A Proposed Hierarchical Safety Goal Framework for Japan

Xiaoyu Zheng[a], Takafumi Narukawa[b], Shogo Takahara[c], Toshimitsu Homma[d], Ken Muramatsu[d], Masashi Hirano[d], Toru Ogino[e] and Toyoshi Fuketa[b,f]

[a] Japan Atomic Energy Agency (JAEA), Ibaraki, Japan, Email address: zheng.xiaoyu@jaea.go.jp
[b] The University of Tokyo, Tokyo, Japan, Email address: narukawa@n.t.u-tokyo.ac.jp
[c] JAEA, Ibaraki, Japan, Email address: takahara.shogo@jaea.go.jp
[d] Retired from JAEA, Japan
[e] Kyoto University, Kyoto, Japan
[f] Nuclear Damage Compensation and Decommissioning Facilitation Corporation (NDF), Tokyo, Japan

Abstract: Following the accident at Tokyo Electric Power Company (TEPCO) Fukushima Daiichi nuclear power station, Japan introduced stringent new regulatory requirements. Concurrently, probabilistic risk assessments (PRAs) conducted by Japanese utilities report core damage frequencies (CDFs) in the range of 10^{-6} to 10^{-5} per reactor-year. Adopting a standard generic target of $\mathrm{CDFs} < 10^{-4}$ for these plants creates a regulatory paradox: a threshold that is easily satisfied provides no incentive for continuous safety improvement and risks fostering a complacent "safety myth." To resolve this issue, this paper proposes a novel, hybrid framework for goal setting in Japan, drawing upon the rigorous optimization philosophies such as the UK's ALARP (as low as reasonably practicable) principle and the US Nuclear Regulatory Commission's (USNRC's) "two-tier" backfit rule. Specifically, we propose a two-layered regulatory architecture. First, the regulatory authority establishes top-down quantitative health objectives (QHOs) as the absolute upper limit of socially acceptable risk. Second, utilities must propose bottom-up, plant-specific surrogate objectives that are mathematically set strictly lower than their own current PRA baselines. To prevent optimistic modeling biases, these self-imposed targets are dynamically recalibrated during decennial periodic safety reviews (PSRs), requiring utilities to commit to tangible physical safety enhancements. By anchoring these dynamic metrics to real-world performance monitoring within a broader risk-informed and performance-based (RIPB) framework, this approach embeds the ALARP principle into Japanese regulation, transforming safety goals from static compliance limits into genuine drivers for continuous risk reduction.

1. INTRODUCTION

The primary lesson drawn from the Tokyo Electric Power Company (TEPCO) Fukushima Daiichi Nuclear Power Station (1F) accident is the absolute necessity of breaking away from the "safety myth", the pervasive fallacy that strict compliance with deterministic regulatory standards equates to absolute safety. A critical reflection on the accident reveals that this myth effectively caused a "halt in thinking" regarding residual risks, paralyzing the proactive pursuit of continuous safety improvement [1]. Furthermore, the historical reluctance to formally and comprehensively integrate Probabilistic Risk Assessment (PRA) into mandatory rulemaking left the infrastructure heavily reliant on deterministic assumptions that failed to account for severe, low-probability external hazards. Ultimately, the fundamental cause of the disaster's occurrence and expansion was a profoundly degraded safety culture. As highlighted in international regulatory discourse, safety goals and safety culture are inextricably linked; a strong safety culture inherently cultivates an operational philosophy consistent with the safety goal objective of continuously minimizing risk, driven by constant vigilance [2][3]. In stark contrast, Japan's degraded safety culture substituted vigilance with complacency. Disturbingly, this cultural deterioration was not a sudden phenomenon; warning signs regarding the lack of continuous safety improvement had already been pointed out following the 1999 JCO criticality accident [4]. The ultimate consequence of this complacency manifested in 2008, when opportunities to implement critical tsunami countermeasures were postponed, leaving the defense-in-depth architecture fatally vulnerable to extreme external hazards [5].

Furthermore, the TEPCO 1F accident forced a paradigm shift in how the consequences of severe nuclear accidents must be evaluated. Historically, traditional international safety frameworks, such as the US Nuclear Regulatory Commission's (USNRC's) 1986 safety goal policy statement [6] and the IAEA's fundamental safety principles (SF-1) [7], have been founded on the narrow premise that the primary risks arising from nuclear power are limited to radiological health effects, such as prompt fatalities and latent cancers. The TEPCO 1F accident fundamentally challenged this premise. While the event presented limited direct radiological health hazards, it inflicted devastating economic, social, and cultural impacts through massive environmental contamination and the long-term relocation of residents [8]. This demonstrates that accident consequences cannot be adequately captured by acute radiation exposure metrics alone. Therefore, it is both ethically and practically necessary to expand the scope of safety goals beyond traditional health-centric boundaries. To prevent unacceptable societal disruptions, Japan's safety goals must explicitly include quantitative limits for large-scale environmental releases, such as the Nuclear Regulation Authority's (NRA's) interim metric limiting Cesium-137 releases to 100 TBq [9][10], which effectively functions as a large release frequency (LRF) target. While Japan implemented some of the world's most stringent regulatory requirements in response to these failures, addressing the remaining residual risk urgently requires a clear, institutionalized policy on national safety goals to scientifically guide regulatory decision-making.

