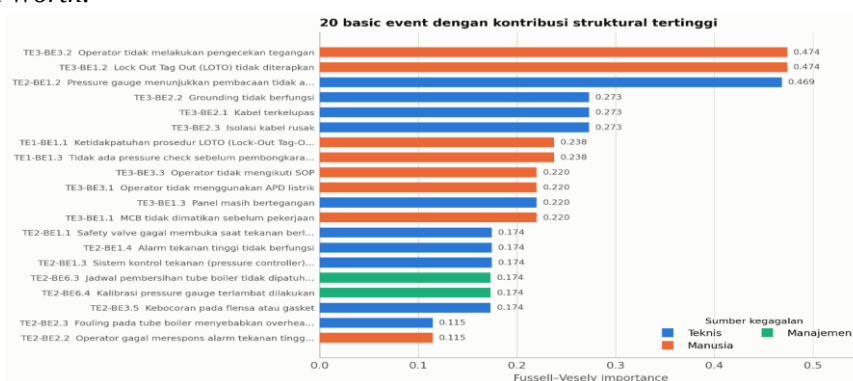
The Fussell–Vesely analysis by failure source (Figure 6b) shows that overall, technical factors accounted for 45.9% of the structural contribution, human factors 37.1%, management factors 14.5%, and environmental factors 2.5%. Thus, the combined human and management factors accounted for 51.6%, or more than half of the total contribution. However, this composition differed significantly across top events: in Top Event 1, the combined human and management factors reached 73.2%, in Top Event 3, human factors alone contributed 58.6%, while in Top Event 2, technical factors dominated with 67.5%. This difference reinforces the results of the previous chi-square test and confirms that control strategies cannot be standardized across boiler areas.

Pareto analysis yields results that differ from the commonly assumed 80/20 pattern. The vital few, which account for 80% of the Fussell–Vesely total, comprise 10 of the 15 basic events (66.7%) in Top Event 1, 16 of the 28 basic events (57.1%) in Top Event 2, and 7 of the 10 basic events (70.0%) in Top Event 3. The risk in the boiler area is therefore relatively evenly distributed, rather than concentrated in a handful of causes. The managerial implications are significant: partial improvements at just a few critical points will not significantly reduce the probability of an accident, so the control program must be comprehensive across the weak layers.

Table 9. Fifteen basic events with the highest structural contribution (Model B)

No	Code	Basic event	Score (cat.)	Pi	FV	Birnbaum	RRW	RAW
1	TE3-BE1.2	Lock Out Tag Out (LOTO) is not implemented	16 (E)	0.1333	0.4740	0.0442	1,901	4.08
2	TE3-BE3.2	The operator did not check the voltage.	16 (E)	0.1333	0.4740	0.0442	1,901	4.08
3	TE2-BE1.2	Pressure gauge shows inaccurate readings	9 (H)	0.0110	0.4685	0.0052	1,882	43.29
4	TE3-BE2.1	Peeled cable	9 (H)	0.0667	0.2732	0.0510	1,376	4.82
5	TE3-BE2.2	Grounding is not working	12 (H)	0.0667	0.2732	0.0510	1,376	4.82
6	TE3-BE2.3	Damaged cable insulation	9 (H)	0.0667	0.2732	0.0510	1,376	4.82
7	TE1-BE1.1	Non-compliance with LOTO (Lock-Out Tag-Out) procedures	16 (E)	0.1000	0.2377	0.2481	1,312	3.14
8	TE1-BE1.3	No pressure check before component disassembly	16 (E)	0.1000	0.2377	0.2481	1,312	3.14
9	TE3-BE1.1	MCB is not turned off before work	12 (H)	0.0667	0.2201	0.0411	1,282	4.08
10	TE3-BE1.3	The panel is still energized	12 (H)	0.0667	0.2201	0.0411	1,282	4.08
11	TE3-BE3.1	Operators do not use electrical PPE	9 (H)	0.0667	0.2201	0.0411	1,282	4.08
12	TE3-BE3.3	Operator does not follow SOP	9 (H)	0.0667	0.2201	0.0411	1,282	4.08
13	TE2-BE1.1	Safety valve fails to open under excessive pressure	10 (H)	0.0041	0.1745	0.0052	1,211	43.29
14	TE2-BE1.3	Pressure controller malfunction	8 (H)	0.0041	0.1745	0.0052	1,211	43.29
15	TE2-BE1.4	High pressure alarm does not work	8 (H)	0.0041	0.1745	0.0052	1,211	43.29

