Received: 11 August 2023 | Revised: 18 November 2023 | Accepted: 18 December 2023

DOI: 10.1002/prs.12579

PROCESS SAFETY PROGRESS

ORIGINAL ARTICLE

How audits fail according to accident investigations: A counterfactual logic analysis

Ben Hutchinson | **Sidney Dekker** | **Andrew Rae**

Safety Science Innovation Lab, Griffith University, Nathan, Queensland, Australia

Correspondence
Ben Hutchinson, Safety Science Innovation Lab, Griffith University, 170 Kessels Road, Nathan 4111, QLD, Australia.
Email: ben.hutchinson@griffithuni.edu.au

Abstract

Despite the reliance on safety auditing within organizations, comparatively limited research has studied the performance of safety auditing. When an investigation laments the "lack of audit quality" following an accident, what is meant by this statement? What contrasts a "good quality" audit from a "poor quality" audit? This study examined counterfactual logics (statements about alternative realities that did not occur but "could have" according to investigators) within 44 major accident reports to assess how audits are supposed to function and how they fall short of the ideal model. The content analysis yielded nine counterfactual auditing failures grouped into four categories. Contrary to the "ideal" model, audits (a) failed to facilitate an accurate understanding of threats by misinterpreting their saliency, (b) failed to facilitate timely action against threats by inadequately addressing the deterioration of known issues, (c) failed to facilitate effective management of issues, leading to confusion around the purpose and scope of audits, and d) failed to facilitate sufficient focus on threats by lacking focus on critical hazards and focusing on paperwork over operational issues or "failing silently" by missing threats while simultaneously praising performance. Practitioners should critically evaluate audits against these criteria and ensure audits effectively identify early warning signs.

KEYWORDS
hazards evaluation, incident investigations, risk assessment, training

1 | INTRODUCTION

Auditing is a widespread practice used, in part, for ensuring expected systems, practices, and controls are in place and functioning. Nevertheless, can auditing drive a false sense of safety in organizations? Auditing was implicated in the 1988 Piper Alpha oil platform accident, which tragically sank into the North Sea after suffering several explosions due to the unplanned release of condensate. Discussing the Piper Alpha accident, Appleton (technical advisor to the inquiry) noted that clearly "*there was no shortage of auditing [at] Piper. What was deficient was the quality of that auditing. Not only were departures from laid-down procedures not picked up, but the absence of critical comment in audit reports lulled senior management into believing that all was well.*"[1(p203)]

Appleton's statement prompts several questions. For one, what do major accident investigators expect from effective audits? Several guidelines cover the implementation of auditing systems (see for instance Reference 2), but there appears to be scarce evidence exploring how investigators construct the successes and failures of auditing in the context of major accidents. Related to the first question: what role have audits played in failing to adequately warn organizations of emerging danger before a major accident? Evidence to this question is also relatively scarce.

Third, what role does counterfactual reasoning play in the construction of investigation reports? Investigation reports make frequent reference to counterfactuals – hypothetical alternate realities that did not occur but may have occurred had some factors been otherwise.[3] Investigations use counterfactual logics to compare what happened in the event against a normative frame, but counterfactuals have come under some debate as scientific evidence in the safety science community.[4–6]

When the counterfactual logic is applied to audits, accident reports allow a view into the idealized model of how accident investigators believe audits should be designed or be capable of achieving. With this knowledge, practitioners can question whether their audits are performing against expectations and whether audits relay (or not) early warning signs. Hence, the three aforementioned questions frame the objectives, and subsequent findings, in this paper: (1) what major accident investigators expected from effective audits have been extracted from accident reports, (2) the contributory role audits played in the accident were categorized based on direct inference from the reports, and (3) counterfactual statements made by the investigators were grouped via thematic coding.

2 | BACKGROUND

2.1 | What is auditing?

Auditing can be defined as a "systematic examination against defined criteria to determine whether activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve the organization's policy and objectives."[7]

Audits may broadly be classified as internal or external.[8] External audits typically involve parties external to the organization, such as auditing firms, clients, or regulators. Internal audits are run by the organization, evaluating how its practices and systems are performing.

2.2 | Do audits have a role in sensitizing or desensitizing organizations to danger?

Audits may play a role as a system performance indicator.[9–11] Resultingly, audits should be able to identify and effectively communicate precursor issues for resolution within organizations and/or amplify weak signals[12]; ultimately creating "safety foresight."[12] The identification and interpretation of signals "rarely arrive pre-packaged in the clear and unambiguous form" but instead must be "actively made and constructed, by relating information on organizational performance to pre-existing concerns."[13(p4)]

Disasters are "essentially organized events"[13(p2)] that require a long incubation period of discrepant gaps in organizational, social, and psychological patterns, as well as prolonged neglect or discounting of potential signs of danger.[13,14] These two elements allow accident precursors to accumulate unnoticed. Theoretically, auditing can help organizations detect accident precursors, but auditing can also provide a false assurance of safety, inadvertently allaying subjective concerns for issues where alleviating concern was not objectively warranted.[15] While ostensibly safety audits can sensitize—or desensitize—organizations to signals of danger, little empirical evidence exists.

2.3 | How do investigation approaches influence the construction of accident reports?

Broadly speaking, an incident investigation report is constructed following several steps after an incident[16(pp3–4)]:

- Assembling the appropriate investigation team.
- Gathering information, like stakeholder interviews and statements or physical evidence.
- Analyzing the data to formulate an accurate understanding of what happened, for example, constructing plausible scenarios to be evaluated.
- Determining causal factors and effective improvements and recommendations.
- Preparing the investigation report that details the facts, findings, and recommendations.
- Communicating findings and conclusions from the investigation.

In safety science, the nature of creating meaning via what exists or is perceived to exist (ontology), and the relationship between the mind and knowledge (epistemology) can be seen along a spectrum. One side of the spectrum is the positivistic perspective. Positivism is said to encompass the "traditional" approach[17] and reasons that an objective reality exists that is separate from the investigator and can be derived via the application of quantitative and scientific methods.[18] That is, facts are simply "out there" waiting to be found by a studious investigator,[19] where "context-free generalizations" and cause–effect relationships ensure that the investigation "converge on the "true" state of affairs."[18(p109)]

The positivist perspective is dominant in construction safety,[20] and reflects the epistemology of engineering theory.[21] This perspective is evident in accident investigation guides, which describe investigations as a "search for the truth,"[22(p3)] or the search for facts, root causes, and causative factors,[16,22,23] or the determination of what is true or false based on hypotheses and analysis techniques.[16] Guides also suggest that physical evidence is a type of "real evidence" that "always tells the same story" and that the investigator's input is less important than the physical evidence, which must be allowed to "present [its] own conclusions."[22(p3)] In sum, from a positivistic perspective, a clear boundary can be drawn around "real" evidence, as it exists in the world with its unbiased story.

Constructivist perspectives sit at the other end of the spectrum. While constructivists agree that knowledge can be constructed through formal methods, logic, and tests,[4] the interpretation and judgment of data still contain an "irreducibly social component."[4(p739)] Moreover, data in the world have inherent uncertainties and ambiguities,[4] which beget

different interpretations of the information's meaning. Applied to investigations, this suggests that facts or causes are not so much found as they are constructed by the analysts[24] and influenced by a range of factors like their existing worldviews.[25] Socio-political factors also influence the information that is included in the investigation, what gets left out, or how information is pieced together, interpreted, or simplified.[24(pp76–77)] Finally, from a critical angle, constructivist perspectives have been argued to capture properties of social systems (of which organizations are made up of)[26] that cannot necessarily be observed via the use of positivist methods related to physical phenomena.[27]

In all, objects of reality do exist. Likewise, investigators invariably construct investigations based on their existing worldviews, assumptions, and preferences and are influenced by environmental factors and constraints. The next section explores how "finding" or "constructing" causes play out in investigations.

2.4 | What is counterfactual reasoning in the context of investigations?

Accident analysis, from both positivist and constructivist perspectives, involves selection and prioritization. The construction of a preferred account of an accident necessarily silences and underemphasizes other competing possibilities.[28,29] This process can be informed by an idealized model of what could or should have happened. In other words, investigations construct, explicitly or implicitly, what did or did not happen using a process of counterfactual reasoning against an idealized model.

