Review

Not peer-reviewed version

Safety Intervention in Technical Operations at Oil and Gas Project Sites: A Structured Review of Barrier, Behavioural, and Management-System Controls, and an Integrated Intervention Framework

Rafiullah Khan *

Posted Date: 15 September 2026

Keywords: process safety; barrier management; permit to work; behaviour-based safety; contractor safety management; oil and gas; turnaround maintenance; leading indicators; ISO 45001

Review

Safety Intervention in Technical Operations at Oil and Gas Project Sites: A Structured Review of Barrier, Behavioural, and Management-System Controls, and an Integrated Intervention Framework

Rafiullah Khan

KBR, United Arab Emirates; rafiullah.khan@kbr.com

Abstract

Background: Technical operations at oil and gas project sites — drilling and well intervention, construction and commissioning, turnaround maintenance, hot work, confined-space entry, hydrotesting, and heavy lifting — concentrate high-energy hazards, transient multi-employer workforces, and abnormal plant states within compressed schedules. Safety intervention here is pursued through three largely separate traditions: barrier management, behaviour-based observation, and management-system certification. **Objective:** To synthesise the evidence and standards base for each, evaluate what is known about intervention effectiveness in technical operations specifically, and propose an integrated framework assigning each intervention class to the failure mode it can plausibly control. **Methods:** Structured review following PRISMA-ScR, searching Scopus, Web of Science, ScienceDirect, and the ASCE and OnePetro libraries for records published 1990–2026, with a purposive corpus of public inquiry reports and normative standards (IOGP, API, ISO, IEC, NORSOK, OSHA, HSE). Forty-nine sources were retained (Figure 1). **Results:** Most intervention evaluation evidence addresses personal safety (recordable injury rates) while the catastrophic risk is process safety — an asymmetry unresolved since Piper Alpha. Contractor fragmentation, not individual error, is the dominant proximate condition in serious project-site events, yet the contracting interface has no controlled evaluation in the retained corpus. Behaviour-based safety has reasonable evidence for repetitive, observable, low-energy tasks and weak theoretical warrant for the judgement-dense work of turnarounds and well operations. Permit-to-work and energy-isolation systems are the most consistently implicated failed barriers, and their failures are failures of administration under schedule pressure rather than of design. The highest-reliability interventions available displace administrative reliance with physical constraint — double-block-and-bleed with monitored bleed, spool removal, individual employee locks, verified inerting, SIL-rated instrumented protection, redundant tested well barrier envelopes, hydro-rather than pneumatic testing, and design-stage elimination of entry and hot work — and all are decided in engineering design and procurement rather than at the site. Three groups of technical operations show traceable intervention success, and each pairs an engineered barrier with a competence requirement and an independently testable verification criterion. **Conclusions:** Intervention effort is misallocated relative to the risk profile: it over-invests in observing individual behaviour and under-invests in barrier verification, interface governance, and competence assurance across the contracting chain. We propose a four-layer framework — barrier integrity, competence, disclosure, governance — with a matched leading-indicator set, five separately manipulable contractor-interface components, and a stepped-wedge design by which they could be tested across multi-site portfolios.

Keywords: process safety; barrier management; permit to work; behaviour-based safety; contractor safety management; oil and gas; turnaround maintenance; leading indicators; ISO 45001

1. Introduction

Oil and gas project sites are among the most intensively regulated workplaces in the world, and among the most persistently lethal. The hazard profile is unusual in that two quite different kinds of harm coexist on the same footprint. Occupational or *personal*safety events — falls from height, dropped objects, struck-by incidents, hand injuries, vehicle collisions — are frequent, individually survivable, and comprehensively counted. Process safety events — loss of containment, blowout, fire, explosion, structural collapse — are rare, catastrophic, and counted only through indicator schemes that most operators adopted long after their injury-reporting systems were mature.[14,22,25] Technical operations sit at the intersection: a shutdown crew breaking containment on a hydrocarbon line is exposed to both hazard classes simultaneously, and the interventions that reduce one do not necessarily reduce the other.

This distinction is not academic. It is the central lesson that four decades of public inquiry have tried and largely failed to institutionalise. The Cullen inquiry into the 1988 Piper Alpha disaster identified the failure of the permit-to-work system during concurrent maintenance activity as the immediate mechanism by which 167 people died, and located the cause of that failure in the platform's management of work rather than in the competence or motivation of any individual.[1] Nineteen years later, the Baker Panel examining BP's U.S. refineries found that the company had tracked and rewarded declining personal-injury rates while process-safety barriers degraded, and stated plainly that a good injury record does not indicate that major-accident risk is under control.[2] The U.S. Chemical Safety Board reached a materially identical conclusion on the same events.[3] After the 2010 Macondo blowout, the Presidential Commission and the CSB again described a decision environment in which schedule and cost pressure was resolved against barrier integrity at multiple sequential decision points, with no single actor behaving irrationally within their local frame.[4,5,13]

Yet the intervention literature has not reorganised itself around this lesson. The bulk of published evaluation evidence concerns interventions whose measured outcome is a recordable-injury metric — behaviour observation programmes, safety-climate initiatives, training packages, supervisor communication trials — and whose target population is a stable, directly employed workforce doing repetitive work.[35–37,41–43] Technical operations at project sites are the opposite of that population: the workforce is transient and multi-employer, the work is non-routine by definition, the plant is in an abnormal state, and the tasks with the greatest catastrophic potential are performed a handful of times per project by crews who may never have worked together before.

This review therefore asks four questions. First, what interventions are actually deployed to control safety risk in technical operations at oil and gas project sites, and how should they be classified? Second, what is the strength and the *relevance*of the evidence for each class — that is, was the intervention tested on work resembling the work it is now applied to? Third, which failure modes recur across the inquiry record, and is the intervention portfolio aligned to them? Fourth, what would a coherent, testable intervention framework for this setting look like, and what indicators would show whether it was working?

Two features distinguish this review. It deliberately places the peer-reviewed intervention-effectiveness literature and the normative standards corpus in a single analytic frame, because in this domain practice is governed far more by standards and recommended practices than by published trials — a reviewer who reads only the journal literature will misdescribe what actually happens on a project site. And it appraises each intervention class against the failure modes documented in inquiry reports rather than against its own stated objectives, which reframes several well-established interventions as addressing hazards that are not the ones killing people.

 doi:10.20944/preprints202609.1107.v1

2. Methods

2.1. Design and Reporting

Because the objective was to map and appraise a heterogeneous intervention landscape rather than to estimate a pooled effect, a structured scoping-review design was adopted and reported in accordance with the PRISMA extension for scoping reviews (PRISMA-ScR).[49] **Protocol and registration.** This review was not prospectively registered. PROSPERO does not accept scoping reviews, and by the time this limitation was identified the review was past the data-extraction stage, so prospective registration was no longer possible. The protocol, search strategy, eligibility criteria, and charting framework were instead deposited retrospectively, and labelled as such, in an OSF Registration (DOI: 10.17605/OSF.IO/Z8MN9), which also documents the screening, extraction, and synthesis procedures in full. Because the registration postdates screening, it should be read as a timestamped record of procedures already executed rather than as a prospective commitment, and the reader cannot independently verify that the eligibility criteria and charting fields were fixed in advance of screening rather than adapted to the retrieved literature; this limitation is restated in Section 7.2. A completed PRISMA-ScR checklist is provided as Appendix B.

2.2. Search Strategy

We searched Scopus, Web of Science, ScienceDirect, the ASCE Library, and OnePetro (SPE/IADC/OTC conference and journal content) for records published between January 1990 and June 2026; all searches were last executed on 18 June 2026. Full database-specific strings, field tags, filters, and per-database yields are given in Appendix A. The start date was set at the publication of the Cullen report, which marks the beginning of the modern safety-case regime in offshore oil and gas.[1] Search strings combined setting terms ("oil and gas," "petroleum," "offshore," "refinery," "gas plant," "drilling," "turnaround," "shutdown," "commissioning," "pipeline construction") with intervention terms ("safety intervention," "barrier management," "permit to work," "lockout tagout," "energy isolation," "behaviour-based safety," "safety observation," "stop work authority," "contractor safety management," "safety management system," "leading indicator," "process safety indicator," "safety climate," "competence assurance") and, where the database supported it, with outcome terms ("incident rate," "loss ofcontainment," "process safety event," "near miss," "fatality").

A parallel purposive search retrieved the normative and inquiry corpus, which is not reliably indexed in bibliographic databases: reports of the International Association of Oil & Gas Producers, American Petroleum Institute recommended practices, ISO management-system standards, U.S. Occupational Safety and Health Administration standards, UK Health and Safety Executive guidance, the Norwegian regulator's annual risk-level programme, and the published inquiry or investigation reports for Piper Alpha, Texas City, and Macondo.[1–5,21–31] Reference lists of retrieved reviews were hand-searched (backward snowballing).

2.3. Eligibility Criteria

Peer-reviewed records were included where they (a) reported empirical evaluation, systematic synthesis, or explicit theoretical development of a safety intervention or control mechanism; (b) addressed either the oil and gas sector directly or a transferable high-hazard or construction setting where the transfer was argued rather than assumed; and (c) were published in English. Standards and inquiry documents were included where they establish a normative requirement, a recommended practice, or an authoritative causal account bearing on technical operations. We excluded records on four grounds, stated here as they were recorded during screening: no intervention or control component; outcome out of scope, meaning the only outcomes reported were environmental, asset-integrity, equipment-reliability, or chronic occupational-health outcomes (noise, hydrocarbon exposure, ergonomics) with no acute human-safety dimension; a computational or reliability model with no implementation component; and insufficient method detail to extract the

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

charting fields, which excluded conference abstracts without extractable method. A fifth ground, non-transferable setting, excluded records whose setting was neither oil and gas nor a high-hazard or construction setting for which transfer was argued in the record itself rather than assumed.