Note: E = Extreme; H = High; M = Medium. FV = Fussell–Vesely; RRW = Risk Reduction Worth; RAW = Risk Achievement Worth.

**Fig. 4. Ranking of 20 basic events according to Fussell–Vesely importance**

3.4 HIRADC and Residual Risk Reduction

Controls were established for all 53 basic events following the ISO 45001:2018 hierarchy; a summary of the 18 most contributing basic events is presented in Table 12, while the complete HIRADC matrix is included in the appendix. The order of control formulation followed the Fussell–Vesely ranking, so that the basic event that most determines the probability of the top event is assigned to the highest technically and economically feasible hierarchy layer. Elimination opportunities were identified at only one point, namely the

relocation of the voltage measurement point outside the panel enclosure, eliminating the need to open the live panel for routine inspections. Substitutions were implemented at five points, primarily the replacement of manual isolation valves with motorized valves with electronic position indicators, the replacement of portable multimeter-based voltage verification with permanent voltage indicators, and the improvement of water treatment chemical formulations. The majority of controls were a combination of engineering and administrative controls (29 basic events); overall, the administrative layer touched 49 basic events and the engineering layer 36 basic events, in keeping with the fact that boilers are essential utilities that cannot be removed from the production process.

Implementation of the control design is estimated to reduce the average risk score from 8.79 to 3.83, or an aggregate reduction of 56.4% (Figure 8). The distribution of risk categories shifted across the board: the Extreme category decreased from 5 to zero, the High category from 32 to 1, while the Medium category increased to 29 and the Low category—which previously had none—to 23 basic events (Figure 5, Figure 6a). The proportion of basic events requiring high-priority or immediate action decreased from 69.8% to 1.9%. All residual figures in this section are estimates based on the hierarchical effectiveness convention in Table 5, that is, the conditions expected to be achieved if controls were implemented as designed, not the conditions measured in the field.

One basic event that remained in the High category after control was 'operator not following boiler operating SOP' (score 12 to 8). This basic event can only be controlled through the administrative layer because it is directly related to behavioral compliance, whereas the administrative layer, according to the convention used, only lowers the likelihood by one level. This finding is not a weakness of the analysis but rather a logical consequence of the hierarchy of controls principle: administrative controls have the lowest effectiveness among the layers acting on the hazard source, so behavioral risk cannot be reduced to a low level without engineering support. For companies, this provides a quantitative argument for complementing SOP enforcement with technical support such as interlocks and operator decision support systems.

At the system level, the reduction in the risk of basic events leads to a much larger reduction in the probability of top events due to the multiplicative effect between the layers of defense. The probability of steam exposure is estimated to decrease from 1.044×10^{-1} to 6.080×10^{-3} per job (a 94.2% or 17.2-fold decrease), the probability of boiler explosion from 1.221×10^{-4} to 2.536×10^{-6} per operating day (a 97.9% or 48.1-fold decrease), and the probability of electric shock from 1.245×10^{-2} to 4.676×10^{-5} per job (a 99.6% or 266.2-fold decrease). Expressed in annual frequency, the probability of a boiler explosion drops from 4.36×10^{-2} to 9.25×10^{-4} per year, moving from just above the acceptable limit of 1×10^{-3} per year (Health and Safety Executive, 2001) to just below. It should be noted that just below this limit leaves a narrow margin, so the ALARP area still requires regular review. The difference in the magnitude of the reduction between top events reflects the number of layers of defense: the more layers reinforced, the greater the multiplicative effect.