Counterfactual reasoning involves mental representations of alternatives to past or future events, actions, or states.[30] Byrne[30(p148)] suggests that counterfactuals typically involve an if-then conditional format, where "if" refers to a person, action, or circumstance change and "then" refers to a better or worse alternative outcome. He et al.[31(p2)] provides examples of how counterfactuals may be broad and general, such as "if only I was more careful," or detailed and specific, such as "if only I had double-checked my safety monitor." These counterfactual thoughts can then modify specific behaviors, traits, or other situational features.[31] Many types of counterfactual thoughts and emotions have been described.[32] Two that have been applied to workplace safety include upward (better alternatives than the outcome) and downward counterfactuals (worse alternatives than the outcome).[31] In sum, it is common practice in industry to explain an event by constructing multiple plausible scenarios and evaluating them against the data.[33]

Safety investigators can fall into the trap of post hoc rationalization—for example, hindsight bias. This is where investigators—with clear knowledge of what happened—can identify points where people could have revised their judgments or actions but failed to do so.[34] Examples of counterfactual statements reflective of post hoc rationalization—for example, would have, should have, could have—are found throughout accident investigations. Examples include the following:

- According to the Nimrod aircraft accident report,[35(p84)] a "cursory examination" of existing improbable likelihood calculations on fuel leak rates "*would have* shown this statement to be unsound" and discussions with Nimrod crews "*would have* shown the claim to be manifestly untrue" (emphasis added).
- A HAZOP study for some pumps was omitted before the Longford plant accident. The investigation suggests that a HAZOP "*would have*" systematically described and questioned deviations from design intents and the consequences that could "conceivably occur" and identify procedures to avoid dangerous low-temperature states. As such, "it is inconceivable that a HAZOP study of GP1 *would not* have revealed factors which contributed to the accident."[36(p204)]
- Investigations of the 2005 BP Texas City explosion cite examples of how: 1) a standard "*should have*" been reviewed during Process Hazard Analysis revalidations and facility siting studies and if it had, a critical hazard "*would have* been evaluated" (emphasis added).[37(p113)]
- Review of the entire health and safety management system before the 2010 Pike River mine disaster "*would have* identified anomalies, many of which could have been readily rectified,"[38(p77)] such as inconsistencies around trigger values for flammable gas accumulation or references to inapplicable codes of practice (emphasis added).

When investigations use language like the above, they are engaging in counterfactual reasoning and making claims about a hypothetical alternate version of events. Yet, the status of such reasoning as scientific evidence is a matter of debate in the safety science community.

Some authors have a different perspective on the value of counterfactual reasoning. Some are critical and dismissive, while others are cautious but see value. Other authors adopt the reasoning as factual with seemingly little hesitation.

In the dismissive camp, Dekker et al.[39] criticize investigations in complex systems as a "construction" of causes rather than objective and unambiguous signs to be found. In their view, investigations reconstruct events through a linear process of parsing complex and interrelated factors into a story broken down into individual components, parts, and actors.[39] Selecting fragments of data can be a process of cherry-picking and story-crafting, whereby similar pieces are lumped together into a coherent narrative.[24] The result is never truly "objective,"[39] since the construction of an event is influenced by the investigator's "background, preferences, experiences, biases, beliefs, and purposes."[39(p5)]

In support, a 2004 meta-analysis of 95 studies found low to moderate effect sizes of hindsight bias on subsequent judgments across a range of settings[40] and a detailed analysis of over 2000 traffic accidents by a multidisciplinary investigation team flagged the role that post hoc rationalizations had on judgments.[41] A later study found that *a priori* "if only" conjectures about the accident can subsequently lock investigators into a narrowed focus that is aimed at correcting people rather than system change.[33]

Further supporting this perspective, investigations are also influenced by a range of other factors including the language used in investigations, where more agentive language can increase the causal attribution of fault toward people,[42] particulars of safety climate and

safety communication within the organization,[43] assumptions inherent within the accident models,[25] and a host of other cognitive, social, and political influences in the organization.[44]

In the middle camp are scholars such as Hopkins, who utilize diverse sources including accident reports, to construct narratives that are informative, perhaps speculative, but logical.[45–47] Hopkins' storytelling employs a repeated narrative sequence, encompassing factual data, organizational assumptions, parallels with other incidents or studies, explanation of phenomena, and recommendations for improvement. This approach incorporates counterfactual arguments, aiming to convince readers of potential accident prevention under different circumstances.[48] This narrative style is said to invite readers on "a journey from a series of problems to a series of solutions while providing novel insights into a set of issues faced daily by [industry] practitioners."[48(p741)] Hopkins contends that investigations fundamentally address causes, either implicitly or explicitly.[3] One meaning of causality involves a "but-for" counterfactual logic: indicating a factor that was necessary for the accident to occur and had it been otherwise, would not have occurred.[3] Instances of "but-for" attributions exhibit logical underpinning in certain scenarios while relying more on expert discernment in others. Logical deductions are better suited for technical matters, whereas nuanced expert judgment is necessary for more remote and organizational factors[3]—reflecting counterfactuals that are more "persuasive rather than conclusive."[3(p6)]

Finally, the third camp takes a more literal approach in interpreting and applying findings from investigations without explicitly recognizing the empirical status of counterfactual arguments. Unlike the other positions, this approach is not explicitly explained and defended in the academic literature. It tends to be the default approach, critiqued by advocates of the alternatives. A combination of perspectives from different camps may also be in effect.

This paper does not aim to criticize the counterfactuals or pass judgment on their veracity. It also does not use counterfactuals as evidence of what should have happened. Instead, counterfactual arguments are used to establish what the idealized model of safety looks like from one cohort of practitioners. Understanding the assumptions and beliefs investigators hold in the causality of accidents is valuable for identifying levers of system improvement (e.g., as highlighted by References 25,44) and closing the gap between assumed safety versus reality.

When we apply this type of analysis to audits, accident reports allow us to see the "counterfactual audit" – the idealized picture of what accident investigators believe audits should achieve in practice. This picture is rarely explained directly but is built up of all of the comparisons when actual audits are criticized for their failure to match the ideal.

The counterfactual audit is an important artifact because it then lets us compare actual audits with this idealized picture. Once we know how audits are expected to function for the prevention of accidents, we can properly ask if, when, and how audits meet these expectations. This provides a tool for both critically examining current audit practices, and, if feasible, developing improved audit practices.

3 | METHODOLOGY

This study examined counterfactual statements of safety audit performance and failure according to major accident reports. Four hundred and twenty-eight reports were identified after searching various internet sources, including the US Chemical Safety Bureau (CSB), UK Health and Safety Executive, regional or national safety and mining regulators, and websites hosting official accident reports and/or royal commissions/investigation inquiries. Most reports came from the CSB due to their large source of publicly available reports.

The inclusion criteria for the accident reports were as follows:

1. English reports from Australia, Canada, New Zealand, The United Kingdom, and The United States of America.
2. Published between 1990 and 2023 (to maintain a contemporary focus).
3. Non-transport onshore and offshore accidents investigating oil & gas, petrochemical, mining, and pipeline sectors (to maintain a manageable sample of major accidents).
4. Final/completed accident reports publicly available on the internet.
5. Mentioned the term "audit" or variations thereof.
6. Of the reports that mentioned "audit", mentioned a type of audit failure or gap in auditing.

Forty four accident reports corresponding to 42 accidents met the inclusion criteria. In two cases, more than one investigation report was available for the same accident (the Macondo well blowout and Texas City accidents). In a single case, a book exploring a major accident by a prominent safety researcher was included due to the unique insights this book offered. Figure 1 shows the breakdown of investigation report sources according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA).[49] In short, an initial sample of 14,576 reports matching the specified search terms were identified. This sample was immediately culled down to 5612 reports for screening. Over 5000 records were removed for not relating to a full safety investigation published in English. Of the remaining sample, 428 reports remained for thorough assessment. Finally, 384 of the 428 reports were excluded for not mentioning an auditing failure in the investigation, leaving a final sample of 44 reports.

All statements relating to auditing within the reports were copied into a Microsoft Excel table. Reports were excluded if they did not meet the above criteria. The most common reasons for excluding reports that mentioned "audit" were either that (a) the statements were vague or simply discussed the importance of auditing or (b) referred to a range of field-based inspections as audits and did not meet the definition for this study (e.g., this was most frequently encountered in mining reports, which more frequently referred to inspections of conveyor belts and other mechanical plant as "audits"). As per guidance from PRISMA,[49] the excluded audit reports were re-evaluated using different keywords ("inspect," "evaluate," "review," "probe," and "investigate") to ensure that different terminology for "audit" were not being used. As per the above item b), none of the excluded studies met inclusion following secondary review.

After a full list of auditing statements was listed in the spreadsheet, a qualitative content analysis approach was used to group the data.