The PRAs currently conducted by Japanese utilities report that operating reactors possess core damage frequencies (CDFs) roughly between 10^{-6} and 10^{-5} per reactor-year, with risks from specific external events like earthquakes and tsunamis evaluated at even lower levels (10^{-7} to 10^{-10} per reactor-year) [11][12]. To align with global best practices, discussions are underway regarding the formal adoption of quantitative safety goals to provide risk metrics that institutionalize continuous improvement. However, if a generic international performance target, such as a CDF of 10^{-4} per reactor-year, is established, the Japanese nuclear industry already satisfies it with a massive margin. When evaluated risks sit comfortably below an established threshold, setting a generic target provides no technical or economic incentive for utilities to pursue further safety enhancements. Instead of driving optimization, such a lenient target risks fostering a new safety myth, inviting a complacent attitude where operators believe that simply "meeting the goal is sufficient". Furthermore, establishing goals that are expected to be easily met can exert undue pressure on the entities evaluating risk, potentially leading to biased, artificially tailored PRA results designed merely to pass the test.

This highlights precisely why Japan requires a redefined framework for safety goals. Because a posture of continuously seeking to improve safety is synonymous with a sound safety culture, safety goals must act as active drivers for optimization, rather than serving merely as passive benchmarks to validate the adequacy of existing regulations. To maintain the incentive for utilities to invest in safety, the regulatory architecture must structurally separate minimum regulatory limits (mandatory compliance) from aspirational safety goals (continuous improvement). Without this structural separation, the industry remains trapped in a paradigm where passing a regulatory test halts further innovation. This situation raises a critical regulatory question: under conditions where evaluated risks are already substantially lower than international norms, how should surrogate objectives for QHOs be set? Achieving true continuous improvement requires abandoning static, uniform thresholds in favor of dynamic institutional frameworks, such as European-style periodic safety reviews (PSR) or US-style backfit rules, which continuously push the boundaries of reasonably achievable risk reduction [13].

The remainder of this paper is organized as follows: Section 2 articulates the regulatory paradox of generic surrogate objectives. Section 3 investigates international frameworks for safety improvement. Section 4 analyzes modern regulatory reforms to justify the proposed framework's philosophy. Section 5 proposes a novel, hybrid two-layered safety goal framework for Japan. Finally, Section 6 concludes the paper.

2. THE REGULATORY PARADOX OF GENERIC SURROGATE OBJECTIVES

Globally, the internationally recognized standard surrogate objective for operating light water reactors is often expressed as a CDF of 10^{-4} per reactor-year [14][15]. Yet, as noted previously, the PRAs currently submitted by Japanese utilities frequently report CDFs in the range of 10^{-6} to 10^{-5} per reactor-year [12]. Adopting 10^{-4} raises a critical question: can a threshold that is already satisfied with massive margins still function as a genuine "goal" for future endeavors?

Conversely, strictly imposing a highly ambitious regulatory target of 10^{-6} presents its own severe technical and cultural dangers due to inherent evaluation biases. Risk evaluations inevitably reflect the assumptions and postures of the evaluating entity; regulators tend to enforce bounding conservatisms, while operators may lean toward optimistic modeling [16]. Imposing a strict 10^{-6} target could induce a "curse of 10^{-6}," where PRA results are artificially tailored, either consciously or unconsciously, by utilities merely to pass the expected numerical threshold. This numerical fixation completely undermines the fundamental purpose of a PRA, degrading it from a realistic, diagnostic engineering tool used for identifying hidden plant vulnerabilities into a hollow, paper-based compliance exercise. When safety is reduced to the mere act of passing a generic mathematical test, organizational safety culture stagnates. According to the Institute of Nuclear Power Operations (INPO), a genuinely healthy safety culture requires the pervasive presence of a conservative bias. Rigid limits inherently discourage this bias because organizations perceive adequate safety has been achieved the moment the numerical threshold is mathematically met [17]. This problem is further exacerbated when evaluating plant modifications. Using traditional approaches for licensing basis changes (e.g., USNRC RG 1.174), modifications are judged by the incremental change in risk ΔCDF. If a heavily modernized Japanese plant with an ultra-low baseline CDF proposes an operational change, traditional rigid thresholds struggle to accurately contextualize the ΔCDF. The safer plant is subjected to vastly disproportionate regulatory scrutiny, creating a perverse incentive where achieving ultra-low baseline risk actually restricts a utility's future operational flexibility.

This paradox is compounded by a gap in institutional capability and regulatory governance. Under Japan's current safety improvement assessment system, the regulatory authority receives utility PRA evaluations but primarily provides general comments on their validity rather than conducting rigorous independent verification [18]. Furthermore, despite widespread advocacy for risk-informed regulation (RIR) and risk-informed decision making (RIDM), actual investment in PRA human resources across Japanese industry, academia, and government remains severely limited. This creates a cyclical dilemma: establishing meaningful safety goals requires high-quality, mature PRAs and robust independent regulatory verification capabilities; yet, without a legally implementable safety goals policy, stakeholders lack the economic and regulatory incentive to invest in developing those very PRA capabilities [13].

Therefore, a properly designed national safety goal policy must resolve these contradictions. It cannot be reduced to a single static number. The central purpose of this paper is to design a comprehensive, multi-layered safety goal framework that serves dual, complementary functions: (1) providing a strict, non-negotiable regulatory judgment threshold to guarantee public safety, and (2) simultaneously acting as a dynamic endeavor goal that incentivizes stakeholders to invest in PRA maturation and pursue continuous innovation. To achieve this, the regulatory architecture must structurally and legally separate the minimum acceptable limit from the surrogate objectives. This concept is deeply rooted in international risk optimization frameworks, which will be examined in the subsequent section.

3. INTERNATIONAL FRAMEWORKS FOR SAFETY IMPROVEMENT

To transition from a static compliance framework to an optimization model, it is necessary to examine how international regulatory frameworks operationalize continuous improvement. A comparative analysis reveals that leading regulatory systems structure their safety governance around a hierarchy of safety goals. These frameworks utilize specific legal mechanisms and quantitative analyses to ensure that achieving a numerical safety target does not halt further safety investments.