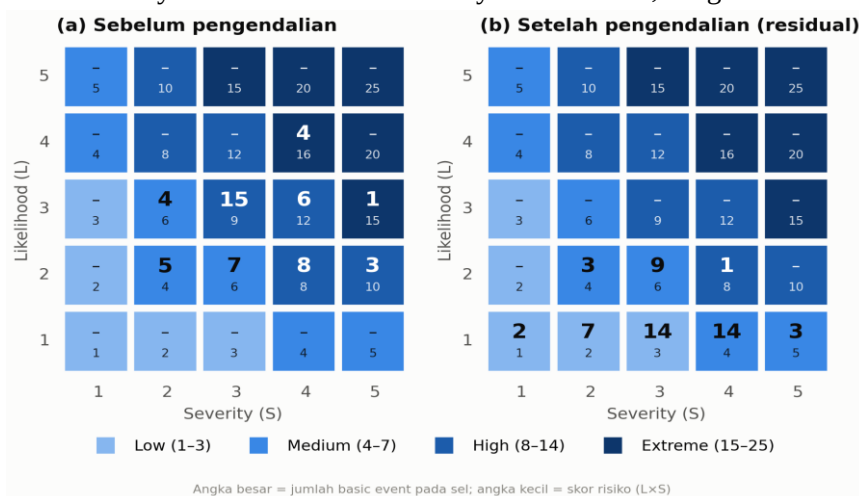


Fig. 5. Distribution of 53 basic events in the 5x5 risk matrix before and after control

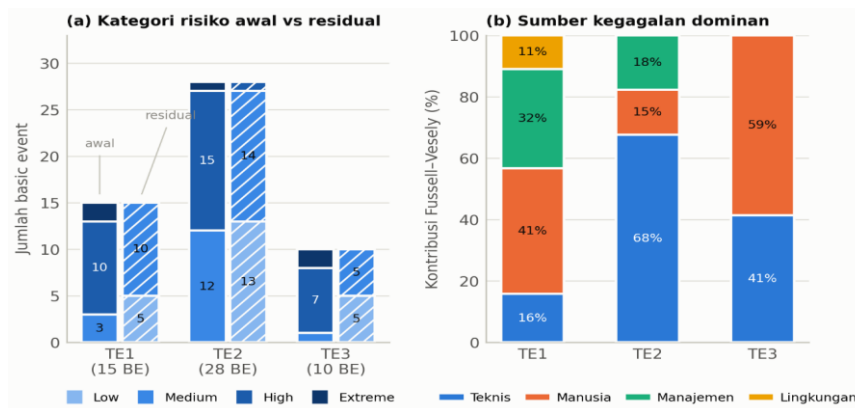


Fig. 6. Distribution of risk categories and contribution of failure sources per top event

Details of the risk score reduction for all 53 basic events are presented in Figure 8, while a summary of HIRADC and recommended controls are presented in Table 12 (both placed on landscape-oriented pages at the end of the manuscript).

3.5 Statistical Testing of Control Design Effectiveness

The Shapiro–Wilk test for the difference between the initial and residual risk scores indicated a non-normal distribution ($W = 0.824$; $p < 0.001$), so the test was continued with nonparametric statistics. A one-sided Wilcoxon signed-rank test showed a significant decrease in risk scores ($Z = 6.360$; $p < 0.001$), with a median decrease from 9.0 to 4.0 and a mean decrease per basic event of 54.5%. The effect size of $r = 0.874$ is considered very large by Cohen's criteria ($r \geq 0.5$ is categorized as a large effect). It should be noted that the effect size in this paired test reflects the consistency of the direction of the decrease across all pairs, not the magnitude of the risk decrease in absolute units. All 53 basic events experienced a decrease in scores without exception, so there were no zero-value pairs in the test; this uniformity of direction resulted in the high r value. A summary of all statistical tests is presented in Table 10.