2.4. Screening and Selection Process

Number of reviewers. Screening, full-text assessment, and charting were performed by the single author. This departs from the PRISMA-ScR expectation of independent duplicate screening and is a material limitation, since single-reviewer screening is known to miss eligible records and to admit reviewer preference into borderline decisions. Three procedures were used to reduce, though not eliminate, that risk. First, eligibility criteria were written and fixed before screening, and every full-text exclusion was recorded against one of the five pre-specified reasons reported in Figure 1 rather than as a free-text judgement. Second, all 135 full-text records were screened twice, in two passes separated by a minimum of fourteen days, with the second pass conducted against the criteria alone and without reference to the first pass's decisions; the eleven records on which the two passes disagreed were resolved by re-reading against the criteria and are flagged in the screening log. Third, the retained corpus was checked for completeness by backward snowballing from all seven retained reviews and by forward citation checking of the two randomised trials, on the reasoning that a single-reviewer search is most likely to have missed evaluations rather than standards. The screening log, including all title/abstract and full-text decisions with reasons and the disagreement flags, is provided as supplementary material. No claim of independent duplicate screening is made anywhere in this review.

2.5. Study Flow

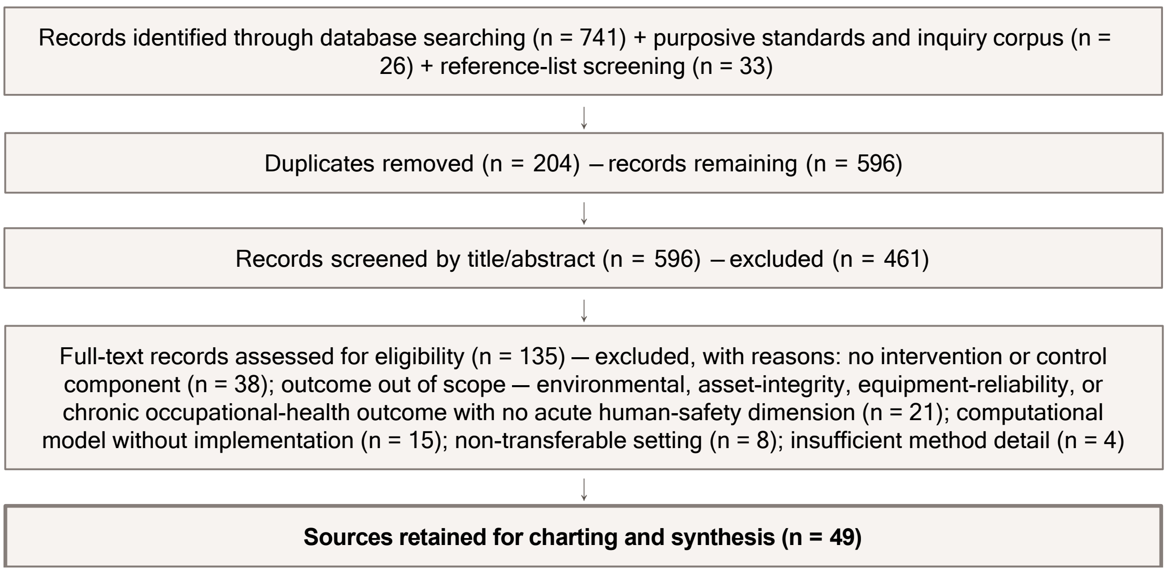

Figure 1. Source selection flow, reported per PRISMA-ScR.49.

2.6. Charting Framework and Appraisal

Retained sources were charted against five fields: intervention class; the failure mode it is intended to control; the setting in which evidence was generated; the outcome measure used; and the design strength of that evidence. Peer-reviewed evaluations were appraised for design strength on a four-level ordinal scale (randomised or cluster-randomised field trial; controlled or interrupted time-series quasi-experiment; uncontrolled pre-post or cross-sectional association; theoretical or descriptive). Meta-analyses were charted at the level of their constituent design mix. Standards and inquiry documents were not appraised for design strength — they are not evidence of effectiveness — but were charted for the failure modes they assert and the controls they require, and are treated

 doi:10.20944/preprints202609.1107.v1

throughout as statements of expert consensus and of legal obligation rather than as empirical findings. This distinction is maintained deliberately: a very large fraction of what the industry treats as established safety knowledge rests on normative consensus, not on evaluation.

Table 1. Composition of the retained corpus by source type, with the strongest evaluation design represented in each.

Source type	n	Strongest design represented	Principal contribution
Meta-analyses and systematic reviews	7	Meta-analysis of mixed designs	Climate–outcome and demand–resource associations; safety-rule theory
Field trials (randomised or cluster-randomised)	2	Cluster-randomised	Supervisor safety communication; behaviour-observation effects
Quasi-experimental and longitudinal evaluations	5	Interrupted time series	Behaviour-based programme durability; regulator trend data
Cross-sectional and observational studies	8	Cross-sectional	Climate measurement; human and organisational factors in QRA
Theoretical and conceptual works	6	—	Barrier theory; systems accident models; resilience framing
Public inquiry and investigation reports	9	—	Authoritative causal accounts of catastrophic events
Standards, recommended practices, regulations	12	—	Normative control requirements and indicator definitions

The distribution is itself the review's first substantive finding. Across the entire retained corpus, two randomised field trials exist, neither conducted on oil and gas technical operations. Every claim about what works during a turnaround, a well intervention, or a commissioning campaign currently rests on transfer from other settings, on quasi-experimental single-site evaluation, or on normative consensus. Two things follow, and neither is an argument for uniform interpretive caution. The absence of trials is a structural access problem — organisational interventions at project sites need the cooperation of parties whose commercial interests are engaged by the result, and their outcomes are rare events — not evidence of ineffectiveness; Section 6 sets out a design that resolves the access constraint. And the interventions with the strongest warrant here are not trial-evidenced at all: the reliability of a blind flange, a proof-tested SIL-rated trip, or a tested dual barrier envelope rests on mechanics and functional-safety mathematics, a firmer basis than a field study rather than a weaker one.[50–53]

3. Conceptual Framework

3.1. Barriers, Energy, and the Anatomy of a Project-Site Event

We adopt a barrier-management framing as the organising theory, because it is the framing that the sector's own regulatory instruments already use and because it maps cleanly onto the physical reality of technical operations. In this account, harm occurs when a hazardous energy or substance is released along a pathway to a target, and safety is the state in which a defined set of barriers — physical, technical, operational, and organisational — stands between the two, each with a specified function, performance requirement, and means of verification.[15,16] The framing is descendant from Reason's organisational-accident model, in which defences are neither perfect nor static but perforated by latent conditions that shift over time,[7,8] and from Turner's earlier demonstration that disasters incubate over long periods through the accumulation ofunnoticed, individually tolerable deviations.[6]

What makes technical operations distinctive within this framing is that they are, by definition, periods during which barriers are *deliberately removed*. A turnaround exists in order to open

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

equipment that is normally closed. A well intervention exists in order to enter a pressure envelope that is normally sealed. Commissioning exists in order to energise systems that have never held their design load. The safety problem is therefore not the maintenance of a static barrier set but the controlled substitution of temporary barriers for permanent ones, the verification that the substitution is complete before energy is admitted, and the restoration of the permanent set afterwards. Permit-to-work and energy-isolation systems are precisely the administrative machinery for managing that substitution,[29,30] which is why their failure recurs so consistently in the inquiry record.[1,3]

3.2. Drift, Pressure, and the Limits of the Barrier Account

A pure barrier account is necessary but insufficient, for two reasons the literature has established firmly. The first is directional drift. Rasmussen's model of migration to the boundary explains why barrier sets degrade systematically rather than randomly: efficiency and workload gradients push practice toward the boundary of acceptable performance, and absent a counter-gradient the boundary is eventually crossed without anyone deciding to cross it.[9] Leveson's systems-theoretic development makes the same point in control terms, and Hopkins's analysis of Macondo is the definitive sectoral demonstration: at every decision point the choice that saved rig time was taken, and each was defensible in isolation.[10,13] The second is that written procedure and actual work diverge. Hale and Borys distinguish a top-down model, in which rules are designed centrally and deviation is violation, from a bottom-up model, in which rules are local representations of competent practice and rigid adherence can itself be unsafe; most organisations manage the first while relying on the second.[32,33] This is directly consequential for permit systems, whose formal compliance rate can be high while their functional integrity is hollow. Hollnagel's Safety-II framing and Dekker's work on error attribution converge on the practical implication — interventions that treat frontline adaptation purely as a deficit remove the mechanism by which the system currently succeeds — while Weick and Sutcliffe supply the positive counterpart in preoccupation with failure and deference to expertise rather than rank.[11,12,34]

3.3. The Intervention Taxonomy Used in This Review

Combining these, we classify interventions into four layers by the failure mode each can plausibly control. The layers are not a hierarchy of value; they are a division of labour, and the review's central argument is that most project-site portfolios are heavily weighted to layers two and three while the inquiry record indicts layers one and four.

Table 2. The four-layer intervention taxonomy used throughout this review, classified by the failure mode each layer can plausibly control.

Layer	Intervention class	Failure mode controlled	Typical instruments
L1 —Barrier integrity	Engineered and procedural controls on hazardous energy	Uncontrolled release during barrier substitution	Permit to work; energy isolation and LOTO; hot-work and confined-space controls; SIMOPS management; barrier verification and override registers [28–30]
L2 — Competence	Task-specific capability and crew assurance	Correct intent, inadequate capability; unfamiliar crew combinations	Competence assurance frameworks; task risk assessment; pre-job briefing and toolbox review; simulation for well control [31]
L3 — Disclosure	Willingness to report, question, and stop	Known hazard not surfaced in time to act	Stop-work authority; near-miss and weak-signal reporting; supervisor safety communication; life-saving rules [23,41]

 doi:10.20944/preprints202609.1107.v1

L4 — Governance	Control of the contracting and decision interface	Accountability diffusion; schedule pressure resolved against barriers	Contractor safety management; bridging documents; management of change; process-safety indicator review at board level [14,22,25–27]

4. Results

4.1. Layer 1 — Barrier Integrity in Technical Operations

Permit-to-work is the single most consistently implicated barrier in serious project-site events, and its evidence base is almost entirely normative rather than experimental. UK HSE guidance sets out the functional requirements — a written record of the work, the hazards, the precautions taken, the isolations verified, the duration authorised, and the identified accepting and issuing persons, with explicit handover at shift change and explicit cancellation on completion.[30] The Cullen inquiry's finding was not that Piper Alpha lacked such a system but that the system was administered in a way that allowed a suspended permit on a removed pressure-safety valve to be unknown to the night crew who restarted the associated pump.[1] The failure was informational and organisational: shift handover, permit suspension conventions, and the physical marking ofisolated equipment.