3.1. The UK's ALARP Principle and Gross Disproportion

In the United Kingdom, the nuclear regulatory framework overseen by the Office for Nuclear Regulation (ONR) explicitly rejects the premise that compliance with a predefined safety target signifies the completion of safety efforts. The UK framework establishes a clear, legally binding hierarchical risk governance structure detailed in the Safety Assessment Principles (SAPs) for Nuclear Facilities, which divides risk into distinct regions bounded by basic safety levels (BSL) and basic safety objectives (BSO) [19]. The BSL represents the upper limit of tolerable risk; exceeding this numerical level places a facility in the "intolerable region," where operations must cease, or new licenses must be denied, regardless of the mitigation costs involved. Conversely, the BSO represents the broadly acceptable region where risks are considered diminutive. For example, under the SAPs, the BSL for the total predicted frequency of accidents leading to a large release might be set at 10^{-4} per reactor-year, while the BSO is set much lower at 10^{-5} per reactor-year, with specific dose bands corresponding to specific frequency limits.

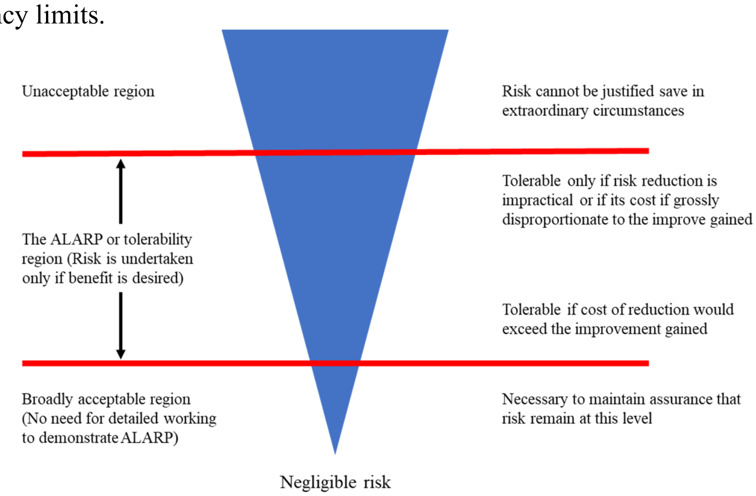

Figure 1 Levels of risk and the requirement of ALARP

As shown in Figure 1, the critical innovation of the UK framework is the "ALARP or tolerability" region situated between the BSL and BSO [20]. Crucially, the legal statutory obligation to reduce risks to a level that is as low as reasonably practicable (ALARP) applies continuously across the entire spectrum. This obligation persists as a strict legal duty even when the evaluated plant risks fall well below the BSO threshold. Within this architecture, quantitative safety objectives (the BSO) do not function as static thresholds or exit criteria for safety investments. Compliance with the safety goal is legally decoupled from the justification to cease risk reduction. The framework places a rigorous statutory obligation of proof firmly on the duty holder to demonstrate that further risk reduction is not reasonably practicable. To justify the rejection of a proposed safety enhancement, the operator must mathematically and qualitatively prove that the sacrifice in terms of cost, time, or effort is "grossly disproportionate" to the safety benefits achieved. Because the regulatory decision-making process is inherently biased toward safety, merely satisfying a quantitative safety objective cannot be used as a standalone defense to forego technically feasible improvements. This mechanism successfully operationalizes the philosophy of continuous risk optimization by transforming safety goals from passive compliance benchmarks into active, persistent drivers.

3.2. The USNRC's Two-Tier Backfit Rule

The USNRC operationalizes continuous safety improvement through a legally formalized regulatory mechanism known as the backfit rule, 10 CFR 50.109 [21]. The backfit framework establishes an objective, systematic process to incorporate new technical knowledge and operational experience into existing facility licenses. Central to this framework is a balance between promoting safety improvements and maintaining regulatory predictability, ensuring that shifts in regulatory staff positions are legally justified and suitably defined [22]. Following the landmark 1987 Union of Concerned Scientists (UCS) v. USNRC court ruling, the rule was structured into a strict "two-tier structure," legally separating the safety dimensions where cost considerations are strictly prohibited from those where cost-benefit balancing is required [23].

- Tier 1 (Mandatory Tier): Focuses on maintaining or re-defining "adequate protection" of public health and safety, as well as ensuring strict "compliance" with existing licensing bases under the Atomic Energy Act (AEA) of 1954. For measures within this tier, cost considerations are legally prohibited, and operators must comply unconditionally. Examples include mandatory post-TEPCO-accident mitigation strategies (Order EA-12-049) [24], reliable spent fuel pool instrumentation under an "administrative exemption" (Order EA-12-051) [25], and corrective actions for previously unconsidered degradation mechanisms, such as pressurizer surge line thermal stratification [26] or intergranular stress corrosion cracking (IGSCC) [27].

- Tier 2 (Discretionary Tier): Addresses safety enhancements beyond adequate protection ("extra-adequate protection"). For these discretionary measures, the USNRC carries a strict burden of proof and must execute a formal "backfit analysis" evaluating nine statutory criteria. A new requirement must be legally justified by proving that it provides a quantitative "substantial safety enhancement" and that its direct and indirect costs are commensurate with the safety benefits gained, as demonstrated by the mandate for reliable hardened containment vents (Order EA-13-109) [28].

