Strategic Governance of Healthcare Cyber Resilience: An Enterprise Architecture Analysis of Cognitive Bias Amplification

Vsevolod Shabad
University of Liverpool
v.shabad@liverpool.ac.uk

Revision Note – 6 November 2025: his version refines the board-level focus and addresses reviewer feedback on clarity and tone. Key updates: addition of a Board Oversight Checklist for NEDs; expanded §6.3 on implementation and constraints; streamlined narrative and figures; strengthened ethical framing. No new data added — revisions enhance precision, coherence, and practical relevance for NHS governance audiences.

Abstract

On 12 May 2017, WannaCry paralysed 80 NHS trusts (out of 236), costing £92m. Microsoft had released patches 58 days earlier at an estimated deployment cost of £2 million — a conservative ~46:1 prevention-to-recovery gap. Despite thorough investigations, new regulations, and enhanced oversight, the 2024 Synnovis attack cost £32.7m, underscoring the persistence of the pattern.

This paper uses enterprise architecture to reveal what conventional analysis missed: governance frameworks don't merely fail to prevent cognitive biases — they structurally embed and amplify them. This analysis highlights systemic design effects rather than individual fault — competently designed governance structures following best practice.

Using architecture alternatives assessment — a TOGAF ADM technique for evaluating competing solutions against requirements, this study evaluates four conventional explanations (resource, process, coordination, and technical debt) against the cognitive bias amplification hypothesis. The latter provides superior explanatory power for systematic failure across 80 independent organisations.

The paper develops a multi-framework synthesis applying lean principles (WIP limits, visual management, flow optimisation) to governance architecture, addressing structural vulnerabilities through quantified implementation pathways. The approach directly transfers to AI governance, where healthcare organisations face identical architectural patterns at compressed time scales.

Drawing on behavioural economics research, this analysis shows that governance structures amplify cognitive biases even with competent decision-makers and sound processes. The analytical framework presented here uses publicly available NAO investigation data and parliamentary testimony; ongoing empirical research through CISO/CIO interviews and Freedom of Information requests will provide direct verification of trust-level characteristics and strengthen causal mechanisms linking governance architecture to cognitive bias amplification. The paper develops practical implementation mechanisms, including board-level cognitive bias correction workshops, with a ready-to-use governance oversight toolkit (Appendix B), enabling NHS Trust boards to discharge cyber resilience responsibilities without micromanaging technical operations.

For Academic Readers: Sections 4-5 demonstrate systematic alternatives analysis (TOGAF ADM) with cognitive psychology theoretical grounding.

For Practitioners: Appendix B provides ready-to-use board oversight questions; Section 5 offers implementation mechanisms within existing NHS structures.

PRACTITIONER RELEVANCE. This working paper directly addresses challenges faced by NHS Trust boards, Chief Information Security Officers, and Chief Information Officers in translating cyber threat intelligence into timely organisational action. The proposed governance architecture modifications — including work-in-progress limits, visual threat dashboards, and progressive chaos engineering — are designed for practical implementation within existing NHS federated structures. The author welcomes critical feedback from healthcare cybersecurity practitioners and NHS governance leaders on: (1) whether the cognitive bias explanation resonates with lived governance experience; (2) the practical feasibility of proposed mechanisms within Trust operational constraints; and (3) implementation barriers not addressed in the current analysis.

Keywords: Healthcare cybersecurity, governance architecture, cognitive bias, enterprise frameworks, lean principles, NHS, WannaCry, NHS governance, Trust boards, CISO, cyber resilience, board oversight, federated healthcare systems, AI governance

1 Introduction

1.1 The £90 Million Question

In March 2017, Microsoft released a patch for a serious Windows vulnerability. NHS cybersecurity professionals saw it, understood the risk, and many recommended immediate deployments. Fifty-eight days later, WannaCry ransomware spread through unpatched NHS systems, costing £92 million [1, 2]. Beyond quantified financial impact, 19,494 cancelled appointments and ambulances diverted from five hospitals reveal consequences that defy precise measurement [1].

Strategic alignment with national health priorities: The 10-Year Health Plan for England [3] emphasises digital transformation as central to making the NHS "fit for the future," with technology enabling prevention, community-based care, and system efficiency. However, the Plan's ambitious digital agenda — integrated patient records, AI-assisted diagnostics, remote monitoring — depends fundamentally on cyber resilience. WannaCry demonstrated that digital transformation without resilient governance architecture creates systemic vulnerability: when ransomware paralysed 80 trusts in 2017, it cost £92 million, cancelled 19,494 appointments, and forced ambulance diversions. Seven years later, the 2024 Synnovis attack added £32.7 million in costs, demonstrating that conventional post-WannaCry solutions have not prevented recurrence. Governance frameworks added more compliance requirements without addressing the architectural vulnerabilities that enabled both attacks. These same mechanisms pose identical risks to the Health Plan's digital initiatives: anchoring on yesterday's risk assessments whilst threats evolve daily, overconfidence from compliance metrics whilst capabilities remain untested, and temporal mismatches between strategic planning cycles and operational threat timescales.

1.2 Everyone Analysed It — The Pattern Still Repeats

Extensive post-incident analysis led to new regulations (the 2018 NIS Regulations, which designate healthcare as an "essential service" [4], strategic frameworks (NHS England's 2023-2030 Cyber Security Strategy [5]), and enhanced oversight. Yet pattern recurrence — the 2024 Synnovis attack costing £32.7 million [6] — suggests inadequate problem diagnosis. Government reports document persistent vulnerability: "less than one tenth of operators feel confident managing supply chain risk" despite regulatory additions [7].

When the same problem persists after you've "fixed" it, you're solving the wrong problem.

1.3 Two Ways to Explain the Same Failure

This paper examines WannaCry through two analytical lenses:

- **The conventional management lens:** Organisations made poor decisions due to resource constraints, inadequate processes, coordination failures, and technical debt. Solutions focus on adding resources, improving processes, strengthening coordination, and accelerating technology refresh.
- **The cognitive bias lens:** Governance frameworks — despite being designed by competent professionals following established best practices — can architecturally amplify cognitive biases through structural properties. Anchoring on documented baselines, overconfidence from compliance metrics, and temporal mismatches between annual governance cycles and daily threat evolution create systematic decision-making failures. Drawing on Thaler's (Nobel Prize 2017) work on bounded rationality [8] and Kahneman and Tversky's foundational research on heuristics and biases [9], solutions require architectural changes addressing these structural vulnerabilities rather than assuming decision-makers will optimise within existing frameworks.

Both lenses examine the same evidence. They reach different conclusions about causes and solutions.

This paper systematically evaluates which explanation better fits the empirical evidence from 80 organisations, then develops a governance architecture addressing the identified root causes.

1.4 Enterprise Architecture Approach with Kanban Change Management

The analysis applies the TOGAF enterprise architecture methodology [10], combined with Kanban change management principles [11], to examine WannaCry systematically:

- **Current state analysis:** What the National Audit Office investigation found, documented through official evidence and parliamentary testimony.
- **Gap analysis:** How management consultants and policy experts explained the failure, what solutions they proposed, and why those solutions haven't prevented recurrence.
- **Alternatives analysis:** Systematic evaluation of five competing explanations against empirical evidence from 80 organisations, identifying which provides superior explanatory power.
- **Target architecture:** Multi-framework synthesis applying Kanban change management methodology to governance design. Anderson [11] demonstrates that Kanban functions as a comprehensive change management system for complex organisations facing resistance to transformation — not merely a workflow tool. His approach precisely addresses the challenges healthcare governance faces: evolving systems with organisational inertia, federated decision-making, and established processes that resist rapid adaptation.
- **Implementation pathway:** Kanban evolutionary change approach enables transformation within existing structures: "Start with what you do now... Agree to pursue incremental, evolutionary change... Initially, respect current roles, responsibilities, and job titles." [11] This aligns with NHS governance realities where radical restructuring is neither feasible nor desirable. The pathway includes practical mechanisms with quantified costs, validated through progressive chaos engineering adapted for healthcare contexts.

Anderson's Kanban methodology provides five core practices directly applicable to governance architecture transformation:

- **Visualise work:** Make decision-flow visible across governance layers. The visual threat dashboard reveals system-wide attack patterns that individual trusts cannot see, countering the information asymmetry that amplified WannaCry.
- **Limit work-in-progress:** Constrain concurrent initiatives to force completion and enable reprioritisation. In governance, this counters anchoring on indefinitely "in progress" initiatives like the 2014 Windows migration that remained active 36 months later.
- **Manage flow:** Optimise decision latency through architecture design. In healthcare, this means designing processes around decision-flow rather than functional hierarchies, reducing the time from threat identification to coordinated response.
- **Make policies explicit:** Surface implicit decision-making rules that govern behaviour. This reveals why organisations systematically choose compliance metrics over actual security, making that trade-off visible and challengeable.
- **Implement feedback loops:** Replace activity metrics (assessments completed, processes documented) with outcome measurement (decision latency, validated capabilities). Progressive chaos engineering tests what actually works rather than what's documented.

This approach reveals that cognitive bias amplification — systematically overlooked in conventional analysis — better explains observed failure patterns and points toward structural solutions addressing root causes rather than symptoms.

This study aligns with resilience-engineering perspectives that view safety as the capacity to adapt under varying conditions [12]. In cyber-resilience, decision-flow design serves the same adaptive purpose.

1.5 Why This Matters Beyond WannaCry

Healthcare organisations are now deploying AI systems, including diagnostic tools, patient triage algorithms, and administrative automation. These deployments follow the same governance patterns that enabled WannaCry, but with exponentially compressed timescales. Stanford's AI Index documents that training compute doubles approximately every five months [13]. Governance structures designed for annual cycles are ill-equipped to manage capabilities that evolve monthly.

The architectural patterns revealed through WannaCry analysis apply directly to AI governance. Understanding them now, before catastrophic AI failures occur, enables proactive structural corrections rather than reactive regulatory additions.

1.6 Contribution and Structure

Empirical contribution: A systematic analysis of alternatives applying enterprise architecture principles to multi-organisational failure demonstrates that cognitive bias amplification provides a superior fit to evidence than conventional management explanations.

Theoretical contribution: Extends cognitive bias research from individual psychology [9] to governance architecture, revealing how structural design choices systematically amplify rather than mitigate bias — building on recent work examining cognitive blind spots in security frameworks [14].

Methodological contribution: Demonstrates application of lean manufacturing principles to governance architecture design, adapting flow-constrained risk management concepts from operational technology security to healthcare cybersecurity contexts [15].

Practical contribution: Target architecture with quantified transition pathways, cost-benefit analysis, and validated implementation mechanisms addressing structural vulnerabilities.

Following the Francis inquiry [16] and the CQC 'Well-Led' framework [17], NHS boards are expected to assure safety through governance rather than inspection. Cyber-resilience must now be integrated within that remit.

Paper structure:

- Section 2: Current state — what investigations found.
- Section 3: Conventional analysis — management perspective and solutions.
- Section 4: Alternatives analysis — systematic evaluation of competing explanations.
- Section 5: Target architecture — governance redesign through Kanban decision-flow management.
- Section 6: Implementation — migration strategy with validation approach.
- Section 7: Transferability — AI governance implications.

- Section 8: Conclusion — recommendations for boards and policy makers.