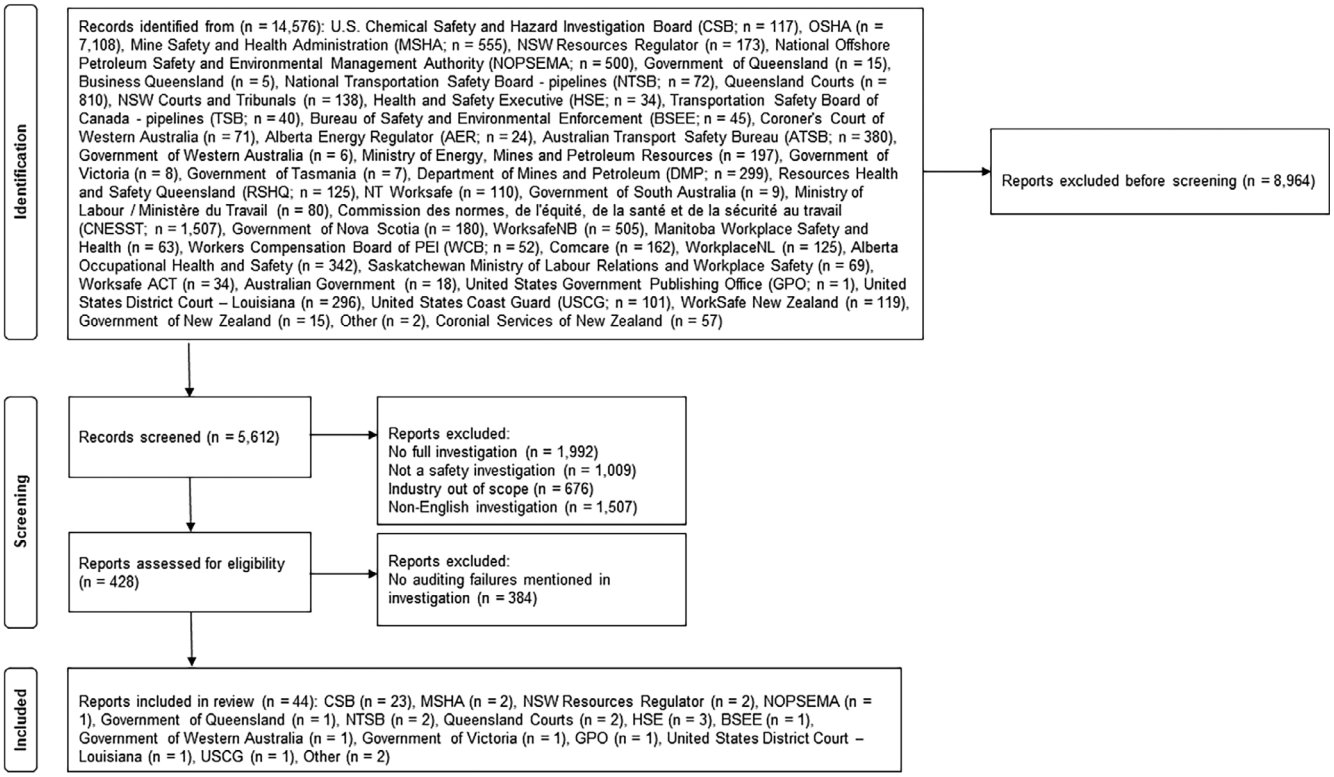

FIGURE 1 Breakdown of investigation report sources based on PRISMA.

Qualitative content analysis involves a process of abstraction[50]—assigning sentences and paragraphs to similar codes, categories, and themes. Specifically, groups of text were assigned under related abstract subcategories, such as at the first pass "Audits focused upon review of processes and documents rather than on verification and assessing whether procedures are being followed or whether work practices were consistent with procedures," then abstracted to a higher level with "Audits focused on "surface compliance"; reflecting stated failure types of audits. Finally, categories were assigned to a counterfactual statement based on the reverse of the audit statement, like "Expectation that audits focus on critical components." All sub-categories were allocated to four overarching failure categories.

By the end of the review, each category had reached saturation,[51] with no new categories being formed within approximately the final 40% of reports. Each audit statement was then rewritten into a briefer description to tabulate all of the findings concisely. Categories, where audits did effectively identify or relay major hazards, were omitted as out of scope.

In many cases, statements to audit failings or involvement in accidents were vague or could be considered to cover more than one category. In these instances, judgment was used to group the factor into the category that best fits the factor based on its context.

4 | RESULTS

Qualitative coding resulted in nine categories of counterfactual reasoning on what auditing is expected to achieve and with associated failure types, shown in Table 1 in summarized format. Table S1 contains the full categorized list of statements.

Note that the frequency of observations per category is not necessarily indicative of how prevalent these factors are in reality, nor their importance according to investigations. Rather, the number of observations reflects the definition of the categories (a structural condition related to the qualitative coding) and also the operating mental models of the accident investigators (what-you-look-for-is-what-you-find).[25] Finally, investigators' independence, competence (including technical and non-technical skills competencies),[52] scopes, and employed investigation methodology may shape the findings of investigation reports.[53]

5 | AUDIT FAILURE TYPES

To understand what investigators expect from audits, this paper evaluated 44 accident reports to (1) infer the idealized model of auditing based on the perspective of investigators, (2) determine how audits departed from this idealized model, and (3) determine how audits can normalize or sensitize organizations to major risks.

The nine sub-categories are contained within four overarching categories:

- Audits involve a failure to understand.
- Audits involve a failure to act.
- Audits involve a failure to manage.
- Audits involve a failure to focus.

TABLE 1 Summary of counterfactual auditing failures grouped by overarching categories.

Failure categories as coded from accident reports	Reconceptualized counterfactual categories	Examples of departures from counterfactual model
1. Failing to understand		
Audits lacked personnel with the necessary skills, knowledge, or expertise	The expectation that audits will involve personnel with appropriate skills, knowledge, or expertise	Internal auditors were said to have not received the training in auditing required to effectively discharge their role
Audit identified a hazard or issue but its relevance or severity was misinterpreted	The expectation that audits will lead to the accurate and salient interpretation of hazards	Issues identified following audits that, in hindsight, should have been clear threats for resolution were understood or interpreted in a way other than as a direct safety threat (e.g., safety threats were interpreted as insurance threats)
2. Failing to act		
Some of the issues relayed by audits were not properly acted on or managed	The expectation that issues identified during audits will be appropriately acted on or managed	For various reasons, issues that were identified during internal or external audits were either not taken as priority issues for resolution, the resolution actions were inadequate or ineffective, or the issues were initially resolved but re-deteriorated over time without further resolution
Inadequate action on improvement opportunities when in hindsight a different response should have been taken	The expectation that auditing will result in clear improvement actions	Issues identified during audits did not result in unambiguous and assigned actions for resolution but rather ambiguous opportunities for improvement, mere observations or ambiguity around how and why some issues were actioned, and others were not
3. Failing to manage		
Misuse or lack of clarity on the purpose and scope of audits	The expectation is that audits will clearly articulate the purpose and scope of the audit	Some degree of uncertainty or lack of clarity existed around what an auditing regime was supposed to entail or achieve, how it would achieve its goals, who should be involved or lead an audit, or uncertainty about what parts of the safety system should be evaluated as part of auditing
General failure to establish an effective auditing program	Expectations that effective auditing programs will be established	A comprehensive or sufficient scheduling of audits and/or guided by appropriate standards was not in place or functioning consistently or effectively
4. Failing to focus		
Audits lacked focus on critical components	Expectations that audits focus on critical components	Audits either failed to consider or explore certain major hazards or lacked the tools or methodologies to appropriately canvass such critical issues
Audits focused on "surface compliance"	The expectation is that audits penetrate through paperwork to measure the effectiveness	Audits verified that certain expected systems, documents, or processes existed but did not adequately evaluate how effectively those deliverables were functioning or whether they had the intended impact on hazards or practices; thus, audits focused on "surface compliance"
Audits failed to identify deficiencies but reinforced a positive view: failing silently	The expectation that audits will identify deficiencies and alert of potential threats to safety – failing loudly	Audits not only missed critical deficiencies but were reported to have praised performance despite issues or reinforced a belief that issues did not exist or were under control, such that audits "failed silently"

5.1 | How do audits fail compared with the idealized model?