Energy isolation is more amenable to engineered rather than administrative control, and the regulatory instruments reflect this. The U.S. lockout/tagout standard specifies the control of hazardous energy through positive isolation, verification of a zero-energy state, and individual employee locks rather than a shared or supervisory lock— a design choice that removes reliance on communication between shifts.[29] The process-safety management standard adds requirements that bear directly on project work: process hazard analysis, management of change, pre-startup safety review, mechanical integrity, contractor obligations, and — notably for turnarounds — hot-work permitting.[28] The presence of these requirements in law since 1992 alongside the continuing recurrence ofthe same failure modes is the clearest available evidence that the L1 problem is one ofverification and assurance rather than ofspecification.

Barrier theory offers a way to make that verification tractable. Sklet's classification distinguishes barrier *function* (what must be achieved) from barrier *system* (what achieves it) and specifies performance requirements — functionality, reliability, availability, response time, robustness — that can be individually verified.[15] Duijm's barrier-diagram method operationalises this as a management tool by making the assignment of each barrier to an accountable owner explicit.[16] Applied to a turnaround, this yields something operationally useful: a live register of which permanent barriers are currently impaired, which temporary barriers are substituting for them, who owns each, and what the accumulated impairment across concurrent work packages amounts to. The Norwegian regulator's risk-level programme provides sectoral evidence that such barrier-condition indicators can be tracked systematically at scale and that they move independently ofinjury rates.[21]

4.1.1. Simultaneous Operations as the Aggregation Problem

The distinctive project-site hazard is not any single impaired barrier but the aggregation of many, held by different employers, in physical proximity, under one schedule. Piper Alpha's proximate mechanism was concurrent maintenance during production;[1] Texas City involved a startup adjacent to occupied temporary trailers.[3] Quantitative risk assessment methods can in principle represent such interactions,[19] and work on incorporating human and organisational factors into offshore QRA has shown that the omission of these factors systematically biases risk estimates in the optimistic direction.[20] Aven's review of the risk-assessment field notes the broader limitation relevant here: assessments conducted at design intent do not represent the plant state that actually exists during a shutdown.[18] The practical implication is that simultaneous-operations control requires a live, site-level aggregation of impairment— not a set ofindependently approved permits.

 doi:10.20944/preprints202609.1107.v1

4.1.2. Engineered Controls: What Actually Holds When Administration Fails

The preceding analysis of permit and isolation failure invites an obvious engineering response, and it is one the standards corpus states clearly while project practice under-applies. If the demonstrated failure mode is informational — a permit not communicated at handover, an isolation believed rather than verified, a gas test assumed still valid — then the highest-reliability intervention is one that removes the dependence on information transfer altogether and replaces it with physical constraint. This is the hierarchy of control applied to an administrative barrier,[17] and it is the single most under-exploited intervention route in technical operations.

Kletz's principle of inherently safer design supplies the governing logic: intensify to reduce inventory, substitute a less hazardous medium, attenuate to reduce stored energy, and simplify to reduce the number of ways the system can be got wrong.[50] Applied to project work the principle bites hardest at design and planning stage, because it changes whether a hazardous technical operation needs to be performed at all. Specifying double-block-and-bleed with a monitorable bleed point at design stage means a positive isolation can be *demonstrated*rather than asserted; a single valve with a tag cannot be. Providing spool-piece removal points converts an isolation that depends on valve integrity into an air gap that is visible from ten metres. Designing for online maintainability, or for module replacement rather than in-situ repair, removes the shutdown that generates the exposure. Locating instrumentation and sample points outside vessels removes confined-space entry from the maintenance scope permanently, for every future turnaround, at no recurring cost. None of these decisions is available to a site safety department, and all of them outperform any behavioural or procedural control on every dimension of barrier performance that Sklet's classification measures.[15,50]

Where the hazard cannot be eliminated, instrumented protection is the next most reliable layer, and unlike most of the intervention corpus it comes with a quantified reliability framework. IEC 61511 establishes the safety-lifecycle and safety-integrity-level (SIL) architecture for process-sector safety instrumented systems, deriving each function's required risk reduction and specifying the proof-test regime that maintains it; IEC 61508 supplies the underlying functional-safety framework.[51,52] The practical significance for technical operations is that a SIL-rated function has a stated probability of failure on demand and a mandated test interval — which means the question "is this barrier currently functioning?" has a defensible answer, whereas the same question asked of a toolbox talk does not. API RP 14C's systematic analysis of process components on offshore facilities performs a comparable service on the design side, specifying the primary and secondary protection required for each component so that protection completeness is auditable rather than judged.[54] CCPS's engineering-design guidance and Lees' compendium collect the equivalent body of practice for onshore process plant.[57,58]

The well-operations domain provides the clearest case in the sector where an engineered barrier philosophy has been formalised to the point of being auditable. NORSOK D-010 requires two independent, tested, and monitorable well barrier envelopes for every well activity, with a barrier schematic maintained for each operational phase and each element's acceptance criteria, monitoring method, and verification stated explicitly.[53] API Standard 53 specifies blowout-prevention equipment systems for drilling wells, including the accumulator capacity, function-test, and pressure-test regime that keeps that equipment credible.[55] This is what a mature barrier-management regime looks like in practice, and it is instructive that it exists in the one technical operation whose catastrophic potential the industry has never been able to mistake for an injury problem. The reason Macondo remains so instructive is not that this framework was absent but that its verification step — the negative-pressure test — was interpreted rather than measured against a pre-declared pass criterion.[4,5,13,20]

Comparable engineered substitutions exist across every operation in scope and are set out in Table 3. Their common structure is worth naming: each replaces a control that depends on a person knowing something with a control that depends on a physical state, and each is specified in a standard rather than evidenced by a trial. That is the honest status of the sector's most reliable interventions — expert-consensus engineering rather than evaluated behavioural science — and it is a firmer foundation than the phrase “no controlled evidence” suggests, because the reliability ofa blind flange is established by mechanics rather than by a field study.

Table 3. Engineered substitutions by technical operation. "Reliability class" records whether the control' s performance can be verified by physical test independent of human report — the property that distinguishes it from the administrative barriers whose failure dominates the inquiry record.

Technical operation	Dominant hazardous energy	Engineered intervention that displaces administrative reliance	Reliability class
Drilling & well intervention	Formation pressure; hydrocarbon influx	Two independent tested barrier envelopes with maintained schematics; [53] BOP with specified function/pressure-test regime; [55] managed-pressure drilling with closed circulating system; pre-declared numeric pass criteria for negative-pressure tests	Engineered, verifiable
Turnaround / shutdown maintenance	Residual pressure, hydrocarbon, stored mechanical energy	Double-block-and-bleed with monitored bleed; spool removal creating a visible air gap; individual employee locks in place of shared supervisory locks; [29] nitrogen purge and inerting verified by oxygen analysis	Engineered, verifiable
Hot work on or near live plant	Ignition of flammable atmosphere	Elimination by cold cutting and mechanical bolting; hot-tapping under the specified procedural and design controls where breaking containment is worse; [56] continuous fixed gas detection interlocked to work stoppage rather than a spot reading with an assumed validity window	Engineered, partly administrative
Confined-space entry	Asphyxiation; toxic atmosphere; inadvertent energisation	Design-stage elimination via external instrumentation and sample points; continuous rather than pre-entry-only atmospheric monitoring; remote inspection by crawler, drone, or borescope in place of entry	Engineered where designed in
Pipeline laying & hydrotesting	Stored pressure energy; stored elastic energy	Water rather than pneumatic testing, reducing stored energy by orders of magnitude at equal test pressure; calculated and physically barricaded exclusion zones; remote pressurisation and remotely read instrumentation	Engineered, verifiable
Lifting & heavy rigging	Gravitational potential energy	Engineered lift plans with load-path exclusion designed rather than marshalled; remote-operated and semi-automated handling; secondary retention on all elevated components; pre-slung and pre-assembled modules reducing lift count	Engineered, partly administrative

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

Facility construction & commissioning	First energisation of untested systems	Off-site modular fabrication removing work from the live facility; staged system handover with documented pre-startup safety review; [28] SIL-rated protective functions proof-tested before first hydrocarbon introduction [51,52]	Engineered, verifiable
Offshore installation & SIMOPS	Aggregated concurrent impairment	Physical separation and blast-resistant siting of temporary accommodation; [2,3] ESD and blowdown systems with tested response times; automated permit-to-barrier linkage preventing energisation while an isolation remains open	Engineered, verifiable

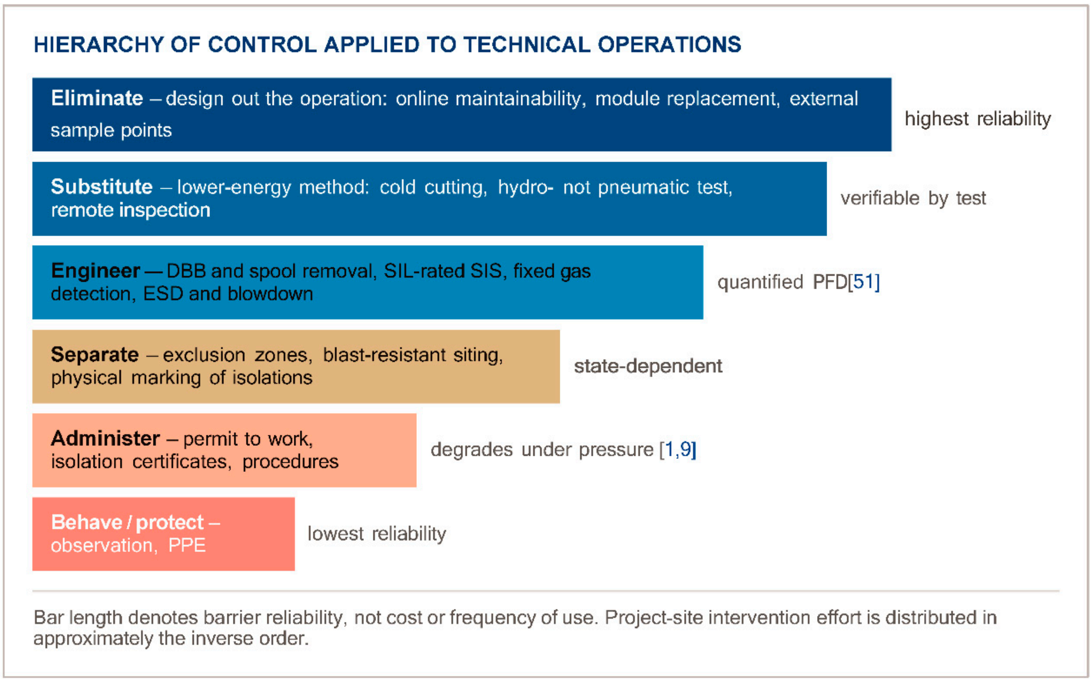

Figure 2. The inversion between reliability and deployment effort. Interventions in the upper three bands are decided in engineering design and procurement; those in the lower three are decided at the site, which is where safety intervention resource is concentrated.