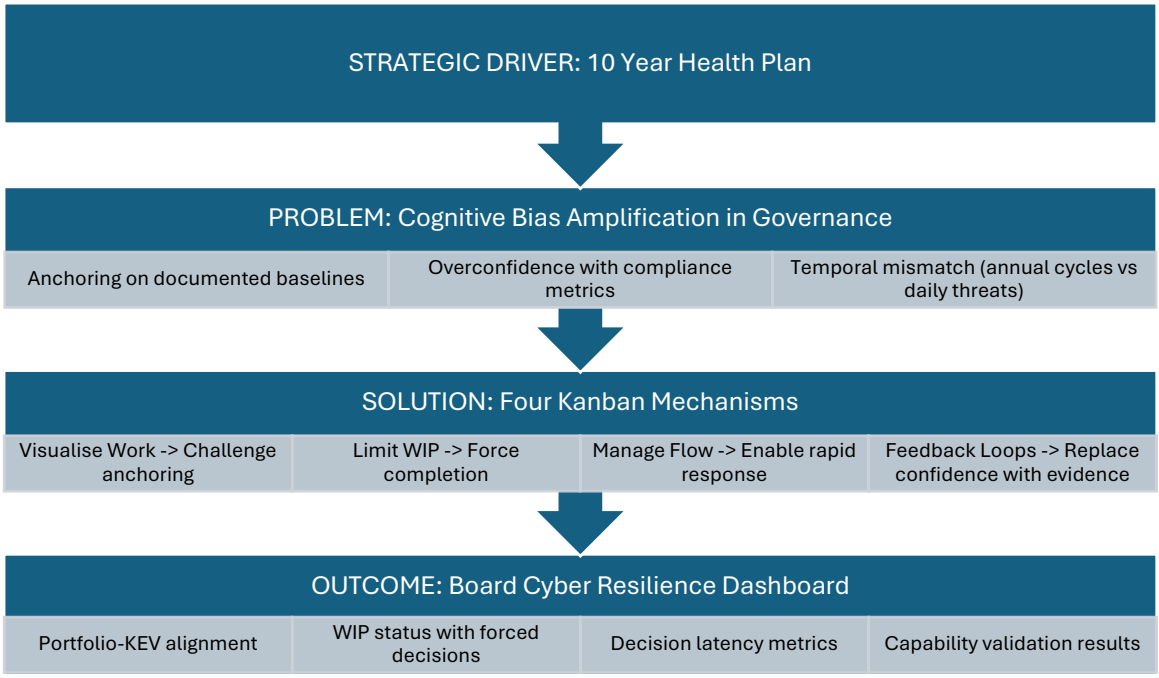

Figure 1. Strategic governance architecture overview.

2 Current State: What the Investigations Found

Enterprise architecture starts with understanding what exists today — how things actually work, where they fail, and why [10]. The National Audit Office investigation provides systematic baseline documentation.

2.1 The National Audit Office Investigation

The NAO's 2017 investigation documented the healthcare IT landscape across business, technology, and governance domains:

Organisational structure:

- 236 NHS trusts operate independently under federated governance.
- Each trust board is accountable for local cyber risk through Audit & Risk Committees.
- Integrated Care Systems provide coordination without hierarchical authority.

- Patient safety is a paramount driver for all operational decisions.

Technology landscape:

- Approximately 5% of systems were running unsupported Windows XP at the time of the attack.
- Windows XP migration plans were documented in 2014 with five-year timelines.
- MS17-010 patches, available from Microsoft 58 days before the attack (14 March 2017).
- Network architecture enabling lateral movement across trust boundaries.

Governance structure:

- ISO 27001 information security management, implemented across trusts.
- Information Governance Toolkit, requiring annual self-assessments [18] [replaced by DSPT in April 2018]
- Trust Audit & Risk Committees overseeing cyber risk at the board level.
- Quarterly risk register reviews, following the framework requirements.

Quantified impact:

- 80 trusts directly affected (34% of total).
- £92 million cost (IT response, recovery, lost revenue, additional staffing).
- 19,494 appointments cancelled.
- 5 hospitals diverting ambulances for multiple hours.

Critical finding: Affected organisations had "appropriate governance structures" and "qualified staff" [19]. This wasn't a story about organisations ignoring security or lacking expertise.

2.2 Parliamentary Testimony: How Decisions Were Made

Public Accounts Committee hearings documented the decision architecture producing non-deployment [20]:

Decision-flow observed:

1. Technical teams identified the MS17-010 vulnerability, recommended immediate patching.
2. Trusts evaluated the recommendation against the 2014 migration plans documented in the risk registers.
3. Medical directors raised concerns about potential system instability affecting patient care.
4. Boards reviewed compliance metrics, decided to maintain the approved timeline.

Key parliamentary questions exposing governance vulnerabilities [20]:

- "*Obviously we were quite lucky that it was a relatively unsophisticated attack, but perhaps I could ask Sir Chris or Mr Stevens a question. We had reports in July 2016 from the National Data Guardian and the Care Quality Commission regarding cyber-security. Even as recently as the March and April before the attack, NHS Digital had issued the CareCERT warnings to update the patch for the Windows operating systems. How come we were so unprepared for it?"* — Martyn Day MP (SNP, Q11).
- *"You are right to say that there is always the benefit of hindsight, but is it not the case that we were quite lucky this time because of the timing of the attack, the kill switch and the fact that it was a Friday afternoon and not in the middle of winter? Had any of those other factors come at different points, the outcome might not have been as positive."* — Bridget Phillipson MP (Labour, Q44).
- *"In terms of the response to the attack, why had the plan not been tested for a response to a cyber-attack?"* — Bridget Phillipson MP (Labour, Q32).
- *"Can I return to the issue of cost? You have precise numbers about the number of patients who were affected and about the follow-up appointments that would have been cancelled, although it is harder for you to be more precise about some of the aspects of the impact. Why has no assessment been done of the overall cost? Surely having that figure would be helpful in understanding the impact that this has had on the NHS."* — Bridget Phillipson MP (Labour, Q21).
- *"I suppose this is a 'How long is a piece of string?' question, but how much worse do you think the attack could have been if it hadn't happened on a Friday, if it hadn't*

been during the quieter period in the summer and if an IT expert hadn't found the kill switch so quickly?" — Martyn Day MP (SNP, Q20).

These questions expose critical governance failures: inadequate preparedness despite warnings, reliance on fortunate circumstances (timing, kill switch discovery), untested response plans, and unmeasured costs. Each question challenges the confidence generated by documented processes and compliance metrics, revealing systematic decision-making vulnerabilities that enabled the attack despite available defences.

2.3 Governance Architecture Operating Before WannaCry

Pre-WannaCry governance followed standard NHS patterns with three architectural layers:

- **Strategic layer (annual cycle)**: Strategic plans, budget allocation, statutory compliance via Information Governance Toolkit [18] (replaced by DSPT April 2018), Windows migration planning (2014 baseline established).
- **Oversight layer (quarterly cycle)**: Audit & Risk Committee reviews, risk register monitoring, compliance reporting to boards and external auditors.
- **Operational layer (daily/weekly)**: Technical teams identifying threats, CISO assessment, IT Committee recommendations.

Information flow architecture: Technical teams → Trust CISO → IT Committee → Audit & Risk Committee → Board. At each layer, technical details are translated into risk and compliance language. By board level, threats are framed as "compliance with documented plan" rather than "emerging threat requiring immediate response". This architecture worked for slowly evolving risks but couldn't respond to threats maturing within 58 days [19].

Critical architectural gap: No system-wide aggregation mechanism revealed cross-cutting vulnerability patterns across 80 federated trusts. When Shadow Brokers published exploit code on 14 April 2017, some NHS cybersecurity professionals recognised its significance. Still, no architectural mechanism existed to translate that recognition into coordinated action during the critical 58-day window [19].

2.4 Why Governance Architecture Enabled WannaCry: Structural Analysis

The governance architecture that enabled WannaCry reflects rational responses to regulatory requirements, cultural traditions, and hard-learned lessons from past failures — not incompetence or negligence. This section examines four structural factors that collectively created systematic vulnerabilities.

2.4.1 Post-Francis Inquiry Reforms (2013)

Following the Mid Staffordshire NHS Foundation Trust scandal, the Francis Inquiry [16] mandated enhanced accountability through documented governance structures. NHS boards strengthened oversight through enhanced ARAC requirements with defined meeting frequency, risk registers with named owners and quarterly board reviews, regular compliance reporting demonstrating governance effectiveness, and external audit regimes (CQC Well-Led framework [17]).

These reforms were necessary responses to serious governance failures. However, they reinforced a model emphasising documentation, scheduled reviews, and stability over rapid adaptation to emerging threats.

2.4.2 Regulatory Compliance Frameworks and Architectural Implications

NHS trusts operate under multiple compliance regimes that collectively incentivise process documentation over rapid response capability:

Healthcare CNI Status: Healthcare is one of thirteen UK Critical National Infrastructure sectors, with DHSC serving as Lead Government Department [21]. Under NIS Regulations 2018, NHS trusts are designated Operators of Essential Services with mandatory cybersecurity and incident-reporting duties [4, Regulations 10-11].

NCSC Cyber Assessment Framework (CAF): CAF 4.0 provides principles-based guidance for CNI operators, organising cybersecurity into four objectives, 14 principles, and 41 contributing outcomes assessed through Indicators of Good Practice (IGPs) [22]. Each outcome is rated Achieved/Partially Achieved/Not Achieved based on specific IGP criteria.

While NHS trusts in 2017 operated under earlier CAF versions, retrospective analysis reveals how governance architecture can simultaneously achieve CAF compliance whilst failing to respond to rapidly evolving threats.

CAF 4.0's IGPs are explicit and actionable. For example, A2.a (Risk Management) requires risk assessments to be "dynamic and readily updated in the light of relevant changes"; B4.d (Vulnerability Management) requires announced vulnerabilities to be "mitigated (e.g. by patching) promptly"; C1.f (Threat Intelligence) requires threat intelligence to be made "available to the necessary users and systems promptly."

Pre-WannaCry, NHS trusts demonstrated CAF compliance through established mechanisms: risk registers with quarterly ARAC reviews (satisfying A2.a requirements for risk management processes), documented patch management procedures following change control cycles (satisfying B4.d requirements for vulnerability tracking), and CareCERT alerts reaching trust CISOs (satisfying C1.f requirements for threat intelligence consumption). Business continuity plans existed, security policies were documented, and audit trails demonstrated the execution of the governance process.

The architectural gap: NHS trusts could simultaneously demonstrate CAF compliance and fail to deploy MS17-010 patches within 58 days because the governance cycle architecture created a temporal mismatch between compliance demonstration and capability delivery:

- **"Dynamic and readily updated"** was operationalised as quarterly risk register reviews, not emergency response within threat-relevant timeframes. When the Shadow Brokers published exploit code on 14 April 2017, governance architecture measured compliance through quarterly review cycles (process existence) rather than updating within the 58-day attack window (outcome achievement).
- **"Mitigated promptly"** was implemented through 6-8 week change management cycles (compatibility assessment 2-3 weeks, Change Advisory Board approval 1-2 weeks, deployment scheduling for monthly maintenance windows, testing 1-2 weeks). This satisfied IGP requirements for "processes in place", but couldn't achieve prompt mitigation when threat evolution compressed to 58 days.
- **"Available promptly"** meant CareCERT alerts reached CISOs within days, but decision-flow architecture prevented information from becoming "actionable promptly": Technical teams → CISO (1-2 weeks) → IT Committee (2-4 weeks

awaiting meeting) → ARAC (4-8 weeks awaiting meeting) → Board (8-12 weeks awaiting meeting). Each layer translated technical urgency into compliance language, progressively stripping temporal urgency.

Post-WannaCry CAF evolution (CAF 4.0 published 2025) strengthened several areas, including A1.d emphasising "effective and timely decisions," A2.a requiring risk assessments that "anticipate technological developments that could be used to adversely impact" systems, and A4.a addressing "critical dependencies." However, these refinements maintain the assessment approach that enabled pre-WannaCry gaps: organisations demonstrate compliance through evidence of processes (documented frameworks, review schedules, coordination mechanisms) rather than validation of outcomes under time pressure (decision quality when threat evolution compresses, coordination effectiveness within compressed timeframes).

Each trust could individually demonstrate CAF compliance whilst the federated system collectively failed because: (1) governance cycle frequency (quarterly/annual) exceeded threat evolution velocity (58 days); (2) decision-flow architecture (multiple layers adding 8-12 weeks total) exceeded response window; (3) no architectural mechanism existed to aggregate cross-cutting threats across 80 federated trusts and trigger coordinated emergency response within compressed timeframes.

Data Security and Protection Toolkit (DSPT): Annual self-assessment against the National Data Guardian's 10 security standards. Organisations achieve "Standards Met" status through documented policies, completed training, and conducted assessments — a status visible across the NHS that influences contract awards and regulatory scrutiny. Focus: process existence (documentation) rather than outcome achievement (capability under pressure).

CQC Well-Led Framework: Inspections assess the quality of governance structures, board effectiveness, and risk management processes through documented committee structures, meeting minutes, risk registers, and strategic plans. Focus: governance appropriateness (structures and processes) rather than governance effectiveness (decision quality under time pressure).