Historically, informal and decentralized backfitting in the early 1980s induced excessive economic burdens (such as unnecessary boron dilution monitors or overly frequent diesel generator cold-start testing that caused premature mechanical wear) without yielding clear safety outcomes. In response, the USNRC established the Committee to Review Generic Requirements (CRGR) for centralized administrative control and institutionalized the regulatory analysis guidelines [29]. Under these guidelines, the quantitative health objectives (QHOs) and plant performance metrics (such as CDF and Large Early Release Frequency, LERF) serve as objective, quantitative screening metrics via the "safety goal evaluation". Before a complex and resource-intensive cost-benefit analysis is initiated, a proposed regulatory action must demonstrate that it achieves a predefined threshold of risk reduction. Consequently, the USNRC safety goals function as an active, structural gatekeeper. They provide a quantitative screening index for the regulator to determine whether a further safety requirement is socially and economically rational, while simultaneously creating a predictable environment that incentivizes utilities to mature their PRA methodologies and proactively optimize plant risk profiles.

3.3. Periodic Safety Review

While ALARP and the backfit rule establish the underlying philosophical boundaries for risk optimization, the periodic safety review (PSR) serves as the primary systematic engine for implementation across the European continent. Rooted in Principle 5 (optimization of protection) of the IAEA fundamental safety principles [7], the PSR is a proactive, multi-factor approach to the regular, comprehensive re-evaluation of overall plant safety over time, governed internationally by standards such as IAEA SSG-25 [30]. Following the 1F accident, the regulatory mandate for the PSR was heavily reinforced. The European Union amended its Nuclear Safety Directive (2009/71/Euratom) [31], legally binding member states to require operators to perform comprehensive safety reassessments at least every ten years. A distinctive feature of this directive is its explicit, uncompromising safety objective: preventing accidents and, should they occur, mitigating consequences to avoid early or large radioactive releases.

In France, this rigorous approach was institutionalized by the Transparency and Nuclear Security (TSN) Law of 2006, now codified under Article L. 593-18 of the Environment Code [32]. This law mandates a decennial PSR structure overseen by the independent regulatory body, ASN (now ASNR). The French PSR framework explicitly splits the assessment process into two distinct phases: a conformity verification (baseline compliance review ensuring current regulations are met) and a safety reassessment. Crucially, the safety reassessment phase benchmarks existing aging facilities not against their legacy design bases from decades prior, but against modern safety requirements applicable to newer, state-of-the-art installations. For instance, during the recent decennial PSR cycles for France's 900 MWe PWR fleet, safety upgrades were targeted to approach the mitigation capabilities of the modern evolutionary power reactor (EPR) design [33]. This resulted in major, capital-intensive hardware modifications, such as the implementation of in-containment molten core stabilization and advanced flooding measures. Following the review, the ASN holds the clear statutory power under Article L. 593-19 to set binding prescriptions and legally impose additional technical requirements on the licensee as an absolute condition for continued operation [34]. This governance model effectively turns the "state-of-the-art" into a continuously moving target, legally compelling operators to modernize older plants systematically.

3.4. Japan's Current Backfit Policy and its Structural Limitations

In the wake of the TEPCO 1F accident, Japan amended the Act on the Regulation of Nuclear Source Material, Nuclear Fuel Material and Reactors (in short, Reactor Regulation Act) to introduce a formal backfit system under Article 43-3-23 [35]. This framework grants the NRA the authority to order the suspension of use, modification, or relocation of a reactor facility if it no longer conforms to newly established technical standards [36]. The Japanese backfit system has demonstrated significant agility in enforcing compliance when new extreme external hazards are identified. For example, when new geological data regarding the eruption scale of the inactive Mt. Daisen volcano emerged, indicating that a massive fall of volcanic ash could exceed original design assumptions, the NRA issued backfit orders requiring utilities like Kansai Electric Power to revise their design policies and apply for license amendments to protect against ash accumulation that could compromise emergency power systems. However, there is a fundamental structural difference between the Japanese and US frameworks. The Japanese system does not legally define a "two-tier" structure, nor does it incorporate a formal cost-benefit analysis mechanism for safety enhancements that exceed minimum regulatory requirements. It functions primarily as a mechanism for enforcing mandatory compliance with updated baselines rather than providing an economic or probabilistic framework for discretionary optimization.

Furthermore, Japan's current regulatory system also has a critical structural gap compared with the European architecture. While Reactor Regulation Act Articles 43-3-29 require operators to conduct "Safety Improvement Assessments" reflecting the latest knowledge, this functions purely as a notification system. The regulator receives operators' assessments and comments on their validity, rather than exercising a rigorous review-and-approval mandate. This severely limits the regulator's ability to issue binding mandates for continuous improvement outside of direct, broad legislative rule changes. In contrast, the European PSR model grants the competent regulatory authority explicit oversight to systematically review, formally approve, and issue legally enforceable licensing decisions regarding the continued operation of the facility. This legal differentiation underscores that true continuous improvement requires a robust, legally binding regulatory structure capable of transforming risk information into mandatory plant enhancements [36].

3.5. Summary

The comparative analysis of regulatory frameworks of Europe, the United States, and the current status of Japan, summarized in Table 1, reveals a critical divergence. Unlike the US two-tier structure that explicitly incorporates cost-benefit justifications, or the UK framework that utilizes "gross disproportion" as a judgment criterion, Japan's current system tends to treat all new regulatory requirements as uniform, mandatory compliance obligations. Consequently, this structure fails to provide institutional incentives for operators to pursue continuous safety improvements after obtaining initial licensing. Furthermore, mechanisms to drive risk reduction through daily regulatory activities remain immature in Japan. To strike a balance between over-regulation (which deprives operators of their sense of ownership) and dangerous under-regulation, establishing a formal safety goal framework is indispensable for driving proactive safety enhancements.