4.2. Layer 2 — Competence Assurance

Competence interventions have the clearest theoretical warrant and the weakest evaluation base in the corpus. HSE's management framework treats competence as a system requirement — identifying the competences each safety-critical role requires, assuring them, and maintaining them — rather than as a training-delivery activity.[31] ISO 45001 imposes a parallel requirement, with the significant addition that it extends to workers who are not the organisation's own employees, which is where project sites fail.[27]

Two structural features undermine competence assurance in ways certification does not capture. The tasks with the greatest catastrophic potential are performed rarely — a crew may execute a specific well-control or isolation sequence once in a project — so competence cannot be maintained by repetition and must be maintained by simulation, drill, and verified procedure. And competence at project sites is a property of *crews*, not individuals: a permit is issued by one employer's authorised person, accepted by a second employer's supervisor, and executed by a third employer's technicians,

 doi:10.20944/preprints202609.1107.v1

none of whom has necessarily worked with the others before. Certification schemes assure the individuals; nothing in the standard corpus assures the combination, and we found no evaluation in the retained corpus measuring a safety outcome against crew-level familiarity or against the integrity of the competence chain across a contracting tier — an absence that is striking given how routinely inquiry reports describe that interface as the locus of failure.[1,3,5]

The closest available evidence concerns pre-task planning. Studies of toolbox talks and pre-job hazard review report improvements in hazard recognition and in the specificity of identified controls, but effects are measured on intermediate cognitive outcomes rather than incident rates, and durability beyond the intervention period is rarely assessed.[42,43] The honest summary is that pre-task planning is well supported as a mechanism for surfacing task-specific hazards and unevidenced as a mechanism for reducing high-energy events.

4.3. Layer 3 — Disclosure: Behaviour-Based Safety and Stop-Work Authority

Behaviour-based safety (BBS) is the most widely deployed and most extensively evaluated intervention class in the corpus, and the one whose applicability to technical operations is most questionable. The method derives from applied behaviour analysis: define observable critical behaviours, sample them through peer observation, feed results back to the work group, and reinforce the safe alternative.[35] Meta-analytic and review evidence reports consistent positive effects on the observed behaviours and generally favourable but smaller and noisier effects on injury outcomes, strongly dependent on implementation fidelity, feedback frequency, and management participation; effects decay when observation frequency falls, so the intervention functions as a maintained programme rather than a permanent control.[35–37]

Three boundary conditions determine whether that evidence transfers to a project site. BBS requires behaviours that are *observable*, *repeated*often enough to sample, and causally proximate to the harm. Personal-safety behaviours satisfy all three: hand placement, three-point contact, fall-protection tie-off, and load-path clearance are visible, continuous, and directly proximate to the injuries they prevent. The high-energy decisions in technical operations satisfy none — whether a blind flange is the correct isolation for this pressure, whether a gas test remains valid after a two-hour break, whether a negative-pressure test result should be accepted, whether a lift plan survives a wind shift — are cognitive judgements made once, often by a single competent person, and invisible to an observer. This distinction, not any general objection to behavioural methods, is why BBS should be understood as a well-evidenced control for the frequent-injury stream and as structurally unable to address the catastrophic stream. Read alongside the Baker Panel's finding, it is a mechanistic explanation of how an organisation can raise observation counts and lower injury rates while its major-accident risk is unchanged.[2]

Stop-work authority addresses disclosure at the moment of exposure rather than through sampling, and its evidence base is thinner but its logic stronger for non-routine work. IOGP's Life-Saving Rules embed an explicit intervention obligation alongside a small set of absolute rules, on the reasoning that a short, memorable rule set travels across the contracting chain and across language barriers better than voluminous procedure;[23] evaluation is largely limited to adoption and self-reported willingness to intervene. The strongest transferable causal evidence for a disclosure intervention is Kines and colleagues' cluster-randomised trial of supervisor safety communication, which improved safety-climate and behavioural outcomes at construction sites and remains the corpus's clearest demonstration that supervisor behaviour is a manipulable causal variable rather than a correlate.[41] Two qualifications matter. The mechanism by which stop-work authority works is the removal of anticipated cost for the person exercising it, and that cost is concentrated where it is least visible to the operator: a subcontracted technician on a day rate, whose next assignment depends on a supervisor's view ofwhether they are difficult to work with, faces a materially different calculus from a directly employed operator. And the safety-climate literature underwriting this layer is measurement-heavy and intervention-light — climate is established as measurable and predictive, and the demand–resource pathway to outcomes is meta-analytically supported,[38–40] but none of it establishes what changes climate at a site that will be demobilised in eleven weeks.

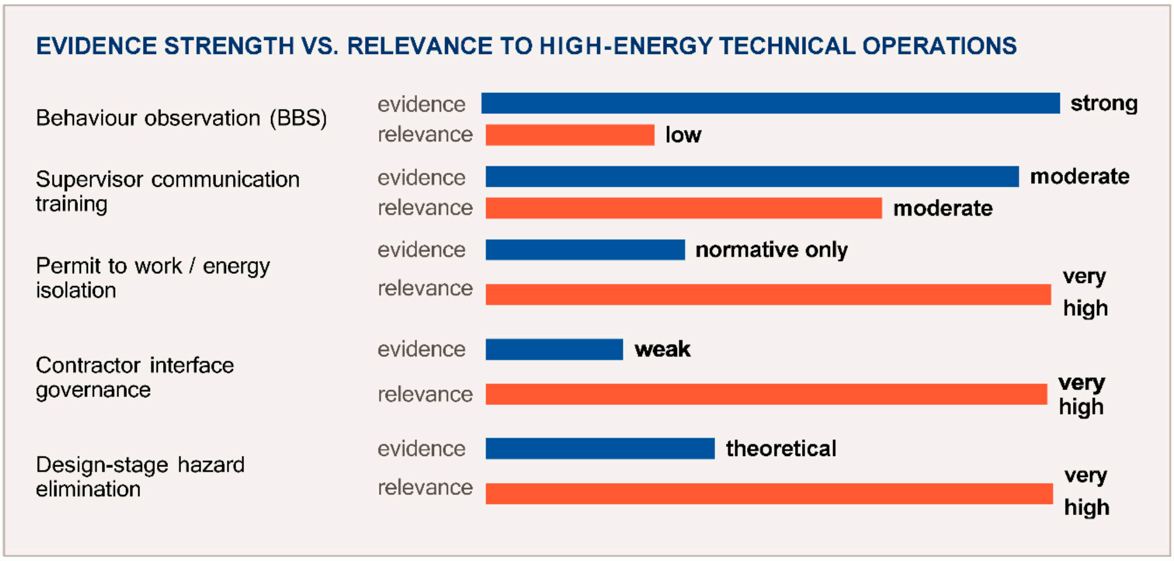

Figure 3. The core misallocation. Bars show the author' s ordinal appraisal of published evidence strength (blue) against relevance to high-energy failure modes in technical operations (orange), charted per Section 2.6. The two rank in near-inverse order: the best-evidenced interventions address the least consequential failure modes.

4.4. Layer 4 — Governance of the Contracting and Decision Interface

Project sites are contractual constructions. An operator holds the licence and the safety case; an engineering, procurement and construction contractor holds the schedule; specialist subcontractors hold the trades; and labour-supply intermediaries hold the individual employment relationships. Each interface is a discontinuity in the flow of safety information, and each is a point at which accountability can be asserted to lie elsewhere. Every inquiry in the retained corpus identifies this structure as causally implicated: Piper Alpha's permit failure occurred across a contractor–operator maintenance interface;[1] Macondo's decision sequence involved an operator, a drilling contractor, and a cementing service company with divergent commercial exposure and no shared risk picture.[4,5,13]

The normative response is well specified. The OSHA process-safety standard obliges host employers to inform contractors of process hazards and to verify contractor employee training and performance;[28] ISO 45001 requires coordination of the management system with contractors and extends consultation duties to non-employed workers;[27] and IOGP's contractor-management guidance sets out a lifecycle model — pre-qualification, selection, pre-mobilisation, execution monitoring, close-out — with bridging documents reconciling differing systems at the interface.[26] IOGP's annual performance data, the sector's most consistent longitudinal series, report that contractor personnel account for a disproportionate share of fatalities relative to hours worked.[22]

Against this, the peer-reviewed evaluation base is nearly empty. We identified no controlled evaluation of a contractor safety-management intervention in the retained corpus. The available studies are cross-sectional or case-based and identify plausible mechanisms — pre-qualification screening on lagging metrics that are themselves gameable, incentive structures that transfer schedule pressure down the tier without transferring authority, and safety costs treated as a competitive variable in bid evaluation — without testing whether altering them changes outcomes.[44,45] This is the corpus's largest single gap: the intervention layer that inquiry evidence most strongly indicts is the layer with the least evaluation evidence, and it is a gap with an obvious structural cause. Contractor governance is not manipulable by a site safety department, so it is not something a site-level researcher can trial.

 doi:10.20944/preprints202609.1107.v1

The indicator problem is the mechanism by which governance failure becomes invisible. API RP 754 defines a tiered process-safety event framework — Tiers 1 and 2 for consequence-classified losses of primary containment, Tier 3 for challenges to safety systems, Tier 4 for operating discipline — explicitly so the higher tiers can be monitored without waiting for the lower ones to occur,[25] and HSE's guidance makes the paired-indicator logic explicit: every lagging indicator should be matched to a leading indicator of the barrier intended to prevent it.[24] Texas City demonstrates what happens when board attention is directed at an indicator set excluding the relevant barriers.[2,3] Hopkins's synthesis — that organisations measure what is easy and infer safety from it — remains the most accurate short statement of the failure.[14] The corrective is not to abandon injury metrics but to stop treating them as the outcome variable for interventions aimed at catastrophic risk. Table 5 sets out the substitution: independently auditable barrier-condition measures — verified-isolation share, concurrent impairment against a declared ceiling, competence-matrix closure at permit issue, tier-disaggregated stop-work origination — each paired with the lagging measure it is theorised to move. Two properties make these usable as evaluation endpoints where injury rates are not: they occur frequently enough to power a trial of feasible size, and they do not depend on voluntary reporting.