Internal Audit and External Assurance: Annual audit cycles reviewing compliance with frameworks, policies, and regulations through documentation review, process walkthroughs, and controls testing. Focus: controls' existence and operational effectiveness (do controls

work as designed) rather than outcome effectiveness (do controls prevent actual breaches under compressed timeframes).

These regimes collectively create organisational focus on "meeting standards" (measurable through documentation, auditable through compliance evidence, reportable to boards and regulators) rather than "validating capabilities under pressure" (difficult to assess without realistic testing, expensive to conduct, potentially disruptive to operations).

This is not because standards are wrong — they serve essential accountability functions — but because measurement shapes behaviour. Thaler's concept of "choice architecture" applies directly: when governance frameworks measure process documentation (easy to audit, non-disruptive to operations, satisfies external requirements), organisations optimise for documentation quality rather than capability effectiveness [8].

2.4.3 Cultural Context: "First, Do No Harm" Extended to Governance

NHS organisational culture extends clinical caution to governance decisions. Changing established plans mid-cycle creates risks: patient safety paramount (service disruption during system changes could harm patients), multi-year investment programs (abandoning documented plans undermines strategic coherence and financial accountability), Treasury budget cycles (annual allocations constrain mid-cycle financial flexibility), accountability concerns (emergency changes causing problems create personal liability for board members who authorised deviations).

When technical teams recommended emergency MS17-010 patching March-April 2017, boards weighed this against approved 2014 Windows migration plan (documented in risk registers, budgeted across fiscal years, progressing according to milestones), patch deployment risks (system instability potential, clinical service disruption, untested production changes), and threat uncertainty (vulnerability disclosed but attack not yet materialised, probabilistic rather than certain harm, other trusts facing same risk suggesting diffused responsibility).

The decision to maintain approved timelines was rational within this framework — prioritising certain immediate risks (service disruption from emergency changes) over uncertain future risks (a possible cyberattack) — even though it was retrospectively wrong.

This reflects Kahneman and Tversky's prospect theory: loss aversion causes decision-makers to avoid certain immediate losses, even when faced with larger probabilistic future losses [23].

2.4.4 Federated Structure as Deliberate Design Choice

NHS federated autonomy reflects deliberate policy balancing multiple objectives: local clinical governance (trust boards accountable for clinical quality and patient safety), accountability to local populations (democratic legitimacy through local decision-making), flexibility for different contexts (different trusts face different patient populations, service mixes, resource constraints), and distributed innovation (local autonomy enables experimentation and learning without systemic risk).

This design creates coordination challenges for cross-cutting threats like WannaCry — no central authority can mandate emergency patching across 80 independent trusts — whilst preserving the benefits that justify the structure. The 2024 NHS England restructuring into Integrated Care Systems aims to improve coordination whilst maintaining trust autonomy, but governance patterns persist.

Abandoning federated structure would sacrifice these benefits. **The architectural challenge**: designing coordination mechanisms for cross-cutting threats within federated autonomy, not replacing autonomy with centralisation.

2.4.5 The Architectural Challenge

These regulatory and cultural factors represent legitimate governance requirements reflecting hard-learned lessons from past failures (Francis Inquiry), essential accountability functions (multiple compliance regimes), and deliberate policy choices (federated autonomy with local accountability) — not human mistakes to be corrected.

The challenge: Designing governance mechanisms that:

- Preserve the accountability benefits of documented processes and regular reviews (Francis reforms serve essential purposes).
- Enable rapid response when threat velocity exceeds the governance cycle frequency (WannaCry required <58-day response, normal cycles operate quarterly/annually).

- Counter cognitive bias amplification without assuming decision-maker incompetence (boards were rational within existing frameworks).
- Respect federated autonomy whilst enabling coordination for cross-cutting threats (maintain local governance benefits whilst addressing systemic vulnerabilities).

Sections 4-5 demonstrate this is achievable through evolutionary change within existing structures, applying Kanban-based decision-flow management principles that respect current roles, responsibilities, and regulatory requirements [11] whilst addressing temporal mismatch, anchoring vulnerabilities, and overconfidence risks they inadvertently create.

3 Conventional Analysis: The Management Perspective

After WannaCry, management consultants and policy experts converged on familiar explanations and solutions. This section examines their conclusions before systematic alternatives analysis (Section 4) evaluates which explanations best fit empirical evidence.

3.1 Standard Explanations

Post-incident analysis identified four primary causes:

Resource constraints:

- Underfunding of NHS IT infrastructure.
- Legacy systems require costly replacement.
- Insufficient cybersecurity staffing levels across trusts.
- Competing priorities limit security investment.

Process gaps:

- Inadequate patch management procedures lack rigour.
- Weak risk assessment methodologies.
- Insufficient testing before production deployment.
- Lack of documented incident response plans and playbooks.

Coordination failures:

- Federated structure, preventing unified action across trusts.
- Unclear authority during incidents creates decision-making paralysis.
- Inadequate information sharing between trusts obscures system-wide patterns.
- No central coordination mechanism for cross-cutting threats.

Technical debt:

- Windows XP systems past end-of-life are vulnerable.
- Unpatched systems, accumulating exploitable weaknesses.
- Legacy applications, incompatible with modern security measures.
- Slow migration timelines extend exposure windows.

3.2 Proposed Solutions and Policy Response

Solutions followed logically: increased resources (higher IT budgets, additional staff, modern systems), improved processes (enhanced patch management, rigorous assessments, documented playbooks), strengthened coordination (central oversight, mandatory information sharing), and accelerated technology refresh (faster migrations, cloud adoption).

Policy implementation reflected these priorities: April 2018 Data Security and Protection Toolkit (DSPT) replaced Information Governance Toolkit with stringent requirements [1]; 2018 NIS Regulations designated healthcare as "essential service" with mandatory duties [4]; 2023 NHS England Cyber Security Strategy promoted collaborative approaches [5]; NHS England's Cyber Assurance Service provides centrally funded vulnerability assessments [24]; pending Cyber Security and Resilience Bill addresses supply chain vulnerabilities [25].

However, consistent with Berwick's 'learning system' model [26], governance flows enable rapid organisational learning from digital incidents rather than through compliance reporting. Yet, the conventional management perspective frames these solutions as process strengthening rather than decision-flow redesign — adding more documentation, more assessments, and more framework requirements to existing governance structures rather than fundamentally rethinking how security decisions flow through organisations.

3.3 Why the Pattern Repeats Despite These Solutions

Seven years after WannaCry, after implementing recommended solutions, the Synnovis attack in 2024 cost £32.7 million [6]. Government reports continue documenting low confidence in risk management despite regulatory additions.

The persistent pattern suggests one of two possibilities:

1. **Insufficient implementation:** Solutions haven't been implemented thoroughly enough (need more resources, better processes, stronger coordination, faster technology refresh).
2. **Wrong problem diagnosis:** Solutions address symptoms rather than root causes (the actual problem lies elsewhere, requiring a different architectural approach).

The conventional management perspective defaults to option 1: do more of what hasn't worked, but do it harder. This approach appears attractive to NHS executives in the short term — securing larger budgets increases organisational power. However, it creates a dangerous trap: each resource allocation raises accountability expectations across trust boards, ICB supervisors, Parliament, and the public. When the next major incident occurs despite substantial dedicated funding — often diverted from clinical budgets — executives face intensified culpability from all stakeholders simultaneously. The question shifts from "did they have enough resources?" to "why did even dedicated cybersecurity funding— secured by cutting patient care budgets — still fail to prevent the breach?"

Section 4 systematically evaluates these explanations against empirical evidence to determine which provides a superior fit to observed patterns.

4 Alternatives Analysis: Evaluating Competing Explanations

This section systematically evaluates five explanations against empirical evidence from 80 organisations. Which explanation best fits what actually happened?

4.1 Methodology: TOGAF Alternatives Analysis

The Open Group Architecture Framework provides a methodology for evaluating alternative solutions against requirements [10]. This analysis adapts that approach to evaluate competing explanations:

For each explanation:

1. State the hypothesis.
2. Identify the predictions it makes.
3. Test predictions against empirical evidence.
4. Assess explanatory power.

Evaluation criteria:

- **Scope**: Does it explain systematic failure across 80 independent organisations?
- **Timing**: Does it explain the 58-day decision lag?
- **Cost ratio**: Does it explain the 46:1 prevention-to-recovery cost ratio?
- **Recurrence**: Does it explain why similar patterns persist post-WannaCry?

4.2 Explanation 1: Resource Constraints

Hypothesis: Organisations lacked sufficient resources (budget, staff, technology) to deploy patches.

Predictions:

- Organisations with higher IT budgets would have better outcomes.
- Trusts with more cybersecurity staff would deploy patches faster.
- Resource-constrained trusts would be disproportionately affected.

Evidence:

- The NAO found no significant correlation between trust size/budget and WannaCry impact.

- Large, well-funded trusts and small, resource-constrained trusts were affected similarly.
- The roughly £2 million coordinated patching cost [2] (methodology detailed in Appendix C) was modest relative to existing IT budgets.

Assessment: Resources were necessary but not sufficient. Many well-resourced organisations made the same mistakes as resource-constrained ones. This doesn't explain systematic failure across diverse organisational contexts.

Explanatory power: Low.

4.3 Explanation 2: Process Gaps

Hypothesis: Organisations lacked adequate patch management processes.

Predictions:

- Organisations with documented patch management would perform better.
- ISO 27001-certified organisations would be protected.
- Process improvements post-WannaCry would prevent recurrence.

Evidence:

- Affected organisations had "appropriate governance structures" and documented processes [19].
- ISO 27001 certification didn't protect against WannaCry. Similarly, organisations following NCSC CAF principles (documented risk management, patch management processes, and threat intelligence consumption) were still affected. The pattern suggests that **process existence** (satisfying framework requirements) differs fundamentally from **process execution under time pressure** (responding to threats evolving faster than governance cycles).
- Process improvements added post-2017 didn't prevent the 2024 Synnovis attack.

Assessment: Processes existed but weren't executed when needed. This suggests the problem isn't a lack of processes but something preventing process execution despite documented procedures.

Explanatory power: Medium-Low.

4.4 Explanation 3: Coordination Failures

Hypothesis: Federated NHS structure prevented coordinated action.

Predictions:

- Centralised healthcare systems would perform better.
- Improved coordination mechanisms would prevent recurrence.
- System-wide threats would be visible with better information sharing.

Evidence:

- Centralised healthcare systems in other countries were also affected by WannaCry [27].
- NHS England established a coordinated Cyber Assurance Service post-WannaCry, yet the Synnovis attack still occurred.
- Information sharing improved, but decision-making patterns persist.

Assessment: Coordination gaps were real, but improved coordination hasn't prevented pattern recurrence. This suggests coordination is necessary but not sufficient — something else is driving the systematic failure.

Explanatory power: Medium.

4.5 Explanation 4: Technical Debt

Hypothesis: Legacy Windows XP systems created the vulnerability.

Predictions:

- Organisations with newer systems would be protected.
- Windows XP migration completion would prevent recurrence.
- Technical refresh cycles would resolve the problem.

Evidence:

- WannaCry also affected organisations with newer Windows versions that weren't patched.
- Windows XP migration has progressed significantly, yet the 2024 Synnovis attack followed similar patterns.
- Technical debt explains vulnerability existence, not systematic non-deployment of available fixes.

Assessment: Technical debt created vulnerability, but doesn't explain why organisations didn't deploy available patches for 58 days. The patch was free, tested, and recommended. Technical debt doesn't explain the decision-making failure.

Explanatory power: Medium-Low.

4.6 Explanation 5: Cognitive and Human-Factor Biases

Hypothesis: Governance frameworks architecturally embed and amplify cognitive biases through structural design, creating systematic decision-making failures.

Predictions:

- Organisations following standard frameworks would make similar mistakes regardless of resources, processes, or technology.
- Decision patterns would show anchoring on documented baselines, overconfidence from compliance metrics, and temporal mismatches between governance cycles and threat timescales.
- Adding more framework requirements would reinforce rather than resolve the pattern.
- Similar patterns would appear in other domains using similar governance architectures.