Table 1: Comparative overview of international regulatory optimization mechanisms

Country /Region	Optimization Mechanisms	Core Philosophy	Legal Authority / Standard
United Kingdom	ALARP / Gross Disproportion	Continuous statutory obligation to keep reducing risk even when the risk level is below the Basic Safety Objective (BSO), unless the cost of risk reduction is grossly disproportionate to the benefit.	Health and Safety at Work etc. Act 1974, Safety Assessment Principles (SAPs)
United States	Two-Tier Backfit Rule	Additional requirements beyond "adequate protection" require proving a substantial safety enhancement and a cost-benefit verification demonstrating that the averted losses outweigh the implementation costs.	The AEA of 1954, Backfit Rule (10 CFR 50.109), Maintenance Rule, Reactor Oversight Process (ROP), Regulatory Analysis Guidelines (NUREG/BR-0058), etc.
France /EU	Periodic Safety Review (PSR)	Re-evaluating existing plants every 10 years by benchmarking them against the modern safety requirements applied to the latest advanced reactors.	Article L. 593-18 (Environment Code), 2014/87/Euratom
Japan	Mandatory Backfit, Safety Improvement Assessment	Strict and deterministic compliance with updated technical standards.	Articles 43-3-23 & 43-3-29 (Reactor Regulation Act)

4. MODERN REGULATORY INSIGHTS

The limitations of applying rigid regulatory risk targets were vigorously debated by the USNRC during the development of 10 CFR Part 53 [37], a risk-informed, technology-inclusive, and performance-based (TI-RIPB) regulatory framework for advanced nuclear reactors. During the drafting process, a central point of contention was whether QHOs should be codified as explicit, legally binding requirements. The industry, represented by organizations such as the Nuclear Innovation Alliance (NIA) and the Nuclear Energy Institute (NEI), strongly cautioned that promulgating QHOs as rigid legal rules would severely limit performance-based flexibility and restrict the use of alternative analysis methodologies for diverse, novel reactor technologies [38]-[41]. They also argued that incorporating QHOs into the rule text as a performance metric introduces unnecessary and unforeseen challenges, primarily because existing dose limits already create sufficient performance standards to protect public health and safety [42].

In a landmark decision formalized in the staff requirements memorandum (SRM) of SECY-23-0021, the USNRC officially disapproved the codification of the QHOs as rigid legal rules within the proposed Part 53 [43]. The Commission astutely recognized that promulgating the QHOs as the exclusive, rigid metric for establishing a cumulative risk standard could introduce unnecessarily inflexible requirements that might conflict with the foundational intent of the AEA of 1954 and inadvertently stifle innovative safety cases for passively safe advanced designs [44]. Furthermore, the Commission explicitly directed the staff not to apply consensus PRA standards as a strict, unyielding checklist for determining PRA acceptability.

Instead, to maintain regulatory flexibility, the Commission proposed requiring applicants to define their own cumulative plant risk metrics, or a set of surrogate metrics, that accurately reflect the unique risk profile of their design and ensure an acceptable level of radiological risk to the public [45]. This regulatory pivot demonstrates a significant shift in modern regulatory philosophy: the regulator should define the high-level philosophical boundaries of public safety, but the licensee must take complete ownership of defining, proposing, and mathematically defending the specific risk metrics that demonstrate exactly how their unique technology achieves that required level of safety. While the initial motivation for this approach in Japan differs from the US context, the USNRC's pivot toward licensee-defined metrics provides a strong practical justification for the proposed hierarchical framework.

5. PROPOSED HIERARCHICAL FRAMEWORK FOR JAPAN

To ensure risk metrics serve as dynamic drivers rather than static limits, this paper proposes a hybrid, two-layered regulatory framework. Drawing heavily on the hierarchical structure of safety goals defined in IAEA TECDOC-1874 [46], as shown in Figure 2, this framework structurally separates regulatory safety goals from voluntary plant-specific surrogate objectives for continuous improvement.

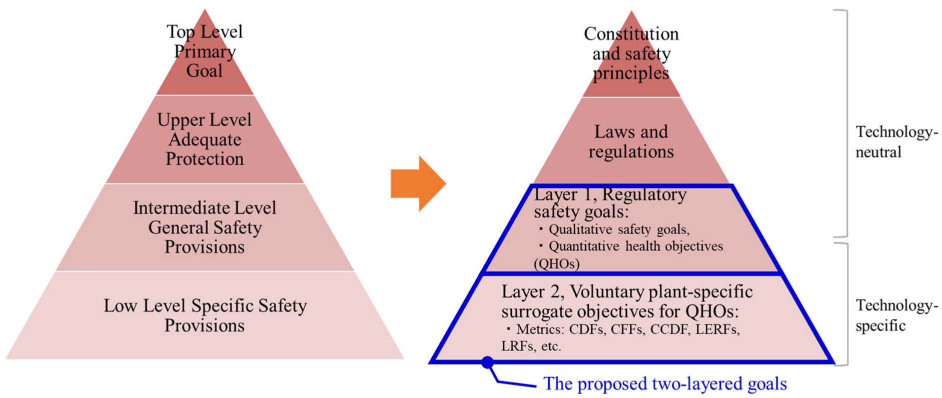

Figure 2 Hierarchical structure of safety goals and the proposed two-layered framework

5.1. Layer 1: Top-Down Regulatory QHOs

In the proposed Layer 1, the regulatory authority establishes top-down qualitative safety goals and quantitative health objectives (QHOs). These QHOs correspond to the intermediate level of the IAEA hierarchy and define the absolute upper limit of socially acceptable risk. By establishing these objectives at the highest regulatory level, Layer 1 serves as a technology-neutral public commitment to ensure the consistency and coherence of national safety regulation, protecting public health without rigidly prescribing specific engineering solutions. In Japan, extensive preliminary discussions regarding the formal establishment of national safety goals have taken place, heavily informed by the desire to prevent any recurrence of the societal disruption caused by the 1F accident. The proposed QHOs focus on suppressing individual annual mortality risks to adequately low levels when compared to the natural background risks of daily life (e.g., interim quantitative safety goals of Japan: suppressing individual acute mortality and latent cancer risks to 10^{-6} per year). Furthermore, the accident indicates that mortality is not the only unacceptable outcome of a severe nuclear accident and massive land contamination poses an existential threat to the socio-economic fabric of the nation (e.g., interim quantitative safety goals of Japan: prompting proposed limits, such as ensuring that the frequency of a Cs-137 release exceeding 100 TBq does not exceed 10^{-6} per year). This setting can establish a firm, non-negotiable regulatory ceiling that guarantees public and environmental protection against catastrophic land interdiction.