4.4.1. Specifying a Testable Contractor-Interface Intervention

Identifying the contracting interface as the least-evaluated and most-indicted layer is only useful if it can be converted into something evaluable. The reason it has not been is partly that "contractor safety management" names a lifecycle rather than an intervention: pre-qualification, selection, mobilisation, monitoring, and close-out[26] are five distinct control points with different owners, and a study of all five simultaneously has no interpretable result. We therefore specify five discrete, separately manipulable components, each stated so that compliance is observable and each targeting a mechanism the inquiry record identifies. This is offered as the intervention definition the field currently lacks, not as validated practice.

1. **Single permit authority irrespective ofemployer.** One named technical authority, employed by the duty holder, issues and closes every permit on the site and owns the impairment register; contractor supervisors accept work but cannot authorise isolation. Targets accountability diffusion at the point Piper Alpha's permit failed.[1]
2. **Barrier scope allocated at bid stage.** Tender documents state which party owns each safety-critical barrier during each work phase, and unallocated barriers block contract award. Targets the assumption — recurrent in investigation findings — that a barrier everyone believed someone else was holding was held by nobody.[3,5]
3. **Contractual stop-work indemnity.** The contract prohibits demobilisation, day-rate reduction, or negative performance evaluation ofany individual within 90 days oftheir exercising stop-work authority, with the operator auditing personnel changes against stop-event records. Targets the cost asymmetry that makes stop-work authority nominal for subcontracted personnel.[23]
4. **Pre-qualification scored on barrier verification, not lagging metrics.** Recordable-injury history is removed from bid scoring and replaced by audited evidence ofisolation-verification practice, competence-matrix closure, and Tier 3 reporting maturity.[25] Targets the gaming incentive that lagging-metric screening creates, and the false assurance it supplies to the operator.[2]
5. **Mandatory tier-disaggregated reporting.** Every near-miss and stop-work event is recorded with the originating party's contracting tier, and a subcontractor-origination share below a pre-set threshold triggers investigation ofthe reporting system rather than congratulation. Targets the invisibility ofthe lowest tier in site safety data.

Components 1, 2, and 4 are contractual or organisational and can be implemented at portfolio level by a single duty holder, which makes them randomisable by site in the design set out in Section 6. Components 3 and 5 require supply-chain agreement and are better suited to stepped implementation across a whole portfolio with an interrupted time-series analysis. Stating them separately also allows a reviewer or a regulator to reject any one without discarding the others — a property the undifferentiated "contractor safety management" construct does not have.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

4.5. Comparative Synthesis: Failure Modes Against Intervention Coverage

Charting the failure modes identified across the nine inquiry and investigation reports against the intervention classes evaluated in the peer-reviewed corpus permits a direct comparison of indictment frequency against evaluation attention. We coded each report for the failure modes its findings assert (a report may assert several) and each peer-reviewed evaluation for the failure mode its intervention targets. The comparison is descriptive and the denominators are small; it is reported as a coverage map, not as an effect estimate.

Table 4. Coverage map. Inquiry-asserted failure modes charted against the failure modes targeted by the 22 peer-reviewed evaluations and reviews in the retained corpus. Counts are of sources, not of effect estimates.

Failure mode	Asserted in inquiries (n/9)	Evaluations targeting it (n/22)	Strongest design available	Coverage verdict
Permit / isolation administration failure	7	1	Cross-sectional	Severely under-evaluated
Accountability diffusion across contracting tiers	7	0	None	Not evaluated
Schedule/cost pressure resolved against barriers	8	1	Case study	Severely under-evaluated
Indicator set blind to process-safety barriers	5	2	Longitudinal observational	Under-evaluated
Simultaneous-operations impairment aggregation	4	1	Modelling only	Under-evaluated
Competence gap on rarely performed tasks	4	3	Pre-post	Under-evaluated
Weak signals not surfaced / intervention withheld	5	5	Cluster-Randomized [41]	Adequately addressed
Unsafe individual behaviour on repetitive tasks	1	9	Meta-analysed trials [35–37]	Over-evaluated relative to indictment

The pattern is unambiguous and is the review's principal result. The three failure modes most frequently asserted by authoritative causal investigations — permit administration, accountability diffusion, and the resolution of commercial pressure against barrier integrity — are supported by a combined total of two peer-reviewed evaluations, neither better than case-study design. The failure mode least frequently asserted — individual unsafe behaviour on repetitive tasks — accounts for nine. An effect-direction reading reinforces this: among the nine behaviour-focused evaluations, direction of effect on observed behaviour is positive in all nine and on injury outcomes positive in seven of nine, so the evidence is not weak; it is simply pointed at a different problem than the one that produces catastrophic loss.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

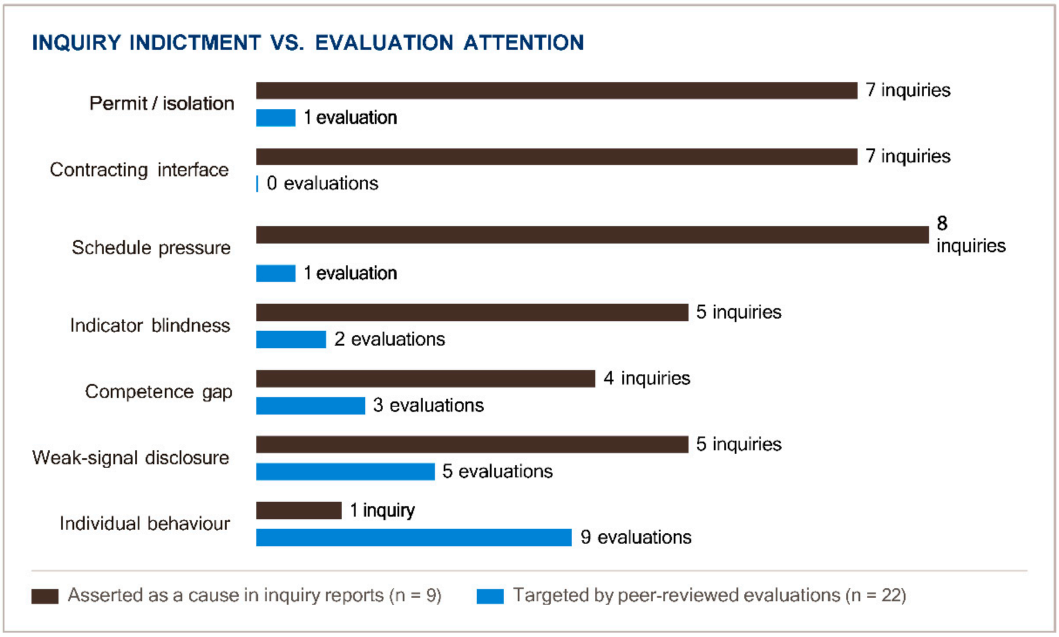

Figure 4. Bars are scaled within each series to its own denominator, so the comparison is of relative attention rather than of absolute counts. The inversion between the two series at the top and bottom of the chart is the misallocation this review reports.

4.6. Where Safety Intervention Demonstrably Works

A review organised around misallocation risks reading as an argument that nothing works, which would be both false and unhelpful to practitioners. Three groups of technical operations in scope show a genuine, traceable record of intervention success, and their common features are more instructive than the failures.

Well control. Blowout frequency per well drilled has fallen substantially over the era covered by this review, and the intervention set responsible is identifiable: mandated redundant barrier envelopes with maintained schematics,[53] prescribed BOP function and pressure testing at defined intervals,[55] mandatory certificated well-control training with simulator-based recertification, and — post-Macondo — pre-declared numeric acceptance criteria for barrier verification tests rather than engineering interpretation at the wellsite.[4,5] This combination pairs an engineered barrier with a competence regime and a verification criterion, which is precisely the layered structure the framework in Section 5 generalises.

Instrumented protection against loss ofcontainment. The SIL architecture works, in the specific and demanding sense that a protective function's probability of failure on demand can be specified in advance, verified by proof test, and re-verified on a mandated interval.[51,52] No behavioural intervention in the corpus can make an equivalent claim. Where the sector has moved detection and shutdown from human observation to interlocked instrumentation — fixed gas detection tied to automatic isolation, high-level trips, ESD and blowdown with tested response times — the barrier stops depending on who is looking.

Dropped objects, lifting, and personal injury. Secondary retention on elevated components, engineered lift plans with designed rather than marshalled exclusion zones, remote-operated handling, and pre-slung modular assembly have produced measurable reductions in dropped-object and lifting incidents in sector performance data — again by substitution: fewer lifts, and a physical tether where previously there was a rule about standing clear.[22] The same data show falls, hand injuries, slips, and struck-by events declining across three decades, and the meta-analytic evidence supports a causal contribution from observation and feedback programmes on exactly this class of frequent, observable, low-energy exposure.[35–37] The error identified in Section 4.3 is not that these programmes fail; it is that their success on this stream is read as assurance on another.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

Three properties recur across all three. Each pairs an engineered or instrumented barrier with a competence requirement and an explicit verification criterion — no case rests on a single layer. Each has a barrier state that can be tested independently of anyone's report. And each is mandated by a standard or regulation rather than adopted at site discretion, which is what makes it survive schedule pressure. The framework that follows is an attempt to state those three properties as a general design rule rather than to introduce anything new.