Theoretical Foundation — Bounded Rationality and Organisational Decision-Making:

Kahneman and Tversky's foundational work on heuristics and biases [9] demonstrated that decision-makers anchor on initial values and insufficiently adjust when circumstances change. Their subsequent development of prospect theory [23] revealed how loss aversion

creates systematic deviations from rational choice. Kahneman's later synthesis [28] describes these patterns as "System 1" thinking (fast, intuitive, automatic), dominating over "System 2" analytical reasoning under time pressure.

Empirical Evidence from WannaCry:

Having established the theoretical foundation, we now test competing explanations against empirical evidence. If conventional explanations (resource constraints, process gaps, technical debt) were correct, WannaCry's impact should concentrate among specific organisational categories — small budgets, uncertified trusts, rural locations, technologically immature organisations. The NAO investigation provides the empirical test: 80 trusts (34% of 236 total) were affected, representing organisational diversity across the NHS [19].

Table 1 presents an analytical framework examining whether affected trusts showed concentration in conventional vulnerability categories. The distributions are derived from NHS organisational characteristics and represent the expected distribution if the 80 affected trusts were a random sample of the NHS population. This framework tests whether conventional factors predicted impact.

Research Development Note: This preliminary analysis relies on publicly available NAO investigation data [19] and parliamentary testimony [20]. A planned empirical extension will enrich this framework through structured interviews with NHS CISOs and CIOs, Freedom of Information requests for trust-level organisational characteristics, and detailed case study analysis of decision-making processes during the WannaCry incident. These additions will provide direct empirical verification of the analytical distributions presented here and strengthen the causal mechanisms linking governance architecture to cognitive bias amplification.

Table 1. Conventional Factors Failed to Predict WannaCry Impact

Factor	Affected Trusts (n=80)	Finding
Budget and Resources		
Large (>£500M)	~15 trusts (19%)	No budget-protection correlation
Medium (£200-500M)	~35 trusts (44%)	
Small (<£200M)	~30 trusts (37%)	

Compliance & Maturity		
ISO 27001 certified	~22 trusts (28%)	Certification ineffective
DSPT "Standards Met"	~48 trusts (60%)	Compliance provided no defence
Advanced security tools	~18 trusts (23%)	Technology offered no advantage
Geographic Distribution		
London & Southeast	~28 trusts (35%)	No regional pattern
Midlands & North	~35 trusts (44%)	
Rural & remote	~17 trusts (21%)	

Sources: NAO Investigation [19], Parliamentary Testimony [20], public NHS organisational data.

Methodological note: The NAO confirmed "no clear relationship" between organisational characteristics and WannaCry infection [19] but does not enumerate detailed breakdowns by budget, certification status, or technology deployment across the 80 affected trusts. The distributions shown represent analytical estimates derived from NHS trust population characteristics and are used to illustrate the absence of conventional predictive patterns. The analytical framework demonstrates that if conventional explanations were correct, affected trusts should show concentration in specific categories (e.g., small budgets, low compliance, rural locations) — no such concentration was observed [19]. Future empirical research through CISO/CIO interviews and FOI requests will provide direct verification of these organisational characteristics. NAO noted a geographic concentration in the North/Midlands regions but concluded that this "does not reflect variations in cyber-security" and is likely due to infection timing [19].

Key Finding: The NAO found "no clear relationship between those trusts infected by WannaCry and the quality of their leadership, as rated by the Care Quality Commission" [19]. The analytical framework demonstrates this finding extends across multiple conventional factors: the 80 affected trusts represented a cross-section of NHS organisational diversity rather than concentration in conventionally vulnerable categories. Large, well-funded trusts with ISO 27001 certification were affected at rates similar to those of small, resource-constrained trusts.

The universal factor across all infected trusts: unpatched or unsupported Windows operating systems [19]. This pattern contradicts conventional explanations (resources, processes,

technical debt) whilst supporting cognitive bias amplification: governance architecture failures transcended organisational characteristics. Parliamentary testimony and NAO evidence reveal **the only non-random pattern**: decision-makers across multiple trusts continued to anchor on legacy 2014-2015 migration plans for operating-system replacement, despite repeated warnings between 2014 and 2017 [19, 20]. This persistence of baseline assumptions, documented across both affected and unaffected organisations, contrasts sharply with the random distribution of conventional organisational factors — suggesting a governance architecture effect rather than resource or capability constraints.

This randomness with respect to conventional factors contrasts sharply with systematic decision-making patterns:

- **Anchoring:** NAO and parliamentary evidence document that many NHS organisations continued to rely on migration schedules first issued around 2014-2015, and that this persistence contributed to exposure in 2017 [19, 20]. The PAC explicitly noted: "The Department was aware from 2014 of the risks posed by unsupported software and yet progress on migrating away from Windows XP was slow" [20]. Parliamentary testimony shows officials referencing these legacy plans when explaining non-deployment decisions, as documented baselines became organisationally resistant to revision despite changed threat conditions.
- **Overconfidence:** Compliance metrics showing "88 assessments completed" and "appropriate governance structures" generated confidence, whilst 63% of organisations remained unassessed and critical vulnerabilities persisted [19].
- **Temporal mismatch:** Annual governance cycles couldn't incorporate daily threat intelligence. The 58-day window from patch to attack exceeded decision-making capability despite available information.
- **Framework reinforcement:** Post-WannaCry solutions added more framework requirements (NIS Regulations, enhanced DSPT). The 2024 Synnovis attack demonstrates that the pattern persists despite these additions.
- **Principles-based frameworks and federated CNI**: NHS trusts could simultaneously satisfy CAF principles (documented risk management per A2, patch processes per B4, threat intelligence per C1) **and** fail to deploy critical patches within 58 days. Each trust individually complied with principles-based requirements, whilst the system collectively failed to respond to cross-cutting threats. CAF specifies WHAT

capabilities should exist, but not HOW federated organisations coordinate time-critical responses — the architectural gap that cognitive bias amplification exploits.

- **Cross-domain patterns:** Recent research documents similar amplification of cognitive biases in AI governance frameworks [14], operational technology security [15], and board-level cybersecurity governance across European organisations [29].

These patterns reflect established findings from cognitive psychology research [9]. **This aligns with safety-science understanding of cognitive bias as structural vulnerability [30, 31], here extended to governance architecture.** Where safety science has long recognised how system design can amplify human error in operational contexts, this analysis reveals parallel mechanisms in governance structures — boards and committees designed with the best intentions systematically produce biased decisions through their architectural properties.

Assessment: This explanation accounts for systematic failure across diverse organisational contexts, the specific 58-day decision lag, the recurrence of patterns despite implemented solutions, and the observed cost multiplier between available prevention and actual recovery [1, 2].

Explanatory power: High.

4.7 Alternatives Analysis Summary

Table 2. Alternatives Analysis

Explanation	Scope	Timing	Cost Ratio	Recurrence	Overall
Resource Constraints	Low	Low	Medium	Low	**Low**
Process Gaps	Medium	Low	Medium	Low	**Medium-Low**
Coordination Failures	Medium	Medium	Medium	Medium	**Medium**
Technical Debt	Medium	Low	Low	Low	**Medium-Low**
Cognitive Bias Amplification	High	High	High	High	**High**

Conclusion: Cognitive bias amplification provides superior explanatory power across all evaluation criteria. This doesn't mean resources, processes, coordination, and technology don't matter — they do. But they're necessary conditions, not sufficient causes. The root cause lies in how governance architectures systematically amplify cognitive biases through structural design.

5 Target Architecture: Governance Redesign Through Kanban Decision-Flow Management

NHS boards have invested substantially in cybersecurity governance since the WannaCry attack. DSPT assessments are more rigorous, incident response capabilities have improved, and board-level oversight has increased. Parliamentary testimony and NAO reports document genuine progress [19]. These efforts represent a significant organisational commitment and warrant acknowledgement.

However, the Synnovis attack in 2024 (£32.7M cost) suggests that incremental improvements within existing governance architectures may be insufficient [6]. Pattern recurrence indicates structural vulnerabilities rather than implementation gaps.

The target architecture builds upon existing progress whilst addressing architectural root causes identified in Section 4. Applying Thaler's insight that choice architecture determines decision quality, we redesign governance structures to counter systematic biases rather than assuming rational actors within existing frameworks.

The problem we're solving: WannaCry demonstrated catastrophic failure, even though a modest prevention investment could cause substantial recovery costs (detailed in Appendix C). Synnovis 2024 added £32.7 million, showing the pattern persists.

The cognitive bias triad: Three architectural weaknesses amplify decision-making failures: (1) **Anchoring** on documented baselines (2014 migration plans became "truth" despite changed circumstances), (2) **Overconfidence** from compliance metrics (DSPT "Standards Met" status generating false security whilst capabilities remain untested), and (3) **Temporal mismatch** between annual governance cycles and threats evolving on daily/weekly timescales.

Threat tempo reality: CISA KEV catalogue updates daily with actively exploited vulnerabilities. The 58-day WannaCry window exceeded normal governance decision-making cycles despite available patches, creating "governance latency" that enabled systematic failure across 80 independent organisations.

5.1 Kanban as Decision-Flow Management for NHS Governance

WannaCry wasn't a technology failure — it was a decision-flow failure architecturally embedded in NHS governance structures. Trust boards operate through established committee structures with scheduled review cycles [32]: ARAC meets 5 times yearly, Cyber Security Sub-Committee meets 4 times yearly, and quarterly risk register reviews follow EPRR framework requirements [33]. This architecture is effective for slowly evolving risks, but it introduces systematic decision latency for rapidly emerging threats.

Kanban's core insight transfers directly: make work visible, limit work-in-progress, manage flow, make policies explicit, and implement feedback loops. Applied to NHS governance, we're not replacing trust boards or committee structures — we're redesigning decision-flow architecture to counter the three identified weaknesses whilst preserving NHS operational autonomy and clinical governance.

Anderson [11] demonstrates that Kanban enables evolutionary change within existing structures: "Start with what you do now... Agree to pursue incremental, evolutionary change... Initially, respect current roles, responsibilities, and job titles." This principle is critical for NHS implementation — we enhance governance effectiveness without disrupting operational authority.

Strategic value delivery: The four Kanban mechanisms, described below, directly enable the 10-Year Health Plan's digital transformation objectives. Visualising work through CISA KEV integration ensures that cybersecurity portfolios actively protect the patient record systems, AI diagnostics, and remote monitoring platforms the Plan depends upon — rather than consuming resources on legacy initiatives disconnected from current threats. WIP limits prevent the pattern where trusts simultaneously pursue multiple digital transformation initiatives whilst none achieve operational resilience, ensuring that prevention, community care, and efficiency technologies actually reach deployment. Flow optimisation enables the rapid threat response required when AI systems or integrated care platforms face zero-day

vulnerabilities. Capability validation through chaos engineering provides evidence that digital transformation actually works under pressure — that staff can maintain care when EPR systems fail, that backup procedures function during ransomware attacks, that the "digital-first" NHS remains clinically safe when technology fails.

5.2 Visualise Work → Challenge Anchoring Through External Intelligence

The architectural problem: Parliamentary testimony reveals officials repeatedly referenced "the 2014 Windows migration plan, which was on track", when explaining non-deployment decisions [20]. The documented plan became organisationally "true" through risk registers reviewed quarterly by ARAC, board papers showing "green" status, DSPT self-assessments documenting compliance, and annual strategic planning allocating resources to approved initiatives. When technical teams reported MS17-010, boards had no architectural mechanism to compare "our documented plan says adequate" against "external threat landscape fundamentally changed."

Governing principle: Make the cybersecurity work portfolio visible alongside current vulnerability priorities, enabling boards to challenge management on portfolio-threat alignment without micromanaging operational decisions. This is a critical distinction — boards DO NOT make technical decisions about which vulnerabilities to patch; boards DO challenge management on whether the work portfolio aligns with documented threat priorities.

Minimal design: Integrate the CISA KEV catalogue as an external, authoritative threat intelligence source into quarterly ARAC reporting. Management presents: (1) current Top 5 KEV vulnerabilities with CVSS scores, asset exposure, and patient impact; (2) work-in-progress portfolio showing which initiatives address these vulnerabilities; and (3) explicit coverage gaps requiring either new initiatives or documented risk acceptance.