The following distinction needs to be made as explicitly as possible. This test measures the significance of risk reduction according to the estimated model, not the empirical effectiveness of the controls after actual implementation. The residual values are derived from the hierarchical effectiveness convention in Table 5, so the Wilcoxon test results indicate that the proposed control design consistently targets the correct risk dimensions across all basic events. These results do not prove that risk reduction has occurred in practice, and should not be interpreted as evidence of control effectiveness. Some of the uniformity of the results is even an arithmetic consequence of the convention used: since each basic event receives at least one layer of control, each pair is guaranteed to experience a reduction. Empirical verification requires re-measuring risk scores and recording incidents after the controls are actually implemented, which is a matter for further research.

Table 10. Summary of statistical test results

Statistical test	Purpose of testing	Statistics	p	Interpretation
Shapiro–Wilk (initial–residual risk difference)	Testing the normality of differences	$W = 0.824$	< 0.001	Not normal → nonparametric test is used
Wilcoxon signed-rank (one-sided)	Initial risk > residual risk (model estimate)	$Z = 6,360$	< 0.001	Significant; $r = 0.874$ (large effect)
Kruskal–Wallis	Difference in risk scores between top events	$H = 9.803$	0.007	Significant; $\eta^2 = 0.156$ (medium effect)
Mann–Whitney U (one-sided)	Human factor risk + management > technical + environmental	$U = 506$	0.002	Significant; $r_{rb} = 0.454$
Spearman Correlation	Relationship between likelihood and severity	$\rho = 0.017$	0.906	Not significant → L and S are independent
Spearman Correlation	The relationship between the 5×5 matrix scores and Fussell–Vesely	$\rho = 0.651$	< 0.001	Significant but only explains 42.4% of the variance

Statistical test	Purpose of testing	Statistics	p	Interpretation
Chi-square	Association of failure sources with top events	$\chi^2 = 17.955$	0.006	Significant; Cramér's V = 0.412
Two-tailed Poisson test (TE1)	Fit of Model B frequency to actual data	16.70 vs 9	0.061	Did not show any significant difference
Two-tailed Poisson test (TE3)	Fit of Model B frequency to actual data	1.49 vs 1	1,000	Did not show any significant difference

Description: 5% significance level. r = Wilcoxon effect size; η^2 = eta-squared; r_{rb} = rank-biserial correlation. The Wilcoxon test tests the consistency of the control design according to the estimated model, not the empirical effectiveness after implementation.

3.6 Sensitivity Analysis

Sensitivity analysis (Table 11, Figure 7) shows that varying all λ values by $\pm 25\%$ results in a probability deviation of -39.2% to $+44.6\%$ for Top Event 1, -57.0% to $+91.6\%$ for Top Event 2, and -55.1% to $+83.3\%$ for Top Event 3. The largest deviations occur for Top Events 2 and 3, which have three layers of defense, because the probability changes at each layer are multiplied. However, all deviations remain within one order of magnitude relative to the base case, and the order of magnitude of probabilities between top events remains unchanged across scenarios.

More importantly, the main conclusion of the study is unchanged by this variation. Even in the most extreme scenario (λ increased by 25%), the probability of the top event of Model B remains significantly below that of Model A: the expected frequency of Top Event 3 in this scenario is 1.37 events per year compared to 32.89 events per year in Model A. Thus, the rejection of the entire OR gate structure is robust to uncertainty in the determination of the value of λ .