5.1.1 | Audits involve a failure to understand

Audits lacked personnel with the necessary skills, knowledge, or expertise

According to investigators, audits are expected to include personnel with appropriate skills, knowledge, or expertise. As an example, the Husky Superior Refinery[54] report referred to an internal audit that did not have the support of a suitable subject matter expert. Without the input of relevant subject matter experts, investigators believe that a critical gap in shutdown procedures was not discovered during audits before the accident. Also, a 2016 process safety audit at the Husky Superior Refinery limited interviews to contractors and emergency responders, omitting refinery operators.

In the case of Piper Alpha,[55] the royal inquiry noted that inspectors did not necessarily have the expertise to thoroughly audit specialized systems, like the permit-to-work system.

Audits identified a hazard or issue, but its relevance or severity was misinterpreted

After identifying an issue, audits are expected to facilitate an accurate understanding of issues and their associated severities. Examples were found where the findings derived during audits were misinterpreted, thereby allowing inadequately managed risks to remain in place.

In the case of the Imperial Sugar dust explosion,[56] a "Good Manufacturing Practice" audit by external parties noted the accumulation of sugar dust in the facility – however, due to the audit's focus, the dust was marked as an issue of food quality and not also as a combustible hazard (with investigators remarking that this type of audit could be suitable for identifying safety hazards). Similarly, an insurance audit before the West Pharmaceutical Services dust explosion also identified dust residue on fire sprinkler heads at the facility but did not assess whether that dust was a combustible hazard.[57]

Other audits evaluated the dust or chemical systems that were in place at the respective facilities. In the case of the toxic chemical release incident at the DuPont La Porte Chemical facility,[58] an external audit identified the presence of a hazardous chemical detection system but did not verify if the system was effective. In contrast, dust control measures and fire protection systems were evaluated during external audits before the AL Solutions metal dust explosion,[59] but mistakenly considered inadequate systems to be acceptable.

5.1.2 | Audits involve a failure to act

Some of the issues relayed by audits were not properly acted on or managed

Audits were expected to effectively ensure that issues identified during the process would be acted on and managed. Multiple examples were found where this idealized statement was not achieved.

For instance, an earlier audit before the Prudhoe Bay oil spill was said to have reaffirmed concerns about ongoing issues of pipeline corrosion, but this finding did not result in a revamp of pipeline control systems.[60] While some audit actions had been closed relating to the blowout preventer at Macondo, some of these "resolved" issues were found to have re-deteriorated.[61]

Inadequate action on improvement opportunities when in hindsight a different response should have been taken

Audits were expected to result in clear and actionable responses to identified issues. Correspondingly, at the El DuPont de Nemours & Co facility, a hose in the phosgene feed system was not aligned with DuPont's recommended standards.[62] The non-compliant hose type remained unresolved, as the audit designated the issue as an "observation" rather than mandating remedial action. The report lacks clarification on why the audit team chose an "observation" (a non-mandatory but preferred response) instead of requiring mandatory action against the company's internal requirements manual (which delineates essential facility mandates). Once categorized as an observation, the manual omitted the necessity for an action plan to rectify the issues.

Elsewhere, audits identified issues, but no recommendations were issued to resolve gaps in work instructions or departures from procedures in the case of the Husky Superior Refinery.[54] Improvements were likewise identified to resolve defects highlighted during audits at the Kuraray Pasadena facility,[63] but no documented information was available to demonstrate why 42% of the improvements were rejected for implementation. In a similar vein, some audit findings before the Macondo blowout were noted to have been outright rejected without demonstration of a formal risk mitigation assessment.[61]

5.1.3 | Audits involve a failure to manage

Misuse or lack of clarity on the purpose and scope of audits

Audits were envisioned to possess clear scopes, purposes, and functions. Instances of misuse surfaced. At Texas City, an overemphasis on audits as the primary driver of continuous improvement impeded the development of a management system that prioritized continuous risk reduction.[64] Similarly, the Zinifex Century mine fatality inquiry unveiled a misguided use, where executives perceived a "complex interlocking system of audits, reports, and rules" as fostering a safety-sensitive culture, despite their partial blindness to hazardous work.[65(p12)] As a different example, Transocean's CEO, responsible for the Deepwater Horizon drilling rig involved in the Macondo disaster, reportedly misconstrued the essence of a safety system management review. The CEO erroneously regarded audits as an "activity carried out by Transocean's consultant, Lloyd's Register", rather than acknowledging it as an "activity he [as the CEO] should [be] leading."[66(p57)]

For lack of clarity examples, there was confusion before the Macondo accident around what the scope of the safety management system was, and thus, confusion surrounding what the auditor believed to be in-scope for the audit.[66] In the Montara oil spill,[67] some uncertainty existed around the difference between performance monitoring and formalized audits and management reviews such that these different activities were conflated as the same thing. Audits were also expected to have sufficient independence from internal and external influences that could influence the results of auditors. In the case of the Moura mine accident,[45] insufficient independence existed, resulting in a lack of candidness and critical assessment from the auditors.

General failure to establish an effective auditing program

Investigators expected that effective auditing programs would be in place to verify the management of systems, processes, and risks.

Comprehensive auditing programs were not implemented in several instances, including before the chemical explosion at the Synthron facility[68] or in the lead-up to the Snapper Mineral Sands mine fatality.[69] Synthron's parent company, Protex, was said to have provided little safety oversight or support to Synthon at their facility—like from a comprehensive auditing program—despite the much larger size of Protex and its available resources. At the Snapper Mineral Sands mine, no safety management system audit was completed in the preceding two-year period, despite being a requirement.

Other instances highlighted specific facets of audit programs that were lacking, like an evaluation of alkylation operations at the CITGO Corpus Christi facility,[70] contractor selection, and the safety oversight program at Xcel energy,[71] or traffic policies and procedures in the Columbia Quarry fatal accident.[72]

Another report directed some culpability toward the respective environmental regulator for not auditing the McKee refinery before the 2007 incident.[73]

5.1.4 | Audits involve a failure to focus

Audits lacked focus on critical components

The largest category of audit failures related to audits, in hindsight, that did not focus on the necessary and critical elements of safety systems, risks, or practices.[1] This suggests that the "ideal" audit is expected to direct acute attention toward systems governing major hazards and/or operational practices interfacing with major hazards.

Different themes were evident in the data. In many cases, specific hazards or hazardous activities did not receive due attention during audits, for example, lockout/tagout procedures involved with the Tosco refinery fire[74] or the cutting of piping before the Tesoro Martinez acid spill[75]; in the latter, audits could not surface gaps between corporate standards and pipe cutting practices. Investigators in the Formosa Plastics explosion identified significant risks embedded within the design and procedures related to PVC resin processing, seemingly missed by audits.[76]

Elsewhere, investigators described weaknesses in audits regarding rule compliance. In the Chevron refinery fire,[77] audits focused too strictly on regulatory compliance—that is, ensuring regulations or standards were adhered to—while omitting the search for other ways to reduce risk over and above regulations or standards. In the Macondo accident, investigators believed that audits were not even compliant with international standards, nor effective for their purposes.[66]

Finally, other notable findings in this category included:

- Auditing at the Moura mine had too great a focus on health and safety matters over catastrophic safety hazards, like explosions.[45]
- Auditors were not provided with suitable auditing tools to assess critical systems, like the permit-to-work system at Piper Alpha.[55]
- Previous incidents relating to phosgene did not feed back into prospective audits, thereby allowing similar issues to persist such as at El DuPont de Nemours & Co.[62]

Audits focused on "surface compliance"

The second most prominent category involved audits appraising safety issues at a superficial level, reflecting two related themes. The first theme pertained to audits overly fixating on documentation. For instance, audits before the Buncefield depot fire[78] and the DPC Enterprises chlorine release[79] faced criticism for prioritizing documented artifacts and the presence of systems over their operational functioning. Similarly, organizations behind the San Bruno pipeline rupture[80] and Airgas Nitrous Oxide explosion[81] used audit protocols that lacked checks for data completeness or accuracy.

The second theme involves inadequate attention to operational practices or states. Preceding the BP Texas City explosion,[82] audits were criticized for insufficiently addressing actual worker practices and their alignment with procedures. At the Tosco facility, while an audit did address practices, it disproportionately focused on worker behavior rather than the efficacy of procedure implementation.[74] Notably, an instance emerged where auditors accepted the proper functioning of a vital gas alarm system at face value, omitting identification of any departures from actual practice.[45]

Audits failed to identify deficiencies but reinforced a positive view: failing silently

Based on the idealized model, audits are expected to "fail loudly" – providing salient cues that the audit itself is not providing an accurate understanding of operational risks and system effectiveness.