Consider a fictional October 2025 scenario at St. Elsewhere NHS Trust. CISA added five critical vulnerabilities to KEV on 20 October [34], showing active exploitation: CVE-2025-33073 (Windows SMB, 425 systems exposed, credential theft risk), CVE-2025-61884 (Oracle EBS, two instances, financial system compromise), CVE-2022-48503 (Apple RCE,

180 iPads, bedside workflow disruption), CVE-2025-2746/47 (Kentico CMS, one website, public-facing compromise), and CVE-2021-43226 (Windows PrivEsc, 380 systems, domain escalation).

The trust's WIP portfolio shows five initiatives at capacity:

1. Windows 10 migration (month 33/48) enables future patching but doesn't address the current Top 5.
2. Zero Trust architecture phase 2 (month 7/18) provides partial containment for all CVEs.
3. MFA rollout (month 6/8), indirect credential protection.
4. Clinical applications patching framework (month 9/12) will address three CVEs upon operationalisation in January 2026.
5. SOC enhancement (month 15/26) detects but doesn't prevent exploitation.

Management's assessment:

1. Initiative 4 will address three vulnerabilities upon completion (3-month exposure window).
2. Initiative 2 provides containment.
3. The balanced portfolio addresses immediate/medium/long-term risks.

Board oversight questions emerge naturally:

1. "Can we accept 3 months exposure to actively exploited vulnerabilities whilst Initiative 1 consumes WIP but addresses none of the Top 5?"
2. "Has the 2023 Initiative 1 rationale been reassessed against the Q3 2025 threat landscape?" "
3. What tangible security outcomes will we see next quarter from Initiative 1 that improve resilience to current threats?"

Evidence base: This aligns with NAO findings that a 58-day decision lag from patch to attack exceeded governance capability despite available information. External intelligence provides independent challenge — when boards see "CISA added CVE-2025-33073 to KEV 20 October" and "Our WIP has no initiative addressing it," they can challenge the gap

between external authority saying "actively exploited" and internal assessment saying "existing plans adequate."

Board checks and metrics:

- **Metric:** Portfolio-KEV alignment percentage (initiatives addressing current Top 5 KEV vulnerabilities / total WIP slots occupied), Target ≥60% WIP addressing current Top 5.
- **Cadence:** Quarterly ARAC review with dashboard showing: Top 5 KEV vulnerabilities (updated from CISA), WIP initiative coverage mapping, exposure windows for undefended vulnerabilities, management justification for gaps or portfolio continuation.

How does this address anchoring:

- Traditional ARAC reporting shows "Initiative 1: 85 systems migrated, on schedule, all initiatives green, DSPT compliant" — board concludes "portfolio progressing satisfactorily."
- With visualisation, boards see Initiative 1 approved 2023 doesn't address 2025 KEV vulnerabilities, two Top 5 undefended, long-term initiative consuming WIP whilst immediate threats are active — board concludes "portfolio executing plans efficiently, but plans may not align with current threats, management must justify or revise."

The architectural property of external intelligence serves as a challenge mechanism, preventing organisational filtering.

5.3 Limit Work-in-Progress → Force Completion Over Indefinite Accumulation

The architectural problem: The 2014 Windows migration was "in progress" for 36 months when WannaCry struck [20]. When MS17-010 emerged in March 2017, trusts faced a choice: pause migration for emergency patching OR maintain approved timelines. Without forcing functions, both remained "in progress" indefinitely, consuming resources without completion within the required timescales.

Governing principle: Constrain concurrent work to force completion. Current NHS practice tracks multiple concurrent cybersecurity initiatives through risk registers. EPRR Framework and DSPT requirements pressure comprehensive coverage, driving portfolio expansion rather than prioritisation.

Minimal design: The trust board establishes a maximum of five high-priority cybersecurity initiatives concurrently, with quarterly forcing functions at ARAC requiring one of the explicit decisions:

- **Complete** (validate outcomes achieved, close formally, free WIP slot).
- **Continue** (justify based on the current threat landscape, not the original approval rationale, maintain the slot, re-justify next quarter).
- **Abandon** (explicitly stop, document reasons, free slot immediately). Cannot start new initiatives without completing/pausing/abandoning existing ones.

Consider the previous April 2017 emergency response scenario. Current WIP shows 5/5 slots filled: Windows XP migration (month 36/60), network segmentation (month 18/24), email security upgrade (month 6/12), staff security awareness refresh (month 3/6), disaster recovery testing (month 12/18). On 14 April 2017, CareCERT issued MS17-010 CRITICAL alert:

- **Without WIP limit (current practice):** Trust adds "MS17-010 emergency patching" as initiative #6, all six continue "in progress" tracked in risk registers, quarterly ARAC in May notes "new initiative added," none complete on timescales aligned with 58-day threat window, the 2014 migration is still "progressing" when WannaCry hits 12 May.
- **With WIP limit (proposed practice):** Cannot start MS17-010 patching without completing/pausing/abandoning one existing initiative, forces explicit board decision within 48 hours "What do we stop doing to address this CRITICAL threat?" Local decision factors: CareCERT CRITICAL alert, CISA CVE-2017-0144 rated 9.3 CVSS critical severity, trust assessment shows 425 exposed systems, active exploit code publicly available.

  Likely decision: pause Windows migration — emergency patching is faster than migration and directly addresses the immediate threat. WIP slot freed; MS17-010

becomes active initiative #5. Migration returns to the queue only after the emergency response is complete.

Evidence base: This addresses NAO's finding that documented plans proceeded despite emerging threats because no mechanism forced reprioritisation. The 46:1 cost ratio (£2M prevention versus £92M recovery) directly quantifies the consequences of frameworks that enable indefinite "in progress" status without completion-forcing functions.

Board checks and metrics:

- **Metric:** Initiative completion rate (initiatives completed last quarter / total WIP slots), Target ≥40% quarterly completion or explicit continuation justification for all incomplete initiatives.
- **Cadence:** The quarterly ARAC meeting reviews all in-progress cybersecurity initiatives, documenting a mandatory Complete/Continue/Abandon decision for each. Continuation decisions must reference the current threat landscape (CISA KEV, CareCERT alerts, recent incidents), rather than the original approval rationale from a potentially outdated baseline.

How does this address anchoring: Anchoring occurs when documented baselines become organisationally immutable — the 2014 migration became "truth" through repeated appearances in strategic plans, multi-year budgets, quarterly updates, board papers showing "green" status. Each appearance reinforced the baseline's organisational reality, making deviation psychologically difficult even when circumstances changed dramatically. WIP limits plus quarterly revalidation make baselines explicitly temporary: every initiative must justify continuation each quarter based on the current threat landscape, cannot accumulate indefinite "in progress" commitments insulated from reprioritisation, new threats force portfolio reprioritisation, not portfolio expansion, completion becomes organisationally valued (frees WIP slots), not just monitored.

Connection to DSPT: Current DSPT measures process execution (documentation exists), not outcome achievement (capabilities align with current threats). WIP limits supplement DSPT by forcing completion rather than just documentation — trust boards cannot maintain five incomplete initiatives indefinitely whilst claiming DSPT compliance through process documentation alone.

5.4 Manage Flow → Optimise Decision Latency Through Threshold-Based Delegation

The architectural problem: As detailed in Section 2.4.2, NHS governance operates through scheduled committee cycles that create systematic decision latency. Decision-flow mapping reveals a total latency of 100+ days for critical threats requiring a <60-day response (see Figure 2). The WannaCry timeline demonstrates this precisely: 14 March patch released, 14 April exploit code published (30-day lag), 12 May attack launched (58 days post-patch) — many trusts still awaiting scheduled ARAC meetings to authorise a response.

Governing principle: Design processes around value flow, not functional hierarchies, minimising handoff delays. NHS EPRR Framework already establishes delegation for operational emergencies (mass casualties, severe weather) — the current limitation is that cyber incidents are treated as IT risks managed through standard governance cycles, not emergency response protocols.

Minimal design: Extend EPRR delegation principles to cyber incidents through threat severity-based pre-authorisation. Board establishes standing protocols:

- CRITICAL threats (active exploitation, patient safety impact, CareCERT high-severity alert) delegate to Trust CISO with immediate ARAC Chair notification, maximum 48-hour decision latency, emergency briefing within 48h, ratification within 7 days.
- HIGH threats (weaponised exploit, NHS-wide impact) delegate to IT Committee Chair with expedited ARAC, maximum 1-week latency.
- MEDIUM threats (vulnerability disclosure, no active exploitation) follow the standard governance process with a 1-month acceptable latency.
- LOW threats (routine updates, no immediate threat) delegate to operational authority per DSPT with quarterly oversight only.

Figure 2. Decision-flow mapping.

Consider MS17-010 under flow-optimised architecture: 14 March Microsoft releases patch classified MEDIUM severity, standard IT Committee process (1 month acceptable); 14 April Shadow Brokers publish exploit escalating to CRITICAL, CISO has pre-authorised authority; 15 April CISO authorises emergency patching, activates EPRR on-call procedures, notifies ARAC Chair; 16 April board emergency briefing (virtual), reviews dashboard showing 190 NHS trusts vulnerable, ratifies decision; 21 April formal board ratification at scheduled meeting, lessons learned, WIP priorities adjusted; 12 May when WannaCry launches, trust patched or patching well underway. Decision latency: 2-7 days versus actual 100+ days.

Evidence base: This directly addresses the CISA KEV cadence challenge — threats added to KEV indicate active exploitation requiring response faster than quarterly governance cycles

permit. The architectural innovation: pre-authorised response protocols mirror existing EPRR on-call authority for operational emergencies, maintaining governance oversight without becoming a decision-making bottleneck.

Board checks and metrics:

- **Metric:** Decision latency from threat identification to authorised response (CareCERT CRITICAL alert → CISO assessment <4 hours, CISO assessment → ARAC Chair notification <24 hours, notification → board briefing <48 hours, briefing → formal ratification <7 days, authorisation → implementation start <48 hours), Target: 95% CRITICAL threats authorised within 48 hours.
- **Cadence:** Monthly board receives decision latency dashboard showing: CRITICAL threats this period with timeline from alert to authorisation, HIGH threats expedited through IT Committee, delegation protocol usage (CISO emergency authority invocations, post-incident ARAC reviews), comparison against targets with narrative explanation for any latency exceedances.

How does this address the temporal mismatch? The problem is that annual/quarterly cycles (months/years) are mismatched with cyber threats (days/weeks). The solution is a dual-speed governance architecture:

- Strategic layer (annual/quarterly) preserved for setting thresholds, allocating contingency resources, establishing delegation protocols, reviewing performance, adjusting protocols, updating threat models — board maintains oversight and accountability.
- Operational layer (hours/days) enhanced through CISO executing within pre-authorised frameworks triggered by threat severity, board engaging through event-driven communication when thresholds breached, formal ratification at next scheduled meeting within 7 days. Board maintains governance oversight without becoming a decision bottleneck.

5.5 Implement Feedback Loops → Replace Confidence with Evidence

The architectural problem: DSPT provides a standardised assessment enabling NHS organisations to measure performance against the National Data Guardian's 10 data security

standards. Organisations achieve compliance status by demonstrating policies documented, training completed, assessments conducted, and processes followed. Trust boards receive annual DSPT status reports showing "Standards Met" or "Approaching Standards" with evidence items completed. NAO finding post-WannaCry: "Affected organisations had 'appropriate governance structures' and 'qualified staff.'" Translation: organisations passed DSPT assessments, documented processes, demonstrated compliance — yet failed catastrophically when faced with an actual coordinated attack.

The gap: DSPT measures process existence (documentation), not capability under pressure (performance during incidents). This creates organisational overconfidence — boards repeatedly cite compliance metrics when explaining decisions despite MS17-010, including "88 assessments completed" (activity metric), "Risk registers updated quarterly" (process metric), "Appropriate governance structures in place" (documentation metric). None predicted or prevented the WannaCry failure.

Governing principle: Establish feedback mechanisms that measure actual outcomes, rather than activity completion. Validate that documented plans actually work under healthcare-specific constraints through progressively realistic testing.

Chaos engineering systematically injects controlled failures into production systems to test their resilience under realistic conditions. Netflix pioneered this approach [35] — terminating random servers during business hours to validate service continuity despite component failures. The insight transfers directly: rather than assuming documented business continuity plans work, deliberately create realistic failure scenarios and observe whether staff execute documented procedures under pressure.