Table 11. Sensitivity analysis of top event probability to variations in the value of λ

Variation of λ	P(TE1)	Deviation	P(TE2)	Deviation	P(TE3)	Deviation
-25%	0.0635	-39.2%	5.248×10^{-5}	-57.0%	0.0056	-55.1%
-10%	0.0872	-16.4%	8.966×10^{-5}	-26.5%	0.0093	-25.2%
+0%	0.1044	+0.0%	1.221×10^{-4}	+0.0%	0.0124	+0.0%
+10%	0.1224	+17.3%	1.612×10^{-4}	+32.1%	0.0161	+29.8%
+25%	0.1509	+44.6%	2.339×10^{-4}	+91.6%	0.0228	+83.3%

Note: the base case (+0%) is the midpoint λ value of the frequency interval in Table 2.

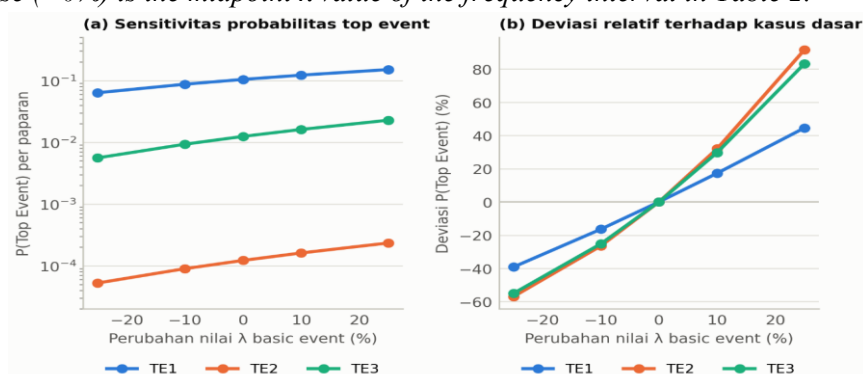


Fig. 7. Sensitivity of the top event probability to variations in the basic event λ value

3.7 Discussion

The primary methodological contribution of this study is the quantitative demonstration that fault tree structures that do not represent layers of defense produce systematically incorrect risk estimates, not simply less precise ones. With identical basic event probabilities, the mere difference in structure results in a 65.8-fold difference in estimates for Top Event 3 and a statistically improbable prediction for Top Event 2. The reason for this can be summarized succinctly: OR gates imply that the layer probability is approximately the sum of the probabilities of its members, so that adding a basic event always increases the probability of the top

event, while AND gates imply multiplication, so that the probability of the top event actually decreases with each layer. Structuring the entire tree with OR gates therefore turns each additional hazard identification result—which should improve the model—into an inflation of the risk estimate.

This finding is relevant beyond the research object, as constructing fault trees with OR gates is a common pattern in FTA applications in practice and in some Indonesian OSH literature, as seen in Ekoanindiyo et al. (2021) and Marsudi (2021), who used FTA as a cause mapping tool rather than as a system logic model. The root of the problem lies in the difference between the questions 'what contributed to this accident' and 'what combination of failures must occur for this accident to occur'. OR gates answer the first question and produce a list of causes; AND gates are needed to answer the second question and produce a model. When FTA is used qualitatively, this difference is not apparent because both produce the same list of basic events. The difference is only revealed when probabilities are calculated and compared with reality—which is precisely why quantification is valuable, despite the availability of perfect component probability data.

The implications for the company are immediate. Model B shows that the strength of a boiler safety system lies in the existence of complementary layers, so that investments in strengthening one already strong layer provide significantly less benefit than investments in improving the weakest layer. In Top Event 2, for example, the pressure protection layer was already relatively reliable (0.0231) while the abnormal condition trigger layer was the weakest (0.0880), so adding pressure instrumentation would provide a lower return than improvements in operating conditions and feedwater control.

The second contribution is the evidence that priority rankings based on the 5×5 matrix and those based on structural contribution are not interchangeable. The correlation between the two explains only 42.4% of the variance in rankings, and in the case of Top Event 2, there is a sharp reversal in ranking between the Extreme and High basic events. This is due to the nature of the risk matrix, which assesses each hazard in isolation: the matrix does not know how many other paths are available to the same top event, whereas the importance measure does. This finding aligns with Cox's (2008) critique of the limited resolution of risk matrices, while extending it: the matrix's limitations stem not only from the compression of ranges at the ordinal scale, but also from the lack of structural information about the hazard's position in the system. These findings thus provide a quantitative justification for the integration of FTA into HIRADC, complementing the conceptual arguments previously presented in the literature.