5. An Integrated Intervention Framework for Technical Operations

The results support a specific reallocation rather than a general call for more safety effort. We set out the framework as four layers, each defined by the failure mode it controls, the verification that demonstrates it is functioning, and the accountable role that owns it. Its claim to integration is not that it combines existing programmes but that it makes each layer's coverage boundary explicit, so that no layer is credited with controlling a failure mode it cannot reach.

INTEGRATED FOUR-LAYER INTERVENTION FRAMEWORK — TECHNICAL OPERATIONS			
Layer	**Control intent**	**Verification (not activity count)**	**Accountable owner**
L1 Barrier integrity	Hazardous energy stays isolated during barrier substitution; impairment is aggregated site-wide, not permit-by-permit	Independent physical verification of isolation before energy admission; live impairment register reviewed daily against a declared ceiling	Operator's technical authority —never delegable to the executing contractor
L2 Competence	Rarely performed safety-critical tasks are executed by verified-capable *crews*, not merely certified individuals	Task-level competence matrix closed out before permit issue; simulation or dry-run for first-execution tasks; crew-familiarity flag on the permit	Discipline authority, jointly with the employing contractor
L3 Disclosure	A worker who sees a degraded barrier can stop the job at no personal cost, regardless of employer or tier	Stop-events traced by originating tier and outcome; contractual protection against demobilisation following a stop; behaviour observation retained for the personal-safety stream only	Site leadership, with contractual force behind it
L4 Governance	Schedule and cost pressure cannot be resolved against barrier integrity without a decision that is visible above the site	Bridging document signed before mobilisation; barrier-impairment escalation authority held off-site; Tier 3/4 indicators on the same report as injury rates	Accountable executive / project sponsor

Figure 5. The framework' s operative feature is column three: each layer is evidenced by a verification of state, not by a count of safety activity. Activity counts are the mechanism by which the Texas City indicator failure occurred.2,3.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

Three design commitments distinguish this from a conventional layered model. L1 verification is *independent* of the executing party — the recurring permit failure is not that no one checked but that the checker held the schedule incentive. L3 narrows behaviour-based safety to the stream where its evidence applies, which frees observation resource and, more importantly, removes the false assurance that observation coverage supplies. And L4's escalation authority sits deliberately off site, because a site facing a deadline resolves ambiguity in favour of the deadline, and no amount ofsite-level exhortation changes the gradient Rasmussen described.[9]

5.1. A Matched Indicator Set

Following the paired-indicator logic of HSE's guidance and API RP 754's tier structure,[24,25] each layer is assigned leading indicators of barrier condition alongside the lagging indicators it is intended to move. The set is deliberately short; the failure mode being guarded against is an indicator suite large enough that no one reads it.

Table 5. Matched leading and lagging indicator set, paired per the logic of HSG254 and API RP 754.24,25.

Layer	Leading indicator (barrier condition)	Matched lagging indicator	Review cadence
L1	% isolations independently verified before energy admission; count of concurrently impaired safety-critical barriers against declared ceiling; overdue safety-critical maintenance on live systems	Tier 3 safety-system challenges; loss-of-containment events (Tier 1/2) [25]	Daily at site; weekly to project
L2	% safety-critical permits with a closed competence matrix; % first-execution tasks preceded by dry-run or simulation; crew-familiarity flag rate	Events attributed to task execution error; rework on safety-critical work packages	Weekly
L3	Stop-work events per 100k hours disaggregated by contracting tier; near-miss reports originating from subcontracted personnel as a share of total	High-potential incidents; recordable injury rate (personal-safety stream only)	Weekly
L4	% contractors mobilised with a signed bridging document; barrier-impairment escalations resolved off site; interval between Tier 3 report and executive review	Tier 1/2 event rate; contractor share of high-potential events vs. hours [22]	Monthly to accountable executive

Two indicators in this set are unconventional and are the ones we would defend most strongly. Disaggregating stop-work events by contracting tier converts a vanity metric into a diagnostic: a site reporting many stops, all originating from directly employed staff, has demonstrated that its disclosure layer does not reach the people doing the highest-exposure work. And counting concurrently impaired barriers against a declared ceiling makes the simultaneous-operations aggregation problem visible as a single number that a project manager must either respect or explicitly override — which is exactly the decision that inquiry reports find was made implicitly.[1,3]

6. Research Agenda: Making the Framework Testable

The framework as stated is a hypothesis, and the review's own appraisal standard requires that it be treated as one. The obstacle to testing interventions of this kind is well understood: individual randomisation is impossible because the interventions are organisational, and parallel-group cluster randomisation is commercially unacceptable because it requires withholding a plausibly protective control from half the sites. A stepped-wedge cluster-randomised design resolves both, because every cluster eventually receives the intervention and the randomisation is of*timing*rather than of allocation.[46,47] This design is now well established in health-services research for exactly this class of problem and has been argued for in occupational safety, where the same political constraint applies.[48]

6.1. Proposed Design

Unit of randomisation. The project site or turnaround campaign, not the individual worker. **Clusters.** A minimum of 12 sites within a single operator's portfolio or a consortium, to hold contracting conventions and regulatory regime approximately constant. **Sequence.** Four steps of three clusters, each step a calendar quarter, with all clusters contributing control periods before their own crossover and intervention periods after it. **Intervention.** Layers L1 and L4 only — independent isolation verification, the site-wide impairment register with a declared ceiling, and the contractor-interface components specified in Section 4.4.1 — single permit authority irrespective of employer, bid-stage barrier scope allocation, pre-qualification scored on barrier verification rather than lagging metrics, and off-site escalation authority. Components 3 and 5 of that set (stop-work indemnity and tier-disaggregated reporting) are excluded from the randomised arm because they require supply-chain-wide agreement and cannot be withheld from control clusters; they are proposed instead for portfolio-wide implementation with interrupted time-series analysis. L2 and L3 are held constant, because bundling all four layers would produce a result that could not be attributed. **Primary outcome.** Tier 3 safety-system challenge rate per 100,000 exposure hours, as defined by API RP 754[25] — chosen because Tier 1 events are too rare for a trial of feasible size, and because Tier 3 is the tier the intervention is theorised to act on. **Secondary outcomes.** High-potential incident rate; independently audited isolation-verification compliance; contractor-originated near-miss share; schedule variance, included deliberately so that the commercial objection can be tested rather than asserted. **Analysis.** Mixed-effects Poisson regression with cluster as a random effect and calendar time as a fixed effect, the latter being essential in stepped-wedge designs to separate intervention effect from secular trend.[47]

6.2. Anticipated Difficulties

Three are worth stating plainly. Contamination is likely and largely unavoidable: personnel and supervisors move between sites within a portfolio, carrying practice with them, which biases the estimate toward the null and should be reported as a conservative-bias limitation rather than corrected away. Blinding is impossible for participants; outcome ascertainment can be blinded by having Tier 3 classification performed by assessors without allocation-sequence knowledge. And the reporting-rate paradox is severe — an intervention that improves disclosure raises the measured rate of the primary outcome. This must be handled at design time, by pairing the Tier 3 rate with an independently audited barrier-condition measure that does not depend on voluntary reporting, and by pre-specifying that a rise in Tier 3 reports alongside improved audited barrier condition is interpreted as success, not failure. We regard the absence of this pre-specification as a likely reason such trials have not been attempted.

7. Discussion

The finding that intervention effort is misallocated relative to the risk profile is not new as an assertion — the Baker Panel made it in 2007[2] — but this review's contribution is to show that the misallocation is reproduced in the *research*literature and not only in industry practice, and to identify why. The causes are structural rather than intellectual. Behaviour-observation interventions are evaluable because they generate high-frequency observable data on a stable population; contractor-governance interventions are not, because they require the cooperation of parties whose commercial interests are engaged by the result and because their outcomes are rare events. Researchers respond rationally to this gradient, with the consequence that the evidence base grows fastest where the risk

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

is smallest. Recognising this as a structural incentive problem rather than an oversight suggests that closing the gap requires funders and regulators to make portfolio-level access a condition of something the industry wants, rather than waiting for investigator-initiated proposals.

A second implication concerns how the sector should read its own good news. Injury rates have fallen substantially over three decades,[22] and it would be perverse to argue the interventions responsible should be abandoned. The argument is narrower: that decline is evidence about the frequent-injury stream and not about catastrophic risk, and the Piper Alpha, Texas City, and Macondo sequence establishes that the two can move in opposite directions.[1–4] The error to avoid is not doing behaviour-based safety; it is counting it as assurance against loss ofcontainment.

Third, framing permit-to-work failure as administrative rather than technical carries a design implication the standards corpus states but practice under-applies. Where reliance on communication between shifts and employers can be replaced by physical constraint — individual employee locks rather than a shared supervisory lock, double-block-and-bleed rather than a single valve with a tag, physically distinctive marking of isolated equipment — the intervention no longer depends on the informational integrity that inquiry reports repeatedly find degraded.[29,30] This is the hierarchy of control applied to an administrative barrier,[17] available at far lower cost than the cultural interventions that dominate discussion. The two-model account of procedure use sets its limit:[32,33] adding verification steps raises administrative load, and beyond some threshold that load is absorbed by frontline adaptation — pre-signing, batching, nominal completion — producing the appearance of a stronger barrier and the reality of a weaker one. The framework's insistence on verification of state rather than counts of activity is intended to be robust to this but is not immune to it, and any implementation should treat the administrative burden of its own controls as a monitored risk.

7.1. Limits of Cross-Sector Transfer

A substantial share of the intervention evidence cited here was generated outside oil and gas — in general construction, manufacturing, healthcare, and other high-hazard settings — and the review's argument depends on treating some of that transfer as sound and some as unsound. That distinction cannot rest on assertion, so we state the criteria used. Transfer is treated as warranted where the intervention's causal mechanism does not depend on features of the originating setting that differ materially from a project site, and as unwarranted where it does. Four features do the discriminating work: whether the workforce is stable or transient; whether the target activity is repeated often enough for the mechanism to operate; whether the hazard's energy density permits recovery from an error; and whether authority over the activity sits with a single employer.

Applied consistently the criteria cut both ways, which is the point of stating them.