Ethical Guardrails for Healthcare Chaos Testing*: Unlike generic IT systems, where failures affect only the user experience, healthcare chaos engineering raises patient safety considerations. Phase 1-2 testing must occur during low-census periods with clinical leadership approval, explicit patient safety monitoring, and immediate abort protocols. Phase 3-4 multi-system failures require ethics committee review, particularly when testing emergency department or critical care dependencies. The progressive approach ensures staff demonstrate capability at each level before advancing to realistic scenarios that could affect care delivery.*

NHS EPRR Framework already requires business continuity testing — current limitation focuses on operational emergencies (mass casualties, severe weather), not cyber resilience (system failures, manual procedure execution during IT outages).

Minimal design: Extend EPRR progressive testing methodology to cyber resilience through four phases, validating whether documented business continuity plans work when critical IT systems fail.

- **Phase 1 Structured Drills** test documentation: single system failure, announced scenario, controlled conditions, off-peak hours (example: Saturday 10 AM disable ePMA for 2 hours, execute documented manual prescribing procedures). Success criteria: staff locate procedures within 10 minutes, manual prescriptions completed, maintaining patient safety, and drug allergy information accessible via alternative routes.
- **Phase 2 Unannounced Exercises** test execution without preparation: random failure injection, realistic hours (evening shift, clinic hours), no advance warning. Success criteria: staff identify failure within 5 minutes, emergency procedures are activated without external prompting, and alternative workflows are executed, maintaining care quality.
- **Phase 3 Multi-System Failures** test cascading impact: compound failures, peak hours, including dependencies (PAS + PACS + email simultaneously during morning clinic). Success criteria: patient registration continues via manual procedures, diagnostic images are accessible via alternative routes, and clinical decisions are made with available information documented appropriately.
- **Phase 4 Coordinated Regional Exercises** test staff redistribution: multiple trust failures simultaneously within the same ICS, patient diversion protocols, staff redeployment, and regional coordination. Success criteria: patients diverted successfully per EPRR protocols, clinical staff redistribute, maintaining care quality, and IT staff remain for recovery coordination.

Testing is event-driven (pulled by system changes) not calendar-driven (scheduled annually regardless of changes): new system deployment triggers Phase 2 test within 30 days validating manual backup procedures before system becomes critical dependency, process change triggers Phase 1 test within 60 days validating staff execute modified procedures, emerging threat class (ransomware targeting healthcare) triggers Phase 3 scenario-specific

test when intelligence identifies NHS-specific targeting, major incident anywhere in NHS (Synnovis) triggers Phase 4 coordinated regional exercise within 90 days incorporating lessons learned, significant WIP completion triggers appropriate phase test validating completed capability before freeing WIP slot.

Evidence base: This implements CAF [22] Objective D capabilities whilst revealing principles-based framework limitations — CAF D1.c requires organisations to "test response and recovery plans" but doesn't specify testing realism levels. Progressive phases (announced drills to unannounced multi-system failures to coordinated regional exercises) provide concrete CAF D1.c implementation, revealing whether documented capabilities actually function under realistic incident conditions. Trusts can demonstrate CAF D1 compliance through annual desktop exercises, whilst failing catastrophically during real incidents if testing doesn't replicate realistic constraints.

Board checks and metrics:

- **Metric:** Capability validation rate (documented procedures tested under realistic conditions annually / total business continuity procedures), Target 100% critical procedures tested annually across progressive phases, with identified gaps remediated and retested before marking complete.
- **Cadence:** Quarterly Board Cyber Resilience Report showing: Compliance status (DSPT Standards Met, CareCERT alerts remediated, board training completed, policies updated), Capability validation (chaos engineering tests conducted with clinical impact, response times against targets, gaps identified with remediation actions, follow-up retests validating fixes), Decision latency metrics (average times for threat identification to response with comparison against targets), Continuous improvement (procedural gaps identified and remediated through testing).

How does this address overconfidence from compliance metrics? The overconfidence mechanism operates through several mechanisms, including external validation (DSPT reviewed by NHS England), peer comparison (trust status visible across NHS), regulatory requirements (contract requirements, CQC consideration), and board assurance (ARAC reviews DSPT status, reports to full board). This creates confidence divorced from capability — "We're DSPT compliant" becomes synonymous with "we're secure" even when documented procedures haven't been validated under realistic conditions.

Chaos engineering breaks this pattern:

1. Testing reveals gaps between documentation and reality — procedures work on paper but fail during unannounced drills, staff are unable to locate materials/access information/execute workflows, dependencies are not understood until systems actually fail, the board sees "Standards Met" coexist with "unable to prescribe safely during ePMA failure".
2. Outcome metrics challenge activity metrics — activity shows "business continuity plan documented, reviewed annually, staff trained" whilst outcome shows "Phase 2 test revealed 15-minute delay to critical drug allergy information," board must reconcile documented plan with failed execution.
3. Continuous validation prevents compliance complacency — cannot remain "compliant" indefinitely whilst capabilities degrade, each system change/threat emergence/major incident triggers new testing, passing last year's DSPT doesn't guarantee current preparedness.
4. Board oversight focuses on actual resilience — quarterly reports show test results/gaps identified/remediation status, board asks "Can we maintain care if systems fail?" not "Are we compliant?"

Why outcome validation resists metric gaming: Prior research [36] demonstrates that quantitative security metrics systematically produce gaming behaviour — vulnerability counts inflate trivial issues, patching percentages exclude critical systems from denominators, and incident closure rates incentivise artificial subdivision. Chaos engineering counters this through unfakeable evidence: staff cannot "document" their way through unannounced system failures (either successfully execute manual procedures or don't), multi-system failure tests reveal actual coordination gaps that cannot be obscured through compliance reporting, and patient safety incidents during realistic exercises provide observable outcomes independent of self-assessment. This architectural property is critical for board oversight — when boards ask, "Can we maintain care when systems fail?", they receive evidence that cannot be gamed through process documentation.

5.6 What the Board Sees Each Month

Board cyber resilience oversight consolidates into a concise monthly dashboard, supplementing quarterly ARAC deep dives. Four artefacts provide comprehensive visibility without overwhelming detail:

1. **Top 5 KEV Vulnerability Alignment** (from Section 5.2): Single-page view showing current CISA KEV priorities, which WIP initiatives address them, and exposure windows for gaps. The board instantly sees whether the portfolio aligns with the current threat landscape or whether documented plans addressing yesterday's threats, whilst today's vulnerabilities remain undefended.
2. **WIP Portfolio Status** (from Section 5.3): Table showing five concurrent initiatives with completion percentage, tangible outcomes achieved this period, Top 5 KEV coverage, and next quarter decision (Complete/Continue/Abandon with justification). The board enforces completion discipline without micromanaging technical details.
3. **Decision Latency Metrics** (from Section 5.4): Timeline chart showing CRITICAL/HIGH threats this period from identification through authorised response with comparison against 48-hour/1-week targets. The board assesses whether delegation protocols facilitate rapid responses or if governance structures introduce systematic delays.
4. **Capability Validation Results** (from Section 5.5): Summary of chaos engineering tests conducted, gaps identified, remediation progress, and retests validating fixes. The board sees evidence that documented procedures actually work under pressure, not just documented compliance assertions.

This consolidated view enables boards to discharge cyber resilience oversight through four simple questions: "Does our current portfolio address current threats?" (alignment), "Are we completing initiatives or accumulating commitments?" (WIP discipline), "Can we respond to emergencies within acceptable timeframes?" (decision flow), "Do our documented plans actually work when needed?" (capability validation).

5.7 Integration Across Mechanisms

The four practices collectively address anchoring, overconfidence and temporal mismatch through complementary mechanisms:

Anchoring is challenged by:

- **Visualisation** providing external intelligence, competing with documented baselines — when boards see CISA adding vulnerabilities to KEV whilst the WIP portfolio doesn't address them, baseline adequacy becomes questionable.
- **WIP limits** forcing quarterly revalidation, preventing indefinite plan anchoring — cannot maintain "2014 migration plan adequate" indefinitely when forced to justify continuation against current threats quarterly.

Overconfidence is challenged by:

- **Feedback loops** revealing gaps in documented procedures through chaos engineering — "procedures documented" proven insufficient when Phase 2 unannounced test reveals staff unable to execute them.
- **Outcome metrics** replacing compliance documentation with capability validation — board receives evidence that procedures work under pressure, not just attestations that procedures exist.

Temporal mismatch is addressed by:

- **Flow optimisation** enabling event-driven decision-making, bypassing scheduled cycles — CRITICAL threats trigger CISO emergency authority within 48 hours, not awaiting quarterly ARAC.
- **Pull systems** triggering testing by changes, not calendar schedules — new system deployment triggers validation testing within 30 days, not an annual schedule, regardless of changes.

Together, these mechanisms create evolutionary governance transformation within existing NHS structures — respecting trust board authority, ARAC oversight, EPRR frameworks, and DSPT requirements whilst addressing the architectural vulnerabilities that enabled WannaCry. The approach doesn't replace existing governance but enhances it through

decision-flow optimisation, forcing functions challenging organisational inertia, and outcome validation, supplementing process compliance.

6 Implementation: Board Oversight Implications

6.1 Strategic Governance Questions

Board oversight requires asking questions that reveal governance effectiveness rather than measuring activity completion:

- **Portfolio Alignment**: "Does our current portfolio address current threats?" Risk registers document baselines and mitigation plans — essential governance functions. However, documentation can create anchoring when circumstances change [20]. Boards should supplement risk registers with forcing functions that trigger fundamental assumption review when threat intelligence changes.
- **Capability Validation**: "Can we prove our plans work?" DSPT provides valuable standardisation across trusts. However, compliance metrics measure process documentation rather than capability under pressure. The gap between "procedures that exist" and "procedures that work during incidents" may not be visible through compliance frameworks. Boards should supplement compliance metrics with progressive chaos engineering to validate continuity plans under realistic conditions.
- **Decision Velocity**: "Can we respond to emergencies within acceptable timeframes?" Enhanced cybersecurity oversight through dedicated committees and regular reporting demonstrates organisational learning. However, boards operate on strategic timescales (annual plans, quarterly reviews) whilst cyber threats evolve tactically (daily disclosures, weekly weaponisation). The 58-day WannaCry window exceeded regular governance cycles despite available information [19]. Boards should design delegation frameworks enabling emergency response outside scheduled cycles whilst maintaining appropriate oversight.
- **Federated Coordination**: "How do we act collectively on cross-cutting threats?" NHS federated structure provides local autonomy and clinical responsiveness — appropriate for healthcare delivery. However, this creates coordination challenges for threats that require coordinated response across all independent trusts. No single authority could mandate WannaCry patching, yet uncoordinated local decisions created system-wide

vulnerability. Boards should design coordination mechanisms enabling collective action on cross-cutting threats whilst preserving appropriate local autonomy.

6.2 Implementation Realities

The target architecture requires significant investment and organisational change. However, cost-benefit analysis justifies this investment: the £2M coordinated patching cost versus £92M WannaCry recovery represents a 46:1 failure ratio [1, 2]; Synnovis added £32.7M [6]. Preventable high-cost incident patterns suggest incremental investment in current governance approaches may cost more than fundamental architectural reform. The question isn't whether to invest but whether to invest proactively in structural solutions or continue paying reactively for incident recovery.

6.3 Progressive Implementation

Full multi-framework synthesis does not need to be implemented simultaneously across all NHS organisations. The implementation pathway proposed in Section 5 enables progressive rollout: pilot at a subset of trusts, validate effectiveness through outcome metrics, expand based on demonstrated results. This reduces risk and enables learning whilst pursuing architectural reform.

Boards might prioritise: external threat intelligence integration (enabling assumption-challenging), outcome-validation mechanisms (supplementing process compliance), and work-in-progress limits (improving completion rates for existing initiatives). These are individually valuable whilst laying groundwork for fuller synthesis.

Three-Phase Pilot (36 months, £62M across 236 trusts):

- **Phase 1** (Months 1-12): 5 volunteer trusts implement the full framework.
- **Phase 2** (Months 13-24): Scale to 40 trusts (ICS-level coordination).
- **Phase 3** (Months 25-36): National rollout with regional variations.