The third contribution is substantive to the research object. Human and management factors accounted for 51.6% of the structural contribution and had a significantly higher risk score than technical factors ($p = 0.002$). The two basic events with the highest Fussell–Vesely risk, namely, failure to implement LOTO and failure to perform voltage checks, are both procedural compliance failures that do not require significant equipment investment to correct. They are also at different layers in the Top Event 3 structure—LOTO at the energy isolation layer and voltage checks at the personal barrier layer—so that their corrections work multiplicatively, not substitute for each other. Simultaneously, the only basic event that remained in the High category after control was also a compliance failure. These three findings lead to a consistent conclusion: the greatest leverage for PT Sehat Alam Segar is not in hardware replacement but in enforcing energy isolation discipline, the effectiveness of which needs to be supported by engineering in the form of interlocks so as not to rely entirely on voluntary compliance.

The Pareto analysis results offer a counterbalance. The vital few in this study account for 57.1% to 70.0% of basic events, rather than around 20% as in the classical Pareto pattern. This is due to the relative homogeneity of basic event probabilities resulting from the use of a five-tiered likelihood scale, which yields only three distinct λ values for the majority of basic events. This finding also highlights the limitations of the approach: anchoring frequencies on an ordinal scale does allow for quantification without a reliability database, but its resulting resolution is limited and tends to flatten differences between basic events. Consequently, the managerial strategy of 'fixing the most important few' is inadequate in this case; a significant reduction in the probability of top events requires a comprehensive strengthening of the defense layer, as reflected in the residual risk estimation results.

Set against the common practice of FTA implementation in industrial settings, the results of this study correct three assumptions that are often accepted without examination. First, the assumption that every basic event is at least a first-order cut set—an automatic consequence of constructing a tree with OR gates throughout—is untenable; Model B, consistent with the incident data, does not contain a single first-order cut set. Second, control priorities cannot be determined from the matrix score rankings alone, as these rankings explain only 42.4% of the variance in the structural contribution rankings. Third, the limitation commonly stated as 'quantitative analysis is not possible due to limited component probability data' is proven to be overcome by utilizing the frequency anchoring inherent in the company's likelihood scale, without requiring a complete component reliability database.

3.8 Research Limitations

Five limitations need to be explicitly stated. First, the basic event probabilities are derived from a likelihood scale based on expert judgment, not from component failure records. Although the conversion is transparent through the scale's inherent frequency anchoring and the results are consistent with aggregate incident data, accuracy at the individual basic event level cannot be guaranteed. Generic industry reference values are used as reasonable limits, not as direct input.

Second, likelihood and severity ratings were recorded only as final aggregate scores from three informants and not as individual ratings, so inter-rater reliability (e.g., through Kendall's coefficient of concordance) could not be calculated. Further research is recommended that records individual ratings to measure inter-rater consistency.

Third, the residual risk is estimated based on the hierarchy of effectiveness convention, not measured after implementation. This convention reflects the widely accepted principle of the hierarchy of controls, but the reduction per layer was determined by the author and has not been calibrated against empirical data. Actual effectiveness in the field depends on the quality of implementation, and the statistical tests performed therefore test the consistency of the control design, not its empirical effectiveness.

Fourth, model validation was conducted on two years of incident data with a small number of events, so the Poisson test's power is low, and the absence of rejection cannot be interpreted as confirmation of model accuracy. Validation is also aggregated at the top event level and does not extend to the accuracy of individual basic event probabilities.