The Longford gas plant explosion[36] highlights the phenomenon where audits failed silently, not loudly. A prior audit was observed to have assessed many elements of the plant's safety management system at the highest possible level, thereby failing to convey bad news. Similarly in the E.I. DuPont De Nemours Co. explosion,[62] prior audits consistently praised the high quality of safety systems while missing critical deficiencies.

The connection to failing silently was less direct in the case of Piper Alpha.[55] The investigation report noted that an unintended by-product of the inadequate auditing approach was the resulting perception that the lack of bad news indicated a successful management approach.

6 | DISCUSSION

This paper evaluated 44 investigation reports, finding several insights about the idealized model of safety auditing according to accident investigators – and how audits failed to meet these ideals. Though primarily derived from major accident reports, the findings are likely relevant to a wider scope of health and safety audits.

6.1 | Audits direct focus on documents and surface issues

The most numerous findings in this sample related to audits failing to (a) direct auditors and auditees toward critical elements of operational practices or safety systems related to major hazards or (b) failing to enable the identification of critical deficiencies, which in hindsight were of importance in the accident's genesis.

The second most prevalent failure type concerns surface compliance,[83] implying that an "ideal" audit should delve beyond the

presence of documentation and system artifacts to verify practical system functioning. For example, audits in the Buncefield[78] and Texas City[82] incidents were criticized for assessing documentation and systems without sufficient validation of their actual implementation or efficacy. Further critiques of audits pertain to their limited assessment of on-site practices.

The current findings support those of Hutchinson et al.,[84] who evaluated 71 safety audit reports within a large engineering contractor in Australia. Audit reports in their study were found to

- overemphasize largely trivial and readily observable physical hazards or paperwork – surface compliance[83];
- treat documents as if they were the issue themselves, for example, a process flowchart being posted on a noticeboard is taken as evidence that the process is used;
- take an overly literal approach, resulting in questions being answered exactly as they are written but not probing why any deficiencies existed;
- place too much stock in the value of signatures, confusing signatures as evidence that something has been read and understood; and
- drive the decoupling of risk, acting as a type of symbolic masquerade activity where serious issues do not get addressed despite audits appearing to be effective.

Comparing the current findings with those of Hutchinson et al.,[84] it is evident that audits can over-prioritize the collection of documents and the presence of artifacts at the expense of probing system functionality. Said differently, a gap exists between how systems are *expected* to function versus how systems *do* function. That gap is argued to drive decoupling, further widening expectation versus reality while providing a false veneer of safety. Based on the alignment between the current paper and prior work on health and safety auditing,[84,85] it is likely that these findings apply to broader health and safety audits and not just process safety audits.

Moura's dataset[45] also underscores an established organizational phenomenon – the distinction between work-as-imagined (WAI) and work-as-done (WAD).[86] WAI describes work based on assumed performance norms, often under ideal and unhindered constraints like training, equipment availability, or time pressures. Conversely, WAD reflects actual work within real constraints, incorporating trade-offs, resource limitations, workarounds, and stressors.[24] At Moura, the mining assistant manager-in-charge had not seen the mining procedures and dismissed the significance of reviewing the procedures based on an assumption that the procedures would have been aligned with ongoing practices.[45] However, the procedures merely "reflected what the writer thought was happening,"[45(p76)] exposing gaps between prescribed procedures (WAI) and actual practice (WAD). Echoing Clarke,[87] Hayes and Hopkins,[88] Hutchinson et al.,[84,89] and the Longford Royal Commission,[36] audits, under certain conditions, conflate document management with issue management, despite their distinct natures.

6.2 | How do audits conceal or normalize major risks?

6.2.1 | Audits reinforce a positive view of safety by failing silently

Three instances in this dataset exemplify how audit programs can convey a positive safety message when it is not justified. Essentially, these auditing programs were failing silently: failing to attain their objectives while providing minimal indication of their ineffectiveness. This suggests that an "ideal" audit should not only pinpoint crucial shortcomings and alert organizations to safety risks but also distinctly signal if the audit program is underperforming – "failing loudly". Cullen[55] noted scant critical references to the permit-to-work system in Piper Alpha platform audits, leading to the reinforcement of an "all is well" assumption due to the absence of adverse feedback on the system's evaluation. Similarly, high praise for system compliance was consistently raised during audits preceding the E. I. DuPont De Nemours Co. explosion.[62]

At Longford,[36] an audit provided high praise for the plant's safety management system, even as the post-accident report criticized the system as needlessly intricate, repetitive, and circular. Additionally, the safety system's language was deemed "impenetrable," and crucially, its management took on a "life of its own, divorced from operations in the field."[36(p200)] A similar conclusion emerged from the San Bruno pipeline explosion.[88] Here, the focus shifted to system management rather than risk management, consequently rendering the system ineffective in disseminating "bad news."

Furthermore, for an audit to fail loudly, an audit is needed in the first place. Instances were identified where audit programs were absent or incomplete, omitting scrutiny of critical facets of safety systems and operational procedures. This failure-via-omission is indicative of "fantasy plans" – unrealistic assumptions or claims in plans that remain unfulfilled.[87–90] Hayes and Hopkins[88] remarked that the asset integrity management system preceding the San Bruno accident was more symbolic than functional. This misplaced symbolism in the effectiveness of the asset management system masked the defective reality that actual operational risks were not adequately addressed. Likewise, safety plans promising an effective audit program's existence, when such a program is absent, can foster false assurance that critical hazards are identified and controlled: decoupled from practice.

Taken together, audits did not relay serious concerns that were present in the management of critical systems or risks such as would be expected from the effective governance of weak signals and risk indicators. Rather, audits inadvertently allayed concerns, a type of probative blindness where subjective confidence in safety was disconnected from objective risk.[15] At a pathological level, audits may be "inadequate to provide early warning of process safety risk".[82(p139)]

6.2.2 | Audits inadvertently reconceptualize hazardous conditions into less concerning conditions

Another area where audits inadvertently silenced concerns (that in hindsight should have sensitized people to potential danger) was

when audits facilitated an ill-calibrated view of danger. This suggests that the "ideal" audit is expected to identify and interpret hazards and issues in a way that is relevant and sensitizing to protecting health and safety and/or accurately estimate the effectiveness of risk controls. Three examples highlight this area. Before dust explosions at the Imperial Sugar[56] and West Pharmaceutical Services facilities,[57] audits identified the accumulation of dust. In the case of Imperial Sugar, dust accumulation impacting food quality was noted but not the possibility as a combustible hazard. Similarly, an insurance audit preceding the West Pharmaceutical Services facility accident noted dust on fire sprinkler heads, but combustibility potential was not assessed. Risk insurer audits at the New Cumberland A.L. Solutions titanium plant also recognized metal dust accumulations, but erroneously presumed adequacy in existing dust control measures and fire protection systems.[59]

The Waterfall train derailment accident also warrants discussion.[91] The official report highlighted the emphasis on safety by staff and safety committees, but this focus centered on personal and occupational safety rather than major hazards. This underscores how accidents can incubate within organizations. Despite activities like audits aimed at uncovering particular issues, they sometimes fail to do so or inadvertently overestimate safety arrangements – a phenomenon termed "interpretive failure."[14] This illustrates how audits can perpetuate a perception of adequacy even when substantial risks remain unaddressed. This study corroborates empirical[84,92,93] and theoretical work,[89] indicating that auditing, under certain conditions, can exacerbate the decoupling of risk – further widening the gap between the organization's perceived and actual degree of major risk control.

6.2.3 | Audits promote unwarranted confidence in risk management

Hutchinson et al.[89(p9)] previously argued that "some plans can be symbolically powerful yet functionally weak." Said differently, some safety artifacts can exert a sufficient influence over the subjective beliefs of people, like how "safe" people believe work to be, even though that artifact may provide little direct functional influence over hazardous work.

The present study substantiates this assertion. For instance, executives at a particular mine held the belief that their safety system's interwoven audits, regulations, and data would foster a more safety-sensitive culture.[65] While audits bolstered their safety confidence, the audits fell short in providing the executives with a comprehensive grasp of the mine's hazardous operations. Likewise, at another mine, an "apparently comprehensive, interlocking system of safety policies and audits" also failed, in hindsight,[2] to identify and rectify "the most basic and … obvious dangers".[94(p11)] This safety management system was used to reassure the site leaseholder that health and safety could be adequately governed.

A parallel can be drawn between the current finding of risk confidence misappropriation and Power's thesis on the audit society.[95] In Power's view, society's need for auditing assurance carries the risk of becoming a ritualistic and "cosmetic practice", hiding "real risk" and replacing it with an inert, transformed auditable outcome or artifact not reflecting reality.[95,96] This paper found evidence corroborating Power's contention, at least partially.