Political and Operational Constraints: *Progressive stress-testing and architecture experimentation in the NHS operate under tight fiscal control and political visibility. Any resilience pilot must therefore demonstrate proportionality, avoid interference with clinical*

delivery, and anticipate reputational sensitivity in the wake of prior incidents. Embedding such pilots within existing assurance cycles (e.g., DSPT or CAF reviews) mitigates these concerns without requiring additional funding lines.

6.4 Measuring Governance Effectiveness

Current board reporting often focuses on process metrics (assessments completed, policies updated, training delivered) — easy to measure and demonstrate activity. However, these may not correlate with actual resilience under pressure.

Outcome-focused metrics revealing whether governance actually works:

- **Decision latency**: Critical threat identification to coordinated response (target <48 hours).
- **Continuity validation**: Plans are tested under realistic conditions annually (target 100% critical procedures).
- **Resource optimisation**: WIP limits (target 50% reduction in initiative completion time).
- **System-wide visibility**: Visual management enabling pattern detection (target detection when 3+ organisations report related indicators).

6.5 Executive Responsibility

Implementation details — which frameworks to integrate, how to redistribute staff, when to schedule chaos tests, and how to coordinate across trusts — remain the responsibility of the executives. They understand local contexts, resource constraints, clinical priorities, and operational realities that this analysis cannot capture.

Board's role: ensure executives have governance structures, resources, and authority to succeed. If governance architecture systematically amplifies cognitive biases, executives cannot succeed solely through better execution. Structural problems require structural solutions — boards are uniquely positioned to mandate and resource these.

7 Broader Implications: AI Governance

Healthcare organisations deploying AI systems face governance patterns identical to those that enabled WannaCry, but with exponentially compressed timescales. Current AI frameworks (EU AI Act, NIST AI RMF, UK sectoral model) rely on pre-incident risk categorisation and distributed accountability — the same structures that failed during WannaCry.

Detailed examination of cognitive bias amplification in AI governance and flow-constrained risk management for operational AI systems appears in related work [14].

The core insight remains: governance frameworks that do not account architecturally for cognitive limitations systematically fail under rapid change and federated decision-making, regardless of domain.

8 Conclusion

This framework enables non-executive directors to discharge cyber-resilience oversight without encroaching on management functions. The four Kanban-based mechanisms provide boards with specific governance capabilities: **(1) challenge without micromanagement** through external threat intelligence (CISA KEV integration) that independently verifies whether management's portfolio aligns with current threats; **(2) outcome validation beyond compliance** through progressive chaos engineering that reveals whether documented procedures actually work under pressure, supplementing DSPT process metrics with capability evidence; **(3) emergency delegation within oversight** through threshold-based protocols enabling CISO response to CRITICAL threats within 48 hours whilst maintaining board accountability via immediate notification and 7-day ratification; and **(4) portfolio governance through WIP limits** via quarterly forcing functions requiring management to complete, continue with justification, or abandon initiatives — preventing indefinite "in progress" accumulation. NEDs should request quarterly Board Cyber Resilience Reports showing decision latency metrics, capability validation results, and WIP portfolio alignment with current threat intelligence.

Strategic imperative: The 10-Year Health Plan commits the NHS to comprehensive digital transformation — integrated records, AI diagnostics, remote monitoring, and prevention-focused care. These ambitions depend entirely on cyber resilience: WannaCry's £92 million cost and 19,494 cancelled appointments demonstrated how governance failures directly undermine patient care and strategic objectives. Seven years later, Synnovis added £32.7 million despite enhanced DSPT requirements, NIS Regulations, and the 2023-2030 Cyber Strategy — proving that conventional solutions have not prevented pattern recurrence.

The Core Finding

WannaCry wasn't caused by insufficient resources or expertise — NHS cybersecurity professionals identified the threat, patches were available, and infrastructure existed. What failed was decision-making, systematically embedded in governance architecture through:

- **Anchoring**: Documented 2014 migration plans became organisationally "true" despite changed circumstances.
- **Overconfidence**: Compliance metrics generated false confidence whilst capabilities remained untested.
- **Temporal mismatch**: Annual governance cycles could not respond to threats evolving within 58 days.

What the Evidence Shows

Systematic alternatives analysis across 80 organisations demonstrates that cognitive bias amplification provides superior explanatory power over conventional explanations. This hypothesis accounts for systematic failure across diverse contexts, the observed decision lag, cost multipliers between prevention and recovery, and the persistence of patterns despite implemented solutions (see Section 4.7 and Appendix C for detailed analysis).

The Architectural Solution

Four Kanban-based mechanisms address structural vulnerabilities:

1. **Visualise work**: External threat intelligence (CISA KEV integration) challenges anchoring on documented baselines — boards see real-time vulnerability exposure across NHS trusts.
2. **Limit work-in-progress**: Maximum five concurrent initiatives with quarterly forcing functions to prevent indefinite "in progress" accumulation.
3. **Manage flow**: Threshold-based delegation enables 48-hour CRITICAL threat response versus 100+ day governance cycles.
4. **Implement feedback loops**: Progressive chaos engineering validates actual capabilities under pressure, replacing compliance documentation with outcome evidence.

What Healthcare Boards Should Do

- Supplement risk registers with forcing functions that challenge assumptions when threat landscapes change.
- Validate capabilities through realistic testing, not just document procedures.
- Implement delegation frameworks enabling emergency response within governance oversight.
- Measure decision latency and validated capabilities, not activity completion.

What Policy Makers Should Do

- Mandate bias correction mechanisms addressing structural causes, not add more compliance requirements.
- Require external intelligence integration, outcome validation, and dynamic reassessment.
- Apply architectural lessons across federated critical infrastructure where similar patterns exist.

Research Limitations and Future Directions

Working Paper Stage: This analysis has several limitations that warrant acknowledgement and will be addressed in the planned empirical extension:

- **Data availability**: The study relies on retrospective analysis using publicly available evidence from NAO reports and parliamentary testimony. Internal board

deliberations, informal communications, and context-specific constraints may not be fully captured in official records. The analytical framework in Table 1 uses estimated distributions based on NHS trust population characteristics rather than verified organisational characteristics of each affected trust. Whilst the NAO's finding of "no clear relationship" between organisational factors and infection supports the framework's logic, direct verification requires trust-level data currently not publicly enumerated.

- **Methodological constraints**: The research focuses specifically on UK healthcare governance structures. The NHS federated model creates particular coordination challenges that may not generalise to hierarchical healthcare systems in other countries. Claims about cognitive bias amplification require validation across different governance architectures before asserting universal applicability. The alternatives analysis demonstrates that cognitive bias amplification provides superior explanatory power compared to conventional explanations, but correlation does not establish definitive causality without experimental validation.
- **Temporal scope**: The analysis examines WannaCry (2017) and Synnovis (2024) as primary evidence points. Pattern recurrence across seven years supports the hypothesis of structural rather than transient vulnerabilities, but longer-term longitudinal tracking would strengthen causal claims. Post-WannaCry governance changes (NIS Regulations, enhanced DSPT, CAF evolution) have had limited time to demonstrate effectiveness, and claims about their inadequacy require ongoing monitoring.
- **Proposed target architecture validation**: The Kanban-based governance mechanisms proposed in Section 5 are theoretically grounded and draw on manufacturing and software engineering precedents, but require empirical validation through controlled pilots within NHS governance constraints. Cost estimates, implementation timelines, and benefit projections are based on scenario analysis and comparative data rather than direct NHS implementation experience. Progressive chaos engineering, WIP limits, and visual management systems need operational testing to identify unforeseen implementation barriers.

Future Research Directions — Empirical Extension: The planned empirical research phase will address these limitations through:

1. **Primary data collection**:
    - Structured interviews with 15-20 NHS Trust CISOs and CIOs examining decision-making processes during WannaCry and subsequent governance evolution.
    - Freedom of Information requests to NHS trusts for organisational characteristics (budgets, staff sizes, IT maturity indicators, compliance certifications) of the 80 affected trusts, enabling direct verification of the distributions in Table 1.
    - Case study analysis of 3-5 trusts with diverse characteristics, accessing board meeting minutes, risk register evolution, and decision timelines during the 2014-2017 period.
    - Comparative analysis of trusts that deployed MS17-010 patches rapidly versus those that delayed, documenting governance architecture differences.
2. **Quantitative validation**:
    - Statistical analysis testing correlations between organisational factors (budget, compliance status, technology maturity, geographic location) and infection outcomes using verified trust-level data.
    - Decision latency measurement across governance layers using timestamped evidence from trust records (threat identification → CISO assessment → board authorisation → implementation).
    - Pilot implementation of proposed mechanisms (WIP limits, visual management, chaos engineering) at 2-3 volunteer trusts, measuring decision latency reduction and capability validation improvements.
3. **Cross-domain extension**:
    - Apply architectural analysis to AI governance frameworks currently being deployed across NHS trusts, testing whether identical bias amplification patterns emerge at compressed timescales
    - Comparative analysis with other federated critical infrastructure sectors (e.g., higher education, local government), assessing whether similar governance patterns yield similar vulnerabilities.

- Longitudinal tracking over 5-10 years, assessing whether bias correction mechanisms prevent incident recurrence or whether new architectural vulnerabilities emerge.

4. **Methodological triangulation**:
    - Combine publicly available investigation reports with internal governance documentation accessed through research agreements with participating trusts.
    - Triangulate official accounts (NAO reports, parliamentary testimony) with first-person narratives from decision-makers involved in the incident.
    - Validate theoretical predictions from cognitive psychology literature against observed decision-making patterns in real organisational contexts.

The Pattern We Must Break

Synnovis 2024 (£32.7 million) demonstrates that the pattern persists despite conventional solutions. The same governance architectures now being applied to AI deployment face identical structural vulnerabilities.

The evidence is quantified. The solutions exist. The implementation pathway is defined with realistic costs and validated mechanisms. What happens next depends on whether we learn from WannaCry or repeat it.

BOARD OVERSIGHT: FIVE CRITICAL QUESTIONS

1. **"Show me our Top 5 CISA KEV vulnerabilities and which WIP initiatives address them"** — Tests portfolio-threat alignment without micromanaging.
2. **"Walk me through our last CRITICAL alert — who knew what, when, and how fast we acted?"** — Reveals decision latency patterns.
3. **"When did we last test [critical procedure] under realistic conditions, and what failed?"** — Distinguishes documented compliance from validated capability.
4. **"What did we stop doing to start [new initiative]?"** — Forces explicit WIP trade-offs, counters indefinite accumulation.
5. **"Has this 2-year-old migration plan been revalidated against today's threat landscape?"** — Challenges anchoring on documented baselines.

9 Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The author, a non-native English speaker, used Claude (Anthropic) for language refinement, argument structure, and literature synthesis assistance. Specific sections that used AI assistance include portions of the Current State section, transition sentences throughout, and overall readability improvements. All substantive intellectual contributions, research design, analysis, and conclusions remain entirely the work of the author.

10 Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

11 Declaration of Competing Interest

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Department of Health and Social Care, "Securing cyber resilience in health and care. Progress update October 2018", DHSC, London, 2018. [Online]. Available: https://assets.publishing.service.gov.uk/media/5bbe1250ed915d732b99254c/securing-cyber-resilience-in-health-and-care-september-2018-update.pdf. [Accessed: 22 Oct 2025].

2. Ponemon Institute, "Costs and consequences of gaps in vulnerability response", Ponemon Institute LLC, Traverse City, MI, USA, 2019. [Online]. Available: https://www.servicenow.com/premium/resource-center/analyst-report/ponemon-vulnerability-survey.html. [Accessed: 22 Oct 2025].

3. Department of Health & Social Care, "10 year health plan for England: Fit for the future", DHSC, 2025. [Online]. Available: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future. [Accessed.

4. UK Government, "The network and information systems regulations 2018", *SI 2018/506*, The Stationery Office, London, 2018.

5. Department of Health and Social Care, "A cyber resilient health and adult social care system in England: Cyber security strategy to 2030", DHSC, London, 2023. [Online]. Available: https://www.gov.uk/government/publications/cyber-security-strategy-for-health-and-social-care-2023-to-2030/a-cyber-resilient-health-and-adult-social-care-system-in-england-cyber-security-strategy-to-2030. [Accessed: 22 Oct 2025].

6. J. Sollof, "Cyber attack cost Synnovis estimated £32.7m in 2024", *Digital Health*, 2025. [Online]. Available: https://www.digitalhealth.net/2025/01/cyber-attack-cost-synnovis-estimated-32-7m-in-2024/ [Accessed: 22 Oct 2025].