5.2. Layer 2: Bottom-Up Plant-Specific Surrogate Objectives

To provide a tangible, mathematical incentive for utilities to invest in safety when their baseline risks are already exceptionally low, the framework shifts the responsibility of goal-setting to the licensees for the second layer. Empowering utilities to take ownership of their safety profiles aligns directly with the modern regulatory philosophy established in the USNRC's 10 CFR Part 53. By adopting this philosophy, the proposed framework empowers Japanese utilities to take ownership of their safety profiles and propose their own risk metrics, rather than passively receiving generic limits from the regulator.

5.2.1. Surrogate Objectives for QHOs (Subsidiary Objectives)

Under the IAEA and USNRC safety goal hierarchy, high-level safety objectives must be translated into low-level specific safety provisions to be operationally useful. Therefore, each utility is required to conduct a plant-specific PRA and translate the regulatory QHOs into technology-specific surrogate objectives, acting as subsidiary objectives or surrogate risk metrics. Common examples include CDF, CFF, conditional containment failure probability (CCFP), LERF, LRF, which serve as the quantitative benchmarks for plant performance. Recognizing the varying maturity of risk assessment tools across the industry, the extent to which external event PRAs are incorporated into these initial metrics may depend on the current development stage of the respective PRA methodologies.

Crucially, instead of the regulator imposing a uniform $CDF < 10^{-4}$ per year target, each utility must propose its own bottom-up, plant-specific surrogate objectives based on the results of their own PRA, and must update them periodically to reflect the state-of-the-art PRA. These proposed goals must strictly meet two non-negotiable conditions:

(1) They must unconditionally satisfy the national QHOs established in Layer 1.
(2) They must be set or updated at a lower risk level than the plant's current PRA baseline results.

For instance, if a utility's internal PRA evaluates its current baseline CDF at 5×10^{-5} per reactor-year, the utility must voluntarily propose a surrogate objective that is strictly lower than that baseline (e.g., 1×10^{-5} or lower). The practical benefit of this approach is that it forces actionable risk reduction. By establishing voluntary risk criteria, the utility creates a self-imposed, legally binding mandate to continually refine its PRA models and optimize plant configurations to meet an increasingly strict target, thereby permanently avoiding the complacency of merely passing a generic test.

5.2.2. Practicability of Voluntary Risk Criteria Determined by Utilities

This approach is supported by international investigations such as reports by the OECD Nuclear Energy Agency (NEA) as shown in Figure 3, which demonstrate that surrogate objectives and risk criteria are diverse across different nations. International precedents confirm that design- and technology-specific risk criteria are a globally accepted norm for robust risk-informed decision making [47].

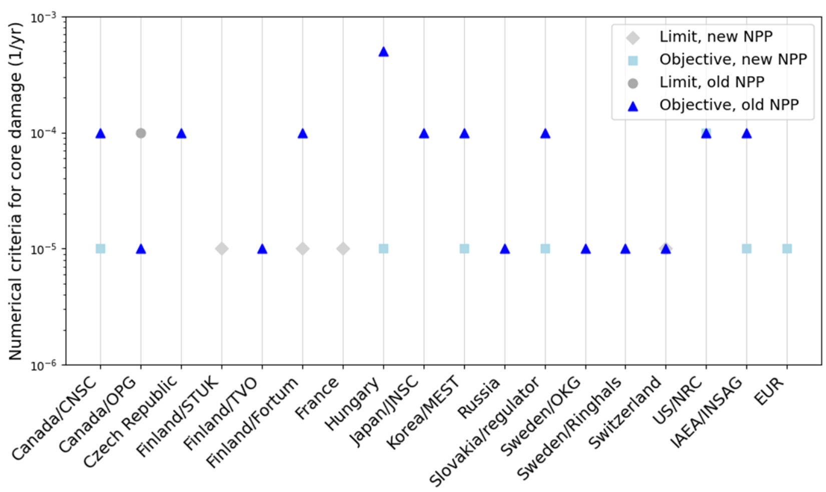

Figure 3 Numerical criteria defined for core damage

5.3. Execution via PSR

To transform this theoretical two-layered framework into tangible safety improvements, the structure must be operationalized through a legally robust, cyclic enforcement mechanism. Drawing upon the well-established European decennial review model, we recommend that this framework utilizes the periodic safety review as its primary implementation engine. At each 10-year PSR reassessment, utilities must submit a comprehensive safety reassessment. Crucially, a plant's baseline PRA should not be static, but it will naturally fluctuate over time due to updated simulation tools, revised external hazard curves, and evolving methodologies. Therefore, at the start of each PSR cycle, the baseline PRA must be formally recalibrated. Utilities are then required to propose specific safety enhancements and dynamically revise their surrogate objectives relative to this newly updated baseline, ensuring the target remains a moving driver for continuous improvement rather than a fixed historical artifact.