The Kines supervisor-communication trial transfers reasonably well: its mechanism—supervisor verbal attention shifting subordinate risk salience — depends on a relationship that exists at a project site, and it was conducted in construction, where transience is comparable.[41] The behavioural-observation meta-analyses transfer to the personal-safety stream on the same reasoning and fail on the catastrophic stream, because their mechanism requires repeated observable behaviour that high-energy decisions do not supply.[35–37] The safety-climate literature transfers least well — not because its findings are weak but because climate takes months to shift and most project sites do not exist long enough at stable membership for the mechanism to run [38–40].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

Table 6. Transfer appraisal by mechanism. Verdicts follow from whether the mechanism depends on workforce stability, activity repetition, energy density, or unified employer authority — not from the perceived quality of the source evidence.

Source setting / intervention	Mechanism relied upon	Discriminating feature	Transfer verdict
Construction — supervisor safety communication [41]	Supervisor attention shifts subordinate risk salience	Supervisory relationship present; transience comparable	Warranted
Manufacturing — behavioural observation [35–37]	Sampling and feedback on repeated observable behaviour	Requires repetition and observability	Personal-safety stream only
Multi-sector — safety climate [38–40]	Shared perception accumulated over time in a stable unit	Project sites demobilise before climate consolidates	Weak
Process industry — SIL / functional safety [51,52]	Quantified probability of failure on demand, proof-tested	Physical, independent of workforce features	Fully warranted
Health services — stepped-wedge evaluation [46,47]	Randomisation of implementation timing across clusters	Method, not intervention; constraint identical	Warranted as method
Aviation / nuclear — high-reliability organizing [12]	Sustained collective mindfulness in a career workforce	Single employer, career tenure, licensed operators	Limited — assumes conditions project sites lack

The final row deserves emphasis because high-reliability organising is the framework most often invoked in this sector's improvement rhetoric. Its empirical base is drawn from aircraft carriers, air traffic control, and nuclear operations — settings with career workforces, single employers, licensed operators, and continuous rather than campaign-based activity. A turnaround crew assembled from six employers for eleven weeks cannot develop collective mindfulness in the sense Weick and Sutcliffe describe,[12] and treating HRO as an available intervention at project sites imports an organisational precondition that is absent. This is a genuine limit on the review's own conceptual apparatus, and it is one reason the framework in Section 5 locates its most important requirements in engineering, contract, and verification rather than in culture.

7.2. Boundaries on These Claims

Three boundaries on these claims should be read alongside them. The coverage map in Section 4.5 rests on the author's coding of failure modes across nine inquiry reports and twenty-two evaluations; inquiry reports are written about the largest disasters and may over-represent organisational relative to technical causation, so the counts should be read as an attention profile rather than as precise frequencies, though the direction of the finding is robust to plausible re-coding. The synthesis is structured and scoping rather than meta-analytic — heterogeneity of setting, intervention definition, and outcome measure precludes pooling, and effect direction is reported as a vote count with its known limitation of weighting all studies equally. Two procedural limitations bear on how much weight the coverage map can carry: the review was not prospectively registered, so the pre-specification of its criteria and charting fields rests on the author's account rather than on a timestamped record, and screening was performed by one reviewer rather than two independently, with the mitigations and the residual risk set out in Section 2.4. Neither limitation is remediable retrospectively, and both are reasons to read the counts in Table 4 as an attention profile rather than as a measurement. And the framework of Section 5 is a synthesis-derived hypothesis: Section 6 exists because we do not claim it validated, and the engineered controls of Section 4.1.2 carry the reliability

 doi:10.20944/preprints202609.1107.v1

of specified mechanics rather than of evaluated field trials — a firmer warrant than behavioural evidence in most respects, but a different kind of warrant, and one that says nothing about whether a given site has actually installed them.

8. Conclusions

Safety intervention at oil and gas project sites is not deficient in volume, expenditure, or sincerity. It is misallocated. The interventions with the strongest published evidence — behaviour observation, safety-climate initiatives, training delivery — were developed and validated on repetitive, observable work performed by stable workforces, and they address the failure mode least frequently implicated in the catastrophic events that define the sector's risk. The interventions the inquiry record most consistently indicts — permit and isolation administration under schedule pressure, competence assurance across the contracting chain, disclosure by workers with the least employment security, and governance of the point at which commercial pressure meets barrier integrity — are supported almost entirely by normative consensus and by a near-empty evaluation literature.

Three conclusions follow. For operators and project managers: the assurance question to ask about a technical operation is not how many observations were completed but whether the isolation was verified by a party without a schedule incentive, whether the site knows how many safety-critical barriers are concurrently impaired, and whether the most junior subcontracted technician on the permit would be materially worse off for stopping the job. For regulators: the tiered process-safety indicator architecture already exists,[24,25] and the remaining gap is enforcement of its presence in the same governance forum where injury metrics are reviewed. For researchers: the field's most important question is also its least evaluated, the reason is a structural access problem rather than a conceptual one, and stepped-wedge designs across multi-site portfolios offer a route through it that does not require withholding protection from anyone. Until that evidence exists, the sector will continue to manage its catastrophic risk on the strength of expert consensus, which is a considerably weaker foundation than its injury statistics make it appear.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Funding: This research received no external funding.

Conflicts: of interest: The author declares no competing interests, financial or non-financial, with any operator, contractor, standards body, or certification scheme named in this review.

Ethics approval: Not applicable: this review synthesises published literature and public documents and involved no human participants, no animal subjects, and no personal data.

Consent: Not applicable.

Author Contributions: Sole author: conceptualisation, methodology, screening, charting, analysis, and writing.

Protocol and registration: Registered retrospectively on OSF (DOI: 10.17605/OSF.IO/Z8MN9); see Section 2.1 for the full disclosure and Section 7.2 for the resulting limitation.

Data: availability: All sources synthesised are cited in the reference list. The screening log (all title/abstract and full-text decisions with pre-specified exclusion reasons and second-pass disagreement flags), the charting spreadsheet underlying Tables 1, 4 and 6 and Figures 3–4, and the ordinal appraisal codebook are provided in full as supplementary material accompanying this preprint; no data are withheld and no request to the author is required to obtain them.

Reporting: Reported per PRISMA-ScR; completed checklist at Appendix B.

Use: of AI tools: No generative artificial intelligence was used to design the review, screen or chart sources, generate or interpret results, or draw conclusions. Generative AI was used only for language editing of author-written prose and for typesetting; the author reviewed and accepts responsibility for all content.

Permissions: No copyrighted figures, tables, or text are reproduced from any source; all tables and figures are original to this review.

Appendix A. Full Search Strategy

All searches were last executed on 18 June 2026. Date limits: 1 January 1990 to 18 June 2026. Language limit: English. Document types: journal articles, conference papers, and reviews; editorials, errata, and abstract-only records were excluded at the database stage where the platform supported it. Where a platform imposed a Boolean-length or wildcard limit, the string was split into the blocks shown and the result sets combined within the platform before export.

Table A1. Search strings as executed, by platform, with per-source yields reconciling to the 800 records entering de-duplication in Figure 1.

Database / platform	Fields searched	String as executed	Records
Scopus	TITLE-ABS-KEY	("oil and gas" OR petroleum OR offshore OR refinery OR "gas plant" OR drilling OR turnaround OR shutdown OR commissioning OR "pipeline construction") AND ("safety intervention" OR "barrier management" OR "permit to work" OR "lockout tagout" OR "lock out tag out" OR "energy isolation" OR "behaviour based safety" OR "behavior based safety" OR "safety observation" OR "stop work authority" OR "contractor safety management" OR "safety management system" OR "leading indicator*" OR "process safety indicator*" OR "safety climate" OR "competence assurance") AND PUBYEAR > 1989 AND (LIMIT-TO (LANGUAGE , "English"))	268
Web of Science Core Collection	TS (topic)	TS=("oil and gas" OR petroleum OR offshore OR refinery OR "gas plant" OR drilling OR turnaround OR shutdown OR commissioning) AND TS=("safety intervention*" OR "barrier management" OR "permit to work" OR "lockout tagout" OR "energy isolation" OR "behavio*r based safety" OR "safety observation*" OR "stop work authority" OR "contractor safety" OR "safety management system*" OR "leading indicator*" OR "process safety indicator*" OR "safety climate" OR "competence assurance") AND TS=("incident rate" OR "loss of containment" OR "process safety event*" OR "near miss*" OR fatalit* OR injur* OR barrier*) — timespan 1990–2026, English	194
ScienceDirect	Title, abstract, keywords	Block 1: ("oil and gas" OR petroleum OR offshore OR refinery) AND ("permit to work" OR "energy isolation" OR "barrier management" OR "lockout tagout"). Block 2: ("oil and gas" OR petroleum OR offshore OR refinery OR construction) AND ("behaviour-based safety" OR "safety observation" OR "safety climate" OR "stop	141

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 September 2026 doi:10.20944/preprints202609.1107.v1

		work"). Block 3: ("turnaround" OR "shutdown" OR "commissioning" OR "well intervention") AND ("safety management" OR "leading indicator" OR "contractor"). Years 1990–2026; article types: research articles, review articles.	
ASCE Library	Anywhere / abstract	("construction safety" OR "safety intervention" OR "job safety analysis" OR "pre-task planning" OR "hazard recognition" OR "safety climate" OR "subcontractor safety") AND ("oil and gas" OR petroleum OR refinery OR "industrial construction" OR commissioning OR "heavy lift*") — 1990–2026	63
OnePetro (SPE / IADC / OTC)	Full text with title/abstract weighting	("safety intervention" OR "barrier management" OR "well barrier" OR "permit to work" OR "energy isolation" OR "behaviour based safety" OR "stop work authority" OR "contractor safety" OR "process safety indicator" OR "competence assurance") AND (turnaround OR shutdown OR "well intervention" OR commissioning OR "simultaneous operations" OR SIMOPS) — 1990–2026	75
Database subtotal			**741**
Purposive normative and inquiry corpus	Issuing-body catalogues	Hand-retrieved from the publication catalogues of IOGP, API, ISO, IEC, Standards Norway, OSHA, UK HSE, Havtil, CCPS, NIOSH, and the U.S. CSB, plus the published inquiry reports for Piper Alpha, Texas City, and Macondo. Not database-indexed; retrieval is documented item-by-item in the screening log.	26
Reference-list screening (backward snowballing)	Reference lists of retained reviews	Reference lists of all seven retained reviews and meta-analyses, plus forward citation checks on the two retained randomised trials.	33
Total records identified			**800**

De-duplication. Records were exported to a reference manager and de-duplicated first automatically on DOI and on title-plus-year, then manually on author-plus-title for records without a DOI; 204 duplicates were removed, leaving 596 unique records for title/abstract screening.