Given the scarcity of examples here, caution should be applied with interpreting these findings.

7 | GENERALIZABILITY AND TRANSFERABILITY

Like any study, a question remains about how to generalize and transfer the findings to organizational audits "in the wild." For example, most reports that directly mentioned audits came from the CSB. How should this be interpreted? The industries served by the CSB may encounter problems with auditing more frequently than other industries. However, this seems unlikely. An alternative possibility is that the CSB and its investigation methods are more prone to finding limitations in auditing systems due to investigatory thoroughness. However, both of these interpretations assume that accident reports document stable facts, "out there" in the world, clearly and unambiguously waiting to be found.

A more constructivist view may lead to conclusions at the other end of the epistemological spectrum. Here, the categories of audit failures found primarily within the CSB's dataset could, at best, demonstrate the mental models and politics of CSB investigations (see for instance[25,44]). Indeed, this study, by necessity, evaluated reports that were, themselves, separate in time from the incidents. As Dekker et al.[39(p5)] argue, the very act of "separating important or contributory events from unimportant ones is an act of construction, of the creation of a story," and therefore, this process of construction is "not the reconstruction of a story that was already there". Whilst we have used the example here of the over-representation of the CSB (most likely explained by the CSB genuinely finding more audit failures), the concern that accident reports construct rather than find problems with audits applies across all of the observations made in this paper. Likewise, the categorization of the audit failures by the study authors could be considered a second-order reconstruction.

In this study, we believe greater utility comes from interpreting the findings from an epistemological midpoint. This approach avoids dichotomizing perspectives, thereby preserving the broader potential of insights into organizational performance from both views. This stance is particularly valuable as epistemological viewpoints can function as "analytical aids" rather than rigid positions.[97] Hence, aligning more with Hopkins' perspective – while it is prudent to exercise caution against an overly literal interpretation of the findings, the overall picture is considered as useful evidence for the efficacy or otherwise of audits. This is significant because of the challenges of ascribing stable states and conditions within complex systems,[98–100] and the propensity of organizations to dismiss literal interpretations of accident findings as inapplicable, citing disparities between their context and the findings.[101,102] This tempered stance also accommodates the shallowness found within most of the included investigation reports

describing auditing failures, sometimes necessitating interpretation by the study's authors.

Overall, these findings can be used to build upon Hopkins' "persuasive over conclusive" stories[3(p6)]; enhancing an organization's "safety imagination"[103]–that is, the ability to break from "overly fixed … patterns of thinking."[103(p22)] While counterfactual reasoning may be a "fruitful exercise when trying to uncover potential countermeasures against … failures in the future,"[34(p375)] they still do not explain why people did what they did. Hence, enhancing safety imagination through counterfactuals demands judicious application but can enable organizations to better confront the uncertainties of risk.

Another consideration relates to the sample size. While the study systematically canvassed investigation reports across a range of onshore and offshore state and national regulatory agencies, few investigation reports overall discussed auditing in depth.

Expanding on the previous point, investigation reports were generally limited in their discussion of audit programs, and their sources of failure or effectiveness. This is problematic since a host of factors influencing audit effectiveness have been found. Factors include biased preconceptions and pre-existing knowledge of auditors,[104] time pressure and task complexity,[105] auditor experience, team composition, audit scope and audit techniques,[106,107] and managerial support.[108] Other factors influencing audit effectiveness include "gaming behavior" by auditees, like delaying information, pre-selecting evidence before the evidence, and directing auditors to areas that the auditees would like auditors to focus on.[109] In all, given the lack of information on audit failures in investigations, this study cannot systematically canvass the influence of such factors in audit failures other than the few items identified in investigation reports. An obvious suggestion is that investigations should place greater priority on how audits contribute to accidents.

8 | CONCLUSION

This examination of accident reports revealed consistent counterfactual themes. Failures to adhere to the idealized model, reflecting various failure types, manifested in audits. Some audits either overlooked crucial safety aspects or lacked methodologies to identify specific issues. Others overly fixated on documented systems instead of actual practices and operational risks, or reinterpreted safety concerns as non-safety matters. Additionally, some audits fostered unwarranted confidence in safety. Characterizing these findings as mere audit quality deficiencies oversimplifies the intricate role audits play in disaster incubation.

Despite the ideal standard, many audits exhibited a "comprehensive shallowness," delving excessively into minor system details and paperwork rather than addressing critical factors. A plausible interpretation is that, in the clarity of hindsight, critical aspects amid a sea of issues become evident, embodying "delusional clarity."[110] The counterfactual auditing ideal risks simplifying the complexity of safety-critical environments into tidy yet oversimplified axioms: should have, could have, would have. Nevertheless, these findings provide insights to drive safety imagination toward major accident prevention. Another implication is that organizations should clarify their audit expectations and implement mechanisms to test and monitor whether auditing achieves those expectations, especially in light of the possibility of silent audit failures. This may require concentrated efforts to implement valid indicators, feedback mechanisms, and critical perspectives on performance.

Moreover, the findings indicate that audits may not be acting as robust indicators of weak signals or early warning signs of danger. Indeed, paraphrasing Dixon-Woods et al.,[111] practitioners should ask whether their audits, in practice, work as tools of problem-solving or instead, the machinery of comfort-seeking.

AUTHOR CONTRIBUTIONS

Ben Hutchinson: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (equal); writing – original draft (lead); writing – review and editing (equal). **Sidney Dekker:** Conceptualization (supporting); methodology (supporting); supervision (supporting); writing – review and editing (supporting). **Andrew Rae:** Conceptualization (equal); investigation (equal); methodology (equal); supervision (lead); writing – review and editing (equal).

ACKNOWLEDGMENT

Open access publishing facilitated by Griffith University, as part of the Wiley - Griffith University agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

The author(s) received no financial support for the research, authorship, and/or publication of this article.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Ben Hutchinson https://orcid.org/0000-0002-6752-1930

ENDNOTES

[1] Many investigation reports lacked comprehensive details of the audit's scope, leading this paper to assume that certain missed issues/hazards fell within the scope based on the accident investigators' decision to discuss those issues in their reports.

[2] Notably, the coroner in the case of Century Resources drilling accident[94] did call out the clarity gained from hindsight.

REFERENCES

1. Appleton B. Piper alpha. *Learning from Accidents*. 3rd ed. Routledge; 2001.
2. Center for Chemical Process Safety. *Guidelines for Auditing Process Safety Management Systems*. 2nd ed. John Wiley & Sons; 2011.
3. Hopkins A. Issues in safety science. *Saf Sci*. 2014;67:6-14. doi:10.1016/j.ssci.2013.01.007
4. Downer J. "737-Cabriolet": the limits of knowledge and the sociology of inevitable failure. *Am J Sociol*. 2011;117(3):725-762.
5. Le Coze JC. Towards a constructivist program in safety. *Saf Sci*. 2012;50(9):1873-1887.