While empowering utilities to set self-imposed targets forces actionable risk reduction, it introduces a critical vulnerability that may incentivize optimistic modeling assumptions. If aspirational goals are treated merely as mathematical hurdles, the process risks degrading into a theoretical numbers game where PRA parameters are artificially tailored without delivering actual safety benefits. To guard against this problem, utilities must propose and commit to tangible physical safety enhancements to satisfy their self-imposed, plant-specific surrogate objectives defined in Layer 2, such as capital-intensive hardware design modifications, updated mitigation strategies, or operational procedure improvements. Because these targets are mathematically set on the lower-risk side of the previous baseline, the utility is structurally and legally compelled to identify vulnerabilities and implement proactive backfits over the entire operational lifetime of the facility. The regulatory body then rigorously reviews these proposals, the underlying PRA models, and the verification data, ensuring that the utility is genuinely adhering to the ALARP principle and continually pushing the boundaries of reasonably achievable risk reduction.

Finally, maintaining the vital connection between numerical risk targets and real-world safety requires acknowledging that PRA is only one cornerstone of a broader RIPB regulatory framework. The structural backbone of the system must remain firmly rooted in continuous performance monitoring, deterministic defense-in-depth, and the maintenance of adequate safety margins. Because a PRA is fundamentally a mathematical approximation, theoretical risk reductions must be continuously validated by real-world operational data. By systematically feeding realistic empirical data back into the PRA during each PSR cycle, the regulator ensures that achieving lower numerical targets translates into tangible, physical safety enhancements rather than mere statistical manipulation.

6. CONCLUSIONS

One of the most profound lessons from the TEPCO 1F accident is that regulatory compliance must never be mistaken for absolute safety. The fact that the utilities are frequently reporting CDFs in the range of 10^{-6} to 10^{-5} per reactor-year has inadvertently created a regulatory paradox. If standard international targets, such as a CDF of 10^{-4} per reactor-year, are adopted as static limits, they risk fostering a new safety myth by removing the technical and economic incentives required for utilities to pursue further safety improvements.

To resolve this contradiction and ensure that safety optimization is not merely a slogan, this paper proposes a paradigm shift from static compliance to dynamic risk optimization. Drawing upon the rigorous philosophies of the UK's ALARP principle, the USNRC's two-tier backfit rule, and the EU's periodic safety review process, we propose a hybrid, two-layered regulatory framework for Japan. In the first layer, the regulatory authority establishes top-down quantitative health objectives to define the immutable upper boundaries of socially acceptable risk, ensuring public and environmental protection. In the second layer, utilities are required to propose bottom-up, plant-specific surrogate objectives. Crucially, these surrogate metrics (such as CDF, CFF, CCDP, LERF or LRF) must be mathematically set on the strictly safer side of the plant's own current PRA baselines.

By operationalizing these self-imposed, ambitious targets through the 10-year PSR cycle, the proposed framework mandates actionable risk reduction and embeds continuous improvement directly into the regulatory architecture. This approach fosters a risk-informed performance-based regulatory environment that emphasizes licensee accountability and continuous optimization. Utilities are incentivized to invest in high-quality PRA methodologies, rigorously evaluate latent uncertainties, and proactively identify vulnerabilities, rather than treating risk assessments merely as compliance exercises bounded by generic thresholds. Ultimately, this approach ensures that quantitative safety goals function not as a finish line to halt safety investments, but as an institutional mechanism for cultivating a robust safety culture and achieving continuous nuclear risk reduction.

Acknowledgements

This paper builds upon the discussions and findings presented in two expert committee meetings of the Safety Goal Study Committee, jointly organized by the JAEA and the University of Tokyo. The authors would like to express sincere gratitude to the committee members for their valuable insights, constructive discussions, and materials that informed the analysis developed in this work. The committee's activities and background information are available through the University of Tokyo website, https://www.safety.n.t.u-tokyo.ac.jp/safetygoal/index.html (accessed June 10, 2026). Any interpretations, analyses, or conclusions presented in this paper are solely those of the authors and do not necessarily represent the views of the committee, JAEA, the University of Tokyo, Kyoto University or the NDF.

References

[1] Nuclear Regulation Authority (NRA). Review of discussions by the study team on continuous safety improvement (In Japanese). Tokyo, Japan (2021)
[2] R.A Meserve. The evolution of safety goals and their connection to safety culture. AESJ/ANS Topical Meeting on Safety Goals and Safety Culture. Milwaukee, WI, USA. June 17-21 (2001)
[3] Working Group of the Committee on Safety Goals. History of safety culture in the United States (In Japanese). Tenth Meeting of the Committee on Safety Goals. Tokyo, Japan, November 14 (2025)
[4] T. Fuketa. Discussion on the introduction of the safety goals report (In Japanese). Thirteenth Meeting of the Committee on Safety Goals, Tokyo, Japan. May 29 (2026)
[5] T. Okuyama. Lessons learned from the TEPCO Fukushima Daiichi nuclear power station accident extracted from the perspective of safety goals (In Japanese). Eleventh Meeting of the Committee on Safety Goals, Tokyo, Japan. January 7 (2026)
[6] USNRC. Safety goals for the operations of nuclear power plants, policy statement. Federal Register 51: 30028 (1986)
[7] IAEA. Fundamental safety principles. No. SF-1, IAEA, Vienna, Austria (2006)
[8] IAEA. The Fukushima Daiichi accident: report by the director general. GC(59)/14, IAEA, Vienna, Austria (2015)
[9] Nuclear Safety Commission. Interim summary of deliberations on safety goals: report by the Nuclear Safety Commission's expert committee on safety goals, in Japanese. Tokyo, Japan (2003)
[10] Nuclear Safety Commission. Performance objectives for light water reactor facilities for power generation: corresponding to proposed safety goals. Tokyo, Japan (2006)
[11] T. Narukawa, et al. Why are Japanese nuclear power plants "so safe"? - Insights from safety improvement assessments. In: Proceedings of PSAM 18, Pittsburgh, PA, USA, July 19-24 (2026)
[12] T. Narukawa et al. Structural impediments to safety goals and risk-informed regulation in Japan. manuscript to be submitted.
[13] T. Fuketa. Discussions surrounding safety goals (In Japanese). 2024 Joint Symposium on Nuclear Energy. Tokyo, Japan. January 20 (2025)
[14] International Nuclear Safety Advisory Group. Basic safety principles for nuclear power plants 75-INSAG-3 Rev.1. INSAG-12, International Atomic Energy Agency. Vienna, Austria (1999)
[15] USNRC. An approach for using probabilistic risk assessment in risk-informed decisions on plant-specific changes to the licensing basis. RG 1.174, Revision 3. Washington DC, USA (2018)
[16] N. J. Diaz. Realism and conservatism. USNRC Speech No. S-03-023, 2003 Nuclear Safety Research Conference, Washington DC, United States. October 20 (2003)
[17] Institute of Nuclear Power Operations. Traits of a healthy nuclear safety culture, INPO 12-012, Rev. 1. Atlanta, GA, United States (2013)