Pre-specified charting fields. Intervention class (per the L1–L4 taxonomy in Table 2); failure mode targeted; setting in which the evidence was generated; outcome measure; design strength on the four-level ordinal scale defined in Section 2.6. For standards and inquiry documents: issuing body; normative status; failure modes asserted; controls required. No field was added after screening began.

Appendix B. PRISMA-ScR Checklist

A completed PRISMA-ScR checklist,[49] mapping each of the twenty-two items to the section and page of this review, is provided as supplementary material. Two items are reported as partially met and are declared as such on the checklist rather than marked compliant: item 5 (protocol and registration), because no protocol was registered — see Section 2.1; and items 8 and 9 (selection and charting process), because screening and charting were performed by one reviewer rather than by two independently — see Section 2.4. No other item is unmet.

References

1. Cullen, W.D. (1990). *The Public Inquiry into the Piper Alpha Disaster*. Department of Energy, HMSO, London.

2. Baker, J.A., III, et al. (2007). *The Report of the BP U.S. Refineries Independent Safety Review Panel*
3. (Baker Panel Report).
4. U.S. Chemical Safety and Hazard Investigation Board (2007). *Investigation Report: Refinery Explosion and Fire, BP Texas City*. Report No. 2005-04-I-TX.
5. National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (2011). *Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling*. Washington, DC.
6. U.S. Chemical Safety and Hazard Investigation Board (2016). *Investigation Report: Explosion and Fire at the Macondo Well*, Volumes 1–4. Report No. 2010-10-I-OS. Washington, DC.
7. Turner, B.A. (1978). *Man-Made Disasters*. Wykeham, London.
8. Reason, J. (1990). *Human Error*. Cambridge University Press.
9. Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Ashgate, Aldershot.
10. Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science*, 27(2–3), 183–213. doi:10.1016/S0925-7535(97)00052-0
11. Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 42(4), 237–
12. 270. doi:10.1016/S0925-7535(03)00047-X
13. Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management*. Ashgate, Farnham.
14. Weick, K.E., & Sutcliffe, K.M. (2007). *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*, 2nd ed. Jossey-Bass, San Francisco.
15. Hopkins, A. (2012). *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*. CCH Australia, Sydney.
16. Hopkins, A. (2009). Thinking about process safety indicators. *Safety Science*, 47(4), 460–465.
17. Sklet, S. (2006). Safety barriers: definition, classification, and performance. *Journal of Loss Prevention in the Process Industries*, 19(5), 494–506.
18. Duijm, N.J. (2009). Safety-barrier diagrams as a safety management tool. *Reliability Engineering & System Safety*, 94(2), 332–341.
19. Manuele, F.A. (2005). Risk assessment and hierarchies of control. *Professional Safety*, 50(5), 33–39.
20. Aven, T. (2016). Risk assessment and risk management: review of recent advances on their foundation.
21. *European Journal of Operational Research*, 253(1), 1–13.
22. Vinnem, J.E. (2014). *Offshore Risk Assessment: Principles, Modelling and Applications of QRA Studies*, 3rd ed. Springer, London.
23. Skogdalen, J.E., & Vinnem, J.E. (2012). Quantitative risk analysis of oil and gas drilling, using Deepwater Horizon as case study. *Reliability Engineering & System Safety*, 100, 58–66.
24. Norwegian Ocean Industry Authority (Havtil; formerly Petroleum Safety Authority Norway) (2024). *Trends in Risk Level in the Norwegian Petroleum Activity (RNNP) 2023 — Summary Report*. Havtil, Stavanger.
25. International Association of Oil & Gas Producers (2024). *Safety Performance Indicators — Annual Report*. IOGP Report 2023s. London.
26. International Association of Oil & Gas Producers (2018). *Life-Saving Rules*. IOGP Report 459. London.
27. UK Health and Safety Executive (2006). *Developing Process Safety Indicators: A Step-by-Step Guide for Chemical and Major Hazard Industries*. HSG254. HSE Books.
28. American Petroleum Institute (2021). *Process Safety Performance Indicators for the Refining and Petrochemical Industries*. API RP 754, 3rd ed.
29. International Association of Oil & Gas Producers (2017). *HSE Management — Guidelines for Working Together in a Contract Environment*. IOGP Report 423. London.
30. International Organization for Standardization (2018). *ISO 45001:2018 — Occupational Health and Safety Management Systems: Requirements with Guidance for Use*. Geneva.
31. Occupational Safety and Health Administration. *Process Safety Management of Highly Hazardous Chemicals*. 29 CFR 1910.119. U.S. Department of Labor.
32. Occupational Safety and Health Administration. *The Control of Hazardous Energy (Lockout/Tagout)*. 29 CFR 1910.147. U.S. Department of Labor.
33. UK Health and Safety Executive (2005). *Guidance on Permit-to-Work Systems: A Guide for the Petroleum, Chemical and Allied Industries*. HSG250. HSE Books.

 doi:10.20944/preprints202609.1107.v1

34. UK Health and Safety Executive & Institution of Engineering and Technology (2007). *Managing Competence for Safety-Related Systems, Part 1: Key Guidance*. HSE, Bootle.
35. Hale, A., & Borys, D. (2013). Working to rule, or working safely? Part 1: A state of the art review. *Safety Science*, 55, 207–221.
36. Hale, A., & Borys, D. (2013). Working to rule or working safely? Part 2: The management of safety rules and procedures. *Safety Science*, 55, 222–231.
37. Dekker, S. (2014). *The Field Guide to Understanding 'Human Error'*, 3rd ed. Ashgate, Farnham.
38. Krause, T.R., Seymour, K.J., & Sloat, K.C.M. (1999). Long-term evaluation of a behaviour-based method for improving safety performance: a meta-analysis of 73 interrupted time-series replications. *Safety Science*, 32(1), 1–18.
39. Sulzer-Azaroff, B., & Austin, J. (2000). Does BBS work? Behaviour-based safety and injury reduction: a survey of the evidence. *Professional Safety*, 45(7), 19–24.
40. Cooper, M.D. (2009). Behavioral safety interventions: a review of process design factors. *Professional Safety*, 54(2), 36–45.
41. Zohar, D. (2010). Thirty years of safety climate research: reflections and future directions. *Accident Analysis & Prevention*, 42(5), 1517–1522.
42. Nahrgang, J.D., Morgeson, F.P., & Hofmann, D.A. (2011). Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. *Journal of Applied Psychology*, 96(1), 71–94.
43. Christian, M.S., Bradley, J.C., Wallace, J.C., & Burke, M.J. (2009). Workplace safety: a meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, 94(5), 1103–1127.
44. Kines, P., Andersen, L.P.S., Spangenberg, S., Mikkelsen, K.L., Dyreborg, J., & Zohar, D. (2010). Improving construction site safety through leader-based verbal safety communication. *Journal of Safety Research*, 41(5), 399–406. doi:10.1016/j.jsr.2010.06.005
45. Albert, A., Hallowell, M.R., Kleiner, B., Chen, A., & Golparvar-Fard, M. (2014). Enhancing construction hazard recognition with high-fidelity augmented virtuality. *Journal of Construction Engineering and Management*, 140(7), 04014024.
46. Rozenfeld, O., Sacks, R., Rosenfeld, Y., & Baum, H. (2010). Construction job safety analysis. *Safety Science*, 48(4), 491–498.
47. Manu, P., Ankrah, N., Proverbs, D., & Suresh, S. (2013). Mitigating the health and safety influence of subcontracting in construction: the approach of main contractors. *International Journal of Project Management*, 31(7), 1017–1026.
48. Nenonen, S. (2011). Fatal workplace accidents in outsourced operations in the manufacturing industry.
49. *Safety Science*, 49(10), 1394–1403.
50. Hussey, M.A., & Hughes, J.P. (2007). Design and analysis of stepped wedge cluster randomized trials.
51. *Contemporary Clinical Trials*, 28(2), 182–191.
52. Hemming, K., Haines, T.P., Chilton, P.J., Girling, A.J., & Lilford, R.J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*, 350, h391. doi:10.1136/bmj.h391
53. Robson, L.S., Shannon, H.S., Goldenhar, L.M., & Hale, A.R. (2001). *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to Show Whether a Safety Intervention Really Works*. NIOSH Publication 2001-119, Cincinnati, OH.
54. Tricco, A.C., Lillie, E., Zarin, W., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. doi:10.7326/M18-0850
55. Kletz, T.A., & Amyotte, P. (2010). *Process Plants: A Handbook for Inherently Safer Design*, 2nd ed. CRC Press, Boca Raton.
56. International Electrotechnical Commission (2016). *IEC 61511 — Functional Safety: Safety Instrumented Systems for the Process Industry Sector*, 2nd ed. Geneva.
57. International Electrotechnical Commission (2010). *IEC 61508 — Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems*. Geneva.
58. Standards Norway (2021). *NORSOK D-010: Well Integrity in Drilling and Well Operations*, Rev. 5. Oslo.

 doi:10.20944/preprints202609.1107.v1

59. American Petroleum Institute (2017). *Recommended Practice for Analysis, Design, Installation, and Testing of Safety Systems for Offshore Production Facilities*. API RP 14C, 8th ed.
60. American Petroleum Institute (2018). *Well Control Equipment Systems for Drilling Wells*. API Standard 53, 5th ed.
61. American Petroleum Institute (2016). *Safe Hot Tapping Practices in the Petroleum and Petrochemical Industries*. API RP 2201, 6th ed.
62. Center for Chemical Process Safety (2012). *Guidelines for Engineering Design for Process Safety*, 2nd ed. AIChE/Wiley, Hoboken.
63. Mannan, S. (Ed.) (2012). *Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control*, 4th ed. Butterworth-Heinemann, Oxford.