6. McCall JR. *Knowing versus Understanding: Adjusting the Contextual Lens in Safety Science*. Vol 314. EDP Sciences; 2020:01008.
7. Workcover NSW. *National Self-Insurer OHS Management System Audit Tool*. 2014. Accessed April 2, 2021. https://content.api.worksafe.vic.gov.au/sites/default/files/2018-06/National_self-insurer_OHS_management_system_audit_tool_Version_3.pdf
8. Standards Australia/Standards New Zealand. AS/NZS ISO 19011: 2019: Guidelines for auditing management systems. Published online 2019. https://www.techstreet.com
9. Alruqi WM, Hallowell MR. Critical success factors for construction safety: review and meta-analysis of safety leading indicators. *J Constr Eng Manag*. 2019;145(3):04019005.
10. Salas R, Hallowell M. Predictive validity of safety leading indicators: empirical assessment in the oil and gas sector. *J Constr Eng Manag*. 2016;142(10):04016052.
11. Lingard H, Hallowell M, Salas R, Pirzadeh P. Leading or lagging? Temporal analysis of safety indicators on a large infrastructure construction project. *Saf Sci*. 2017;91:206-220. doi:10.1016/j.ssci.2016.08.020
12. Dobrowolski Z. Forensic auditing and weak signals: a cognitive approach and practical tips. *Eur Res Stud J*. 2020;XXIII(Special Issue 2):247-259.
13. Macrae C. Early warnings, weak signals and learning from healthcare disasters. *BMJ Qual Saf*. 2014;23(6):440-445.
14. Turner BA, Pidgeon NE. *Man-Made Disasters*. Butterworth-Heinemann; 1997. Accessed September 4, 2017.
15. Rae AJ, Alexander RD. Probative blindness and false assurance about safety. *Saf Sci*. 2017;92:190-204.
16. Center for Chemical Process Safety (CCPS). *Guidelines for Investigating Process Safety Incidents*. Wiley; 2019.
17. Oswald D, Sherratt F, Smith S. Problems with safety observation reporting: a construction industry case study. *Saf Sci*. 2018;107: 35-45.
18. Guba EG, Lincoln YS. Competing paradigms in qualitative research. *Handbook of Qualitative Research*. Vol 2(163-194). Sage Publications, Inc.; 1994:105.
19. Dekker SW. Illusions of explanation: a critical essay on error classification. *Int J Aviat Psychol*. 2003;13(2):95-106.
20. Zou PX, Sunindijo RY, Dainty AR. A mixed methods research design for bridging the gap between research and practice in construction safety. *Saf Sci*. 2014;70:316-326.
21. Baillie C, Douglas EP. Confusions and conventions: qualitative research in engineering education. *J Eng Educ*. 2014;103(1):1-7.
22. Department of Mines, Industry Regulation and Safety. Information sheet for safety and health representatives. Investigating incidents. Published online 2018. https://www.dmp.wa.gov.au/Documents/Safety/MSH_IS_InvestigatingIncidents.pdf
23. National Heavy Vehicle Regulator. Incident investigation – Quick Guide-v1.0. Published online 2021. https://www.nhvr.gov.au/files/201809-0895-sms-quick-guide-incident-investigation.pdf
24. Dekker S. *The Field Guide to Understanding "Human Error."*. Ashgate Publishing, Ltd.; 2014.
25. Lundberg J, Rollenhagen C, Hollnagel E. What-You-Look-For-Is-What-You-Find–The consequences of underlying accident models in eight accident investigation manuals. *Saf Sci*. 2009;47(10): 1297-1311.
26. Nassehi A. Organizations as decision machines: Niklas Luhmann's theory of organized social systems. *Sociol Rev*. 2005;53(1_suppl): 178-191.
27. Reiman T, Rollenhagen C. Does the concept of safety culture help or hinder systems thinking in safety? *Accid Anal Prev*. 2014;68:5-15.
28. Antonsen S. Safety culture and the issue of power. *Saf Sci*. 2009; 47(2):183-191.
29. Dekker SW, Nyce JM. There is safety in power, or power in safety. *Saf Sci*. 2014;67:44-49.
30. Byrne RM. Counterfactual thought. *Annu Rev Psychol*. 2016;67: 135-157.
31. He Y, Payne SC, Yao X, Smallman R. Improving workplace safety by thinking about what might have been: a first look at the role of counterfactual thinking. *J Safety Res*. 2020;72:153-164.
32. Kahneman D. Varieties of counterfactual thinking. *What Might Have Been*. Psychology Press; 2014:387-408.
33. Morris MW, Moore PC, Sim DL. Choosing remedies after accidents: counterfactual thoughts and the focus on fixing "human error". *Psychon Bull Rev*. 1999;6:579-585.
34. Dekker SW. Reconstructing human contributions to accidents: the new view on error and performance. *J Safety Res*. 2002;33(3): 371-385.
35. Haddon-Cave C. *Return to an Address of the Honourable the House of Commons Dated 28th October 2009 for an Independent Review into the Broader Issues Surrounding the Loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006*. Stationery Office; 2009. Accessed June 16, 2017. http://www.official-documents.gov.uk/document/hc0809/hc10/1025/1025.pdf
36. Longford Royal Commission. *The Esso Longford Gas Plant Accident: Report of the Longford Royal Commission*. Govt. Printer for the State of Victoria; 1999.
37. U.S. Chemical Safety and Hazard Investigation Board. Investigation Report–Refinery Fire and Explosion and Fire. BP Texas City. March 23, 2005. Published online 2007. http://www.csb.gov
38. Royal Commission on the Pike River Coal Mine Tragedy. *Royal Commission on the Pike River Coal Mine Tragedy*. Government of New Zealand; 2012.
39. Dekker S, Cilliers P, Hofmeyr JH. The complexity of failure: implications of complexity theory for safety investigations. *Saf Sci*. 2011; 49(6):939-945.
40. Guilbault RL, Bryant FB, Brockway JH, Posavac EJ. A meta-analysis of research on hindsight bias. *Basic Appl Soc Psychol*. 2004;26(2-3): 103-117.
41. Treat JR, Tumbas NS, McDonald ST, et al. Tri-Level Study of the Causes of Traffic Accidents, Final Report. Report DOT-HS-034-3-535. Published online 1979.
42. Vesel C. Agentive language in accident investigation: why language matters in learning from events. *ACS Chem Health Saf*. 2020;27(1): 34-39.
43. Hofmann DA, Stetzer A. The role of safety climate and communication in accident interpretation: implications for learning from negative events. *Acad Manage J*. 1998;41(6):644-657.
44. Lundberg J, Rollenhagen C, Hollnagel E. What you find is not always what you fix—How other aspects than causes of accidents decide recommendations for remedial actions. *Accid Anal Prev*. 2010;42(6): 2132-2139.
45. Hopkins A. *Managing Major Hazards: The Lessons of the Moura Mine Disaster*. Allen & Unwin; 1999.
46. Hopkins A. *Disastrous Decisions: The Human and Organisational Causes of the Gulf of Mexico Blowout*. CCH Australia Limited; 2012. Accessed February 7, 2018.
47. Hopkins A. *Lessons from Longford: The Esso Gas Plant Explosion*. CCH Australia Ltd; 2015. Accessed August 5, 2017.
48. Le Coze JC. Storytelling or theory building? Hopkins' sociology of safety. *Saf Sci*. 2019;120:735-744.
49. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
50. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Educ Today*. 2004;24(2):105-112.
51. Corbin J, Strauss A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Inc.; 2008.