7. Department for Science, Innovation and Technology, "Second post-implementation review of the network and information systems regulations 2018", Department for Science, Innovation and Technology, 2022. [Online]. Available: https://www.gov.uk/government/publications/second-post-implementation-review-of-the-network-and-information-systems-regulations-2018/second-post-implementation-review-of-the-network-and-information-systems-regulations-2018. [Accessed: 22 Oct 2025].

8. R.H. Thaler and C.R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT, USA: Yale University Press, 2008.

9. A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases", *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974. doi: 10.1126/science.185.4157.1124.

10. The Open Group, "The TOGAF® standard, 10th edition", Reading, UK, 2025. [Online]. Available: https://publications.opengroup.org/standards/togaf/c220. [Accessed 22 Oct 2025].

11. D.J. Anderson, *Kanban: Successful evolutionary change for your technology business*. Sequim, WA, USA: Blue Hole Press, 2010.

12. E. Hollnagel, *Safety-ii in practice: Developing the resilience potentials*. London: Routledge, 2018.

13. N. Maslej, et al., "Artificial intelligence index report 2025", Institute for Human-Centered AI, Stanford University, Stanford, CA, USA, 2025. [Online]. Available: https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf. [Accessed: 13 Sep 2025].

14. V. Shabad, "Cognitive blind spots in security frameworks: From cybersecurity to AI governance", SSRN Working Paper No. 5525340, Oct 2025. [Online]. Available: https://dx.doi.org/10.2139/ssrn.5525340. [Accessed: 22 Oct 2025].

15. V. Shabad, "Flow-constrained risk management for operational technology security: A multi-criteria framework for critical infrastructure", SSRN Working Paper No. 5389934, Aug 2025. [Online]. Available: https://dx.doi.org/10.2139/ssrn.5389934. [Accessed: 22 Oct 2025].

16. R. Thorlby, et al., "The francis report: One year on — the response of acute trusts in England", The Nuffield Trust, 2014. [Online]. Available: https://www.nuffieldtrust.org.uk/sites/default/files/2017-01/francis-report-one-year-on-web-final.pdf. [Accessed: 22 Oct 2025].

17. NHS England, "Well-led framework", 2024. [Online]. Available: https://www.england.nhs.uk/well-led-framework/. [Accessed: 22 Oct 2025].

18. Department of Health, "Information security management: NHS code of practice", London, 2007, Rep. 7974, 2007. [Online]. Available: https://assets.publishing.service.gov.uk/media/5a7c7c31e5274a559005a304/Information_Security_Management_-_NHS_Code_of_Practice.pdf. [Accessed 22 Oct 2025].

19. National Audit Office, "Investigation: WannaCry cyber attack and the NHS", HC 414, NAO, London, 2018. [Online]. Available: https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf. [Accessed: 13 Sep 2025].

20. UK Parliament, House of Commons, Public Accounts Committee, "Oral evidence: Cyber-attack on the NHS", House of Commons, London, UK, 2018. [Online]. Available: https://committees.parliament.uk/oralevidence/10786/pdf/. [Accessed: 22 Oct 2025].

21. National Protective Security Authority, "Critical national infrastructure", 2025. [Online]. Available: https://www.npsa.gov.uk/about-npsa/critical-national-infrastructure [Accessed: 22 Oct 2025].

22. National Cyber Security Centre, "Cyber assessment framework v4.0", 2025. [Online]. Available: https://www.ncsc.gov.uk/files/NCSC-Cyber-Assessment-Framework-4.0.pdf. [Accessed: 22 Oct 2025].

23. D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk", *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979. doi: 10.2307/1914185.

24. NHS England, "Cyber assurance service," https://digital.nhs.uk/cyber-and-data-security/managing-security/cyber-assurance-service. [Accessed: 25 Oct 2025.

25. Department for Science, Innovation and Technology, "Cyber security and resilience policy statement", DSIT, London, 2025. [Online]. Available: https://www.gov.uk/government/publications/cyber-security-and-resilience-bill-policy-statement/cyber-security-and-resilience-bill-policy-statement. [Accessed: 22 Oct 2025].

26. D. Berwick, "A promise to learn – a commitment to act: Improving the safety of patients in England", National Advisory Group on the Safety of Patients in England, 2013. [Online]. Available: https://assets.publishing.service.gov.uk/media/5a7cc74540f0b6629523bc31/Berwick_Report.pdf. [Accessed: 22 Oct 2025].

27. R. Collier, "NHS ransomware attack spreads worldwide", *Canadian Medical Association Journal*, vol. 189, no. 22, pp. E786–E787, 2017. doi: 10.1503/cmaj.1095434.

28. D. Kahneman, *Thinking, fast and slow*. New York, USA: Farrar, Straus and Giroux, 2011.

29. G. Barre, et al., “Towards understanding cognitive biases in cybersecurity governance”, in *Proc. 38th Bled eConference*, Bled, Slovenia, 2025, pp. 737–744. doi: 10.18690/um.fov.4.2025.46.

30. J.T. Reason, *Human error*. Cambridge, UK: Cambridge University Press, 1991.

31. S. Dekker, *Drift into failure from hunting broken components to understanding complex systems*. Boca Raton, FL, USA: Taylor & Francis Group, 2011.

32. NHS England, “Board governance”, n. d. [Online]. Available: https://www.england.nhs.uk/about/nhs-england-board/board-governance/ [Accessed: 20 October 2025].

33. NHS England, "NHS emergency preparedness, resilience and response framework: Version 3”, London, 2022. [Online]. Available: https://www.england.nhs.uk/wp-content/uploads/2022/07/B0900-NHS-Emergency-Preparedness-Resilience-and-Response-Framework-version-3.pdf. [Accessed: 22 Oct 2025].

34. Cybersecurity and Infrastructure Security Agency, "The U.S. Cybersecurity and infrastructure security agency (CISA) added five new vulnerabilities to its known exploited vulnerabilities (KEV) catalog, based on evidence of active exploitation”. [Online]. Available: https://www.cisa.gov/news-events/alerts/2025/10/20/cisa-adds-five-known-exploited-vulnerabilities-catalog. [Accessed: 22 Oct 2025].

35. A. Basiri, et al., “Chaos engineering”, *IEEE software*, vol. 33, no. 3, pp. 35–41, 2016. doi: 10.1109/MS.2016.60.

36. V. Shabad, “From crisis to control - a retrospective enterprise architecture analysis of cybersecurity transformation“, SSRN Working Paper No. 5393328, Aug 2025. [Online]. Available: https://dx.doi.org/10.2139/ssrn.5393328. [Accessed: 22 Oct 2025].

Appendix A: Acronyms

ARAC — Audit and Risk Assurance Committee

ADM — Architecture Development Method

CAF — Cyber Assessment Framework

CareCERT — Care Computer Emergency Response Team

CISA — Cybersecurity & Infrastructure Security Agency

CNI — Critical National Infrastructure

CSIRT — Computer Security Incident Response Team

DHSC — Department of Health and Social Care

DSPT — Data Security and Protection Toolkit

EOL — End-of-Life

ePMA — Electronic Prescribing and Medicines Administration

EPR — Electronic Patient Record

EPRR — Emergency Preparedness Resilience and Response

HMG — Her Majesty's Government

HSJ — Health Service Journal

ICS — Integrated Care System

ICB — Integrated Care Board

IG — Information Governance

IGP — Indicator of Good Practice

ISP — Internet Service Provider

ISO — International Organisation for Standardization

NAO — National Audit Office

NED — Non-Executive Director

NHS — National Health Service

NIS — Network and Information Systems

NIST — National Institute of Standards and Technology

NPSA — National Protective Security Authority

OES — Operator of Essential Services

PACS — Picture Archiving and Communication System

PAS — Patient Administration System

TOGAF — The Open Group Architecture Framework

WHO — World Health Organisation

WIP — Work-In-Progress

Appendix B: Practical Governance Oversight Toolkit for Non-Executive Directors

Purpose: This toolkit translates the architectural analysis (Sections 4-5) and cognitive bias theory into actionable board oversight mechanisms. NEDs can use these questions to discharge cyber resilience responsibilities without micromanaging technical operations.

Strategic Context: The 10-Year Health Plan positions digital transformation as central to NHS sustainability — prevention, community care, and efficiency all depend on resilient technology. These oversight questions connect directly to strategic delivery: "Does our portfolio address current threats?" ensures cybersecurity investments protect the digital infrastructure the Plan requires. "Can we prove our plans work?" validates that transformation initiatives achieve actual resilience, not just documented compliance.

CRITICAL CAVEAT: These questions provoke narrative decision-making evidence, not quantitative metrics. The moment boards track "percentage compliance" with these targets, they become gameable. The purpose: challenge management on decision quality, not measure activity completion.

The distinction matters: Asking "Walk me through the last critical alert — who knew what, when?" reveals decision-flow patterns. Creating a dashboard showing "Decision latency: 98% within 24h target" incentivises gaming without improving actual response capability.

Table 3. Board Governance Oversight Framework

Governance Domain	Board Oversight Question	Evidence Type	Red Flags
Decision Flow	"Walk me through the last critical alert. Who knew what, when?"	Narrative timeline with decision points	Gaps >24h unexplained, multiple handoffs
Capability Validation	"Show me the last time this backup procedure actually worked"	Chaos engineering test results with identified gaps	No tests conducted, tests showing only successes

Portfolio Alignment	"Why the initiative X is still consuming resources when threat landscape changed?"	Management justification against current CISA KEV	Answers referencing old baselines, no revalidation
Bias Detection	"What did we stop doing to start initiative Y?"	WIP decisions with explicit trade-offs documented	Everything marked "in progress", no completions

Interpretation Guide

Healthy Governance Patterns:

- Management can explain decision-making for recent incidents with evidence.
- WIP initiatives complete or get explicitly abandoned with documented rationale.
- Chaos tests reveal gaps that subsequently get addressed and retested.
- Portfolio composition changes when threat landscape shifts (with justification).

Warning Signs:

- Latency increasing or exceeding targets.
- WIP accumulating, initiatives ageing>12 months.
- Compliance is improving, but tests are failing.
- Portfolio unchanged despite new threats.

Board Role

Ask: *"Does our portfolio address current threats, and can we prove our plans work?"*

Do NOT: Direct technical decisions, specify patches, micromanage operations.

DO: Challenge portfolio-threat misalignment, require capability evidence, and assess governance effectiveness.

Adoption Path

- **Q1**: Establish baseline measurements.
- **Q2**: Implement latency and WIP tracking.

- **Q3**: Conduct the first capability test.
- **Q4**: Full framework with trend analysis.

This framework supplements existing DSPT and EPRR requirements by adding outcome validation to compliance documentation.

Appendix C: Methodological Foundation for Prevention-to-Recovery Cost Ratio

Purpose and Scope

This appendix documents the calculation methodology for the "46:1 prevention-to-recovery cost ratio" cited throughout the paper. This ratio quantifies the financial consequence of governance architecture failures, establishing that prevention was economically trivial whilst recovery proved catastrophically expensive.

What this ratio represents: Cost multiplier resulting from governance architecture failure enabling WannaCry despite available patches, qualified staff, and documented processes.

What this ratio does NOT represent: Business case for implementation costs of proposed governance mechanisms. Determining whether Kanban-based governance (Section 5) and cognitive bias workshops (Appendix D) achieve cost-effective prevention requires empirical validation through progressive pilots, rather than projections from historical incident costs.

Ponemon Institute Baseline Data

Ponemon Institute's 2019 study surveyed 2,900 IT security professionals across eight countries [2]:

- Average annual vulnerability management expenditure: $1.4M (≈£1.1M)
- Average time to patch critical vulnerabilities: 16 days
- Coordination delay across organisational silos: 12 days
- **Critical finding: 60% of breaches occurred where patches were available but not applied**

Weekly resource allocation (average organisation):

- Patching applications and systems: 206 hours/week
- Monitoring systems: 139 hours/week
- Total: 443 hours/week at $62.50/hour fully-loaded

NHS WannaCry Context

National Audit Office documentation [19]:

- Total disruption cost: **£92 million**
- Affected: ~80 NHS trusts (34% of 236)