[18] M. Hirano, History of periodic safety review (PSR) in the EU (In Japanese). Thirteenth Meeting of the Committee on Safety Goals, Tokyo, Japan, May 29 (2026)
[19] Office for Nuclear Regulation. Safety assessment principles for nuclear facilities, 2014 Edition, Revision 1. Bootle, UK (2020)
[20] Health and Safety Executive. The tolerability of risk from nuclear power stations. Her Majesty's Stationery Office, London, UK (1992)
[21] USNRC. Backfitting of nuclear power plant licensees. 10 CFR 50.109. Federal Register, 53(132): 20603–20608 (1988)
[22] USNRC. Backfitting guidelines. NUREG-1409, Rev. 1. U.S. Nuclear Reactor Regulation, Washington DC, United States. (2020)
[23] UCS v. NRC. 824 F.2d 108. United States Court of Appeals, District of Columbia Circuit (1987)
[24] USNRC. Order modifying licenses with regard to requirements for mitigation strategies for beyond-design-basis external events, Order EA-12-049 (2012)
[25] USNRC. Order modifying licenses with regard to reliable spent fuel pool instrumentation, Order EA-12-051 (2012)
[26] USNRC. Pressurizer surge line thermal stratification, NRC Bulletin 88-11 (1988)
[27] USNRC. NRC position on IGSCC in BWR austenitic stainless steel piping, Generic Letter 88-01 (1988)
[28] USNRC. Order Modifying licenses with regard to reliable hardened containment vents capable of operation under severe accident conditions, Order EA-13-109 (2013)
[29] USNRC. Regulatory analysis guidelines of the USNRC. NUREG/BR-0058, Revision 4. USNRC, Washington DC, USA (2004)
[30] IAEA. A framework for an integrated risk informed decision making process. INSAG-25, IAEA, Vienna, Austria (2011)
[31] Council of the European Union. Council Directive 2009/71/Euratom of 25 June 2009. Official Journal of the European Union L 172 18-22 (2009)
[32] Code de l'environnement. Articles L. 593-18 and L. 593-19, Republic of France (2006)
[33] R. Cozeret, et al. Improvement on 900 MWe NPPs in the occasion of the 4th 10-year Periodic Safety Review on severe accidents. EUROSAFE 2019, Cologne, Germany, November (2019)
[34] Autorité de sûreté nucléaire. Resolution 2021-DC-0706 of 23 February 2021 setting prescriptions for ... fourth periodic safety review, ASN, France (2021)
[35] Act on the Regulation of Nuclear Source Material, Nuclear Fuel Material and Reactors (In Japanese), Act No. 166 of 1957, as amended, Japan (2019)
[36] T. Ogino. Function of the backfit system in current Japanese law (In Japanese). Thirteenth Meeting of the Committee on Safety Goals, Tokyo, Japan, May 29 (2026)
[37] USNRC. Risk-informed, technology-inclusive regulatory framework for advanced reactors, Federal Register 91(6): 15696-15881 (2026)
[38] C. Draffin, U.S. Nuclear Industry Council perspectives on risk metric, Public Workshop on Technology Inclusive Risk Metrics for Advanced Reactors (2024)
[39] Nuclear Innovation Alliance, NIA public comment on the proposed 10 CFR Part 53 rule (2025)
[40] A. Stein, Quantitative health objectives in a performance-based regulation, The Breakthrough Institute (2022)
[41] The Breakthrough Institute, Nuclear Innovation Alliance, N.P. Kadambi, Stakeholder consensus on proposed Part 53 major topics (2025)
[42] Letter from Doug True (NEI) to Dan Dorman (EDO/NRC). Subject: Comprehensive industry comments on the NRC’s rulemaking on 10 CFR Part 53, NRC-2019-0062, Aug 31 (2022)
[43] USNRC. Staff requirements – SECY-23-0021 (2024)
[44] A. Caputo. Commissioner Caputo’s comments on SECY-23-0021, USNRC (2023)
[45] C.T. Hanson. Chair Hanson’s comments on SECY-23-0021, USNRC (2023)
[46] IAEA. Hierarchical structure of safety goals for nuclear installations. IAEA TECDOC-1874, IAEA, Vienna, Austria (2019)
[47] Nuclear Energy Agency. Probabilistic risk criteria and safety goals. NEA/CSNI/R(2009)16, Organisation for Economic Co-operation and Development, Paris, France (2009)