52. Flaherty S. *Understanding Accident Investigators: A Study of the Required Skills and Behaviours for Effective UK Inspectors of Accidents*. PhD Thesis. Published online 2008.
53. Roed-Larsen S, Stoop J. Modern accident investigation–Four major challenges. *Saf Sci*. 2012;50(6):1392-1397.
54. U.S. Chemical Safety and Hazard Investigation Board. Husky Energy Superior Refinery Explosion and Fire. Published online 2022. https://www.csb.gov
55. Cullen W. *The Public Inquiry into the Piper Alpha Disaster*. H.M. Stationery Office; 1990.
56. U.S. Chemical Safety and Hazard Investigation Board. Imperial Sugar Company Dust Explosion and Fire. Published online 2009. Accessed December 20, 2022. https://www.csb.gov
57. U.S. Chemical Safety and Hazard Investigation Board. West Pharmaceutical Services Dust Explosion and Fire. Published online 2004. https://www.csb.gov
58. U.S. Chemical Safety and Hazard Investigation Board. DuPont La Porte Facility Toxic Chemical Release. Published online 2019. https://www.csb.gov
59. U.S. Chemical Safety and Hazard Investigation Board. *AL Solutions Fatal Dust Explosion*; 2014. https://www.csb.gov
60. US Congress - Subcommittee on Oversight and Investigations. *BP's Pipeline Spills at Prudhoe Bay [Microform]: What Went Wrong? Hearing before the Subcommittee on Oversight and Investigations of the Committee on Energy and Commerce, House of Representatives, One Hundred Ninth Congress, Second Session, September 7, 2006*. U.S. G.P.O.: For sale by the Supt. of Docs., U.S. G.P.O; 2006. http://purl.access.gpo.gov/GPO/LPS79063
61. US Coast Guard. Report of Investigation into the Circumstances Surrounding the Explosion, Fire, Sinking and Loss of Eleven Crew Members Aboard the Mobile Offshore Drilling Unit Deepwater Horizon in the Gulf of Mexico, April 20–22, 2010. Published online 2011.
62. U.S. Chemical Safety and Hazard Investigation Board. E. I. DuPont De Nemours Co. Fatal Hotwork Explosion. Published online 2012. https://www.csb.gov
63. U.S. Chemical Safety and Hazard Investigation Board. Kuraray Pasadena Release and Fire. Published online 2022. https://www.csb.gov
64. The BP US Refineries Independent Safety Review Panel. The Report of the BP US Refineries Independent Safety Review Panel. Published online 2007. Accessed September 20, 2017. http://www.csb.gov/assets/1/19/Baker_panel_report1.pdf
65. Queensland Courts. Inquest into the Death of Peter Whitoria Marshall; 2006. https://www.courts.qld.gov.au
66. Mitchell A. In the Matter of DEEPWATER HORIZON. Report Regarding Transocean's Safety Management System and the ISM Code. Published online 2011. http://www.mdl2179trialdocs.com/releases/release201305171200030/TREX-040011.pdf
67. Noetic Solutions Pty Ltd. Review of PTTEP Australasia's Response to the Montara Blowout. For the Department of Energy, Resources and Tourism. Published online 2010. https://www.industry.gov.au/sites/default/files/2018-11/review-of-pttep-australasias-response-to-montara-blowout.pdf
68. U.S. Chemical Safety and Hazard Investigation Board. Synthron Chemical Explosion. Published online 2007. https://www.csb.gov
69. NSW Resources Regulator. *Report into the Death of Mr Andrew Bray at Snapper Mineral Sands Mine 12 August 2019*; 2021. https://www.resourcesregulator.nsw.gov.au
70. U.S. Chemical Safety and Hazard Investigation Board. CITGO Refinery Hydrofluoric Acid Release and Fire. Published online 2009. https://www.csb.gov
71. U.S. Chemical Safety and Hazard Investigation Board. Xcel Energy Company Hydroelectric Tunnel Fire. Published online 2010. https://www.csb.gov
72. Mine Safety and Health Administration. *Report of Investigation: Fatal Powered Haulage Accident October 14, 2020. Columbia Quarry Vulcan Construction Materials, LLC Columbia, Richland County, South Carolina. Mine ID No. 38-00013*. United States Department of Labor; n.d. https://www.msha.gov
73. U.S. Chemical Safety and Hazard Investigation Board. Valero McKee Refinery Propane Fire. Published online 2008. https://www.csb.gov
74. U.S. Chemical Safety and Hazard Investigation Board. Tosco Avon Refinery Petroleum Naphtha Fire. Published online 2001. https://www.csb.gov
75. U.S. Chemical Safety and Hazard Investigation Board. Tesoro Martinez Sulfuric Acid Spill. Published online 2016. https://www.csb.gov
76. U.S. Chemical Safety and Hazard Investigation Board. Formosa Plastics Vinyl Chloride Explosion. Published online 2007. https://www.csb.gov
77. U.S. Chemical Safety and Hazard Investigation Board. Chevron Richmond Refinery Fire. Published online 2015. https://www.csb.gov
78. COMAH Competent Authority. Buncefield: Why did it happen? The underlying causes of the explosion and fire at the Buncefield oil storage depot, Hemel Hempstead, Hertfordshire on 11 December 2005. Published online 2011. http://www.hse.gov.uk/comah/buncefield/buncefield-report.pdf
79. U.S. Chemical Safety and Hazard Investigation Board. DPC Enterprises Glendale Chlorine Release. Published online 2007. https://www.csb.gov
80. National Transportation Safety Board. Pacific Gas and Electric Company Natural Gas Transmission Pipeline Rupture and Fire, San Bruno, California, September 9, 2010. Pipeline Accident Report NTSB/PAR-11/01. Published online 2011.
81. U.S. Chemical Safety and Hazard Investigation Board. AirGas Facility Fatal Explosion. Published online 2017. Accessed December 21, 2022. https://www.csb.gov
82. Mogford J. Fatal accident investigation report: isomerization unit explosion final report. *Accident Report*. Published online 2005.
83. Hu X, Yeo G, Griffin M. More to safety compliance than meets the eye: differentiating deep compliance from surface compliance. *Saf Sci*. 2020;130:104852.
84. Hutchinson B, Dekker S, Rae A. Audit masquerade: how audits provide comfort rather than treatment for serious safety problems. *Saf Sci*. 2024;169:106348.
85. Blewett V, O'Keeffe V. Weighing the pig never made it heavier: auditing OHS, social auditing as verification of process in Australia. *Saf Sci*. 2011;49(7):1014-1021.
86. Hollnagel PE. *Safety-I and Safety–II: The Past and Future of Safety Management*. Ashgate Publishing, Ltd.; 2014.
87. Clarke L. *Mission Improbable: Using Fantasy Documents to Tame Disaster*. University of Chicago Press; 1999.
88. Hayes J, Hopkins A. *Nightmare Pipeline Failures: Fantasy Planning, Black Swans and Integrity Management*. CCH Australia Limited; 2015. Accessed September 2, 2017.
89. Hutchinson B, Dekker S, Rae A. Writing plans instead of eliminating risks: how can written safety artefacts reduce safety? *Saf Sci*. 2022;151:105738.
90. Hutchinson B, Dekker S, Rae AJ. Fantasy planning: the gap between systems of safety and safety of systems. 2018. Accessed November 5, 2018. https://www.researchgate.net/publication/325395758_Fantasy_planning_the_gap_between_systems_of_safety_and_safety_of_systems
91. McInerney P. Special Commission of Inquiry into the Waterfall Rail Accident. Final Report. Vol 1. Published online January 2005. Accessed December 20, 2022. https://nraspricms01.blob.core.windows.net/assets/documents/Waterfall-Rail-Accident/Waterfall-final-report-Volume-1.pdf
92. Batalden BM, Sydnes AK. Auditing in the maritime industry: a case study of the offshore support vessel segment. *Saf Sci Monit*. 2015;19(1):3.

93. Hohnen P, Hasle P. Making work environment auditable–A 'critical case' study of certified occupational health and safety management systems in Denmark. *Saf Sci*. 2011;49(7):1022-1029.
94. Queensland Courts. *Inquest into the Death of Scott Meho Karajic*; 2006. https://www.courts.qld.gov.au
95. Power M. *The Audit Society: Rituals of Verification*. OUP; 1997.
96. Power M. Modelling the micro-foundations of the audit society: organizations and the logic of the audit trail. *Acad Manage Rev*. 2021;46(1):6-32.
97. Dekker S. The criminalization of human error in aviation and healthcare: a review. *Saf Sci*. 2011;49(2):121-127.
98. Cook RI. How complex systems fail. Cognitive Technologies Laboratory. University of Chicago. Published online 1998. Accessed July 10, 2017. https://www.researchgate.net/profile/Richard_Cook3/publication/228797158_How_complex_systems_fail/links/0c96053410db96a89c000000.pdf
99. Perrow C. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press; 1999.
100. Woods DD, Branlat M. Basic patterns in how adaptive systems fail. *Resilience Engineering in Practice*. CRC Press; 2017:127-143.
101. Cook RI, Woods DD. Distancing through differencing: an obstacle to organizational learning following accidents. In: Hollnagel E, Woods DD, Leveson N, eds. *Resilience Engineering Concepts and Precepts*. Ashgate Publishing, Ltd., Published online; 2006:308-316.
102. Nicol J. *Have Australia's Major Hazard Facilities Learnt from the Longford Disaster?: An Evaluation of the Impact of the 1998 Esso Longford Explosion on Major Hazard Facilities in 2001*. Institution of Engineers, Australia; 2001.
103. Pidgeon N, O'Leary M. Man-made disasters: why technology and organizations (sometimes) fail. *Saf Sci*. 2000;34(1):15-30.
104. Kennedy J. Debiasing the curse of knowledge in audit judgment. *Account Rev*. Published online. 1995;70:249-273.
105. Bowrin AR, King J. Time pressure, task complexity, and audit effectiveness. *Manag Audit J*. 2010;25(2):160-181.
106. Shamki D, Alhajri TA. Factors influence internal audit effectiveness. *Int J Bus Manag*. 2017;12(10):143-154.
107. Arena M, Azzone G. Identifying organizational drivers of internal audit effectiveness. *Int J Audit*. 2009;13(1):43-60.
108. Singh KSD, Ravindran S, Ganesan Y, Abbasi GA, Haron H. Antecedents and internal audit quality implications of internal audit effectiveness. *Int J Bus Sci Appl Manag*. 2021;16(2):1-21.
109. Sweeney B, Pierce B. Audit team defence mechanisms: auditee influence. *Account Bus Res*. 2011;41(4):333-356.
110. Henriksen K, Kaplan H. Hindsight bias, outcome knowledge and adaptive learning. *Qual Saf Health Care*. 2003;12(suppl 2):ii46-ii50.
111. Dixon-Woods M, Baker R, Charles K, et al. Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study. *BMJ Qual Saf*. 2014;23(2):106-115.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hutchinson B, Dekker S, Rae A. How audits fail according to accident investigations: A counterfactual logic analysis. *Process Saf Prog*. 2024;43(3):441-454. doi:10.1002/prs.12579