Fifth, the model assumes independence between basic events. In practice, common cause failures can occur, such as weak safety culture that simultaneously reduces LOTO compliance, PPE use, and voltage inspection performance. The presence of common cause failures would increase the probability of the top event above the calculated value, so the estimates in this study are lower bounds for this scenario. Modeling with beta factors or equivalent approaches is recommended for further research.

IV. Conclusion

study quantifies occupational safety risks in the boiler area of PT Sehat Alam Segar through the integration of quantitative Fault Tree Analysis with HIRADC. The four research objectives were achieved with the following results.

The first objective, quantifying the probability of top events and testing the fit of the fault tree structure to the actual data, was achieved with the finding that the fault tree structure determined the validity of the risk estimates. With identical basic event probability values, the all-OR gate structure overestimated the event frequency by a factor of 9.6 for steam exposure and 65.8 for electric shock, and resulted in a prediction of 118.68 boiler explosions in two years that was completely rejected by the data. The layered defense structure produced estimates within a factor of 1.49–1.86 of the actual events and showed no significant difference according to the Poisson test ($p = 0.061$ and $p = 1.000$). The top event probabilities consistent with the actual data are 1.044×10^{-1} per steam line maintenance job, 1.221×10^{-4} per boiler operating day (4.36×10^{-2} per year), and 1.245×10^{-2} per electrical job. In the data-consistent structure, there is not a single minimum cut set of order one, so the assumption that all basic events are single points of failure cannot be maintained.

The second objective, determining the basic events with the greatest structural contribution, was achieved by finding that control priorities based on structural contribution differed significantly from priorities based on 5×5 matrix scores; the correlation between the two explained only 42.4% of the variance in rankings. The three basic events with the highest contributions were failure to implement LOTO (FV = 0.4740), failure to check voltage (FV = 0.4740), and inaccurate pressure gauge readings (FV = 0.4685; RAW = 43.29). Human and management factors contributed 51.6% of the overall structural contribution, with significantly different compositions across top events ($\chi^2 = 17.955$; $p = 0.006$).

The third objective, developing a HIRADC control plan and its risk reduction estimates, was achieved with a design that eliminated all Extreme categories and left one basic event in the High category, reducing the proportion of high-priority basic events from 69.8% to 1.9%, and reducing the average risk score from 8.79 to 3.83. At the system level, the probability of the top event was estimated to decrease by 17.2 times; 48.1 times; and 266.2 times, respectively, so that the frequency of boiler explosions moved from 4.36×10^{-2} to 9.25×10^{-4} per year. All residual values are design estimates, not post-implementation measurement results.

The fourth objective, testing the statistical significance of the risk reduction, was achieved with a significant Wilcoxon signed-rank test result and a very large effect size ($Z = 6.360$; $p < 0.001$; $r = 0.874$). This test assesses the consistency of the control design according to the estimation model, not the empirical effectiveness of the control in the field. The main conclusions of the study were proven robust to variations in the λ value of $\pm 25\%$ in the sensitivity analysis.

For PT Sehat Alam Segar, the first implementation priority is the enforcement of energy isolation procedures (LOTO) supported by isolation valve interlocks and permanent voltage indicators, followed by redundancy and scheduled calibration of pressure instrumentation. For scientific development, this study shows that FTA quantification can be carried out in companies that do not yet have a component reliability database, as long as the likelihood scale used includes frequency anchoring and the results are validated against actual incident data. Further research is recommended to measure the effectiveness of controls empirically after implementation, record each informant's assessment to measure inter-rater reliability, and model common cause failures.

Thank-you note

The author would like to thank the management and all staff of the Utilities Department of PT Sehat Alam Segar for permission, data, and assistance during the implementation of the research, as well as to the supervisor of the Industrial Engineering Study Program, Muhammadiyah University of Gresik for direction and corrections in the preparation of this article.

REFERENCE

- [1]. Alizadeh, S.S., & Moshashaei, P. (2015). The Bowtie method in safety management systems: A literature review. *Scientific Journal of Reviews*, 4(9), 133–138.
- [2]. Bell, J. L., & Williams, J. C. (2018). Evaluation and consolidation of the HEART human reliability assessment principles (pp. 3–12). https://doi.org/10.1007/978-3-319-60645-3_1
- [3]. Cox, L.A., Jr. (2008). What's wrong with risk matrices? *Risk Analysis*, 28(2), 497–512. <https://doi.org/10.1111/j.1539-6924.2008.01030.x>
- [4]. Ekoanindiyo, FA, Yohanes, A., Prihastono, E., & Hayati, EN (2021). The fault tree analysis method approach in occupational health and safety and its impact on productivity. *Journal of Industrial Systems Engineering*, 7(1), 20–25. <https://doi.org/10.33884/jrsi.v7i1.4456>
- [5]. Health and Safety Executive. (2001). Reducing risks, protecting people: HSE's decision-making process. HSE Books. ISBN 0-7176-2151-0.
- [6]. Henny, H., Budi, AHS, Andriyansyah, M., Ar Rozzak, MR, Baru, MM, & Masek, A. (2025). Hazard identification, risk assessment, and determining control (HIRADC) for workplace safety in manufacturing industry: A risk-control framework complete with bibliometric literature review analysis to support Sustainable Development Goals (SDGs). *ASEAN Journal for Science and Engineering in Materials*, 4(2), 267–284.
- [7]. International Organization for Standardization. (2018a). ISO 31000:2018 Risk management — Guidelines.

- [8]. International Organization for Standardization. (2018b). ISO 45001:2018 Occupational health and safety management systems — Requirements with guidance for use.
- [9]. Ministry of Manpower of the Republic of Indonesia. (2016). Regulation of the Minister of Manpower of the Republic of Indonesia Number 38 of 2016 concerning Occupational Safety and Health of Aircraft Power and Production.
- [10]. Mahyunis, M., Effendi, Z., & Asrianto, F. (2023). Analysis of work accident risks at the boiler station at the PTPN VI-South Solok Unit PKS using the HIRARC method. *IRA Journal of Mechanical Engineering and Its Applications (IRAJTMA)*, 2(2), 37–46. <https://doi.org/10.56862/irajtma.v2i2.57>
- [11]. Marsudi, S. (2021). Boiler system failure risk analysis using fault tree analysis. *Marine Science and Technology Research*, 81–84. <https://doi.org/10.62012/sensistek.v4i1.19751>
- [12]. OREDA. (2015). OREDA: Offshore and onshore reliability data handbook (6th ed.). OREDA Participants.
- [13]. Dutch East Indies Government. (1930). Steam Act 1930 (Stoom Ordonnantie 1930) and Steam Regulations 1930 (Stoom Verordening 1930).
- [14]. Rahmanto, I., & Ihsan Hamdy, M. (2022). Analysis of employee occupational accident risk using the Hazard and Operability (HAZOP) method at PT PJB Services PLTU Tembilahan. *Journal of Applied Industrial Technology and Management*, 1(2), 53–60. <https://doi.org/10.55826/tmit.v1i2.15>
- [15]. Republic of Indonesia. (1970). Law Number 1 of 1970 concerning Occupational Safety.
- [16]. Standards Australia/Standards New Zealand. (2004). AS/NZS 4360:2004 Risk management. Standards Australia, Sydney & Standards New Zealand, Wellington.
- [17]. Vesely, W.E., Goldberg, F.F., Roberts, N.H., & Haasl, D.F. (1981). Fault tree handbook (NUREG-0492). US Nuclear Regulatory Commission.
- [18]. Wiradara, AG, Nirmala G., F., & Saptaputra, SK (2024). Analysis of potential OHS hazards using the HIRARC method in the boiler section of the Nii Tanasa Kendari PLTU in Lalonggasumeeto District, Konawe Regency in 2023. Detector: *Journal of Health Science Research Innovation*, 2(1), 243–262. <https://doi.org/10.55606/detector.v2i1.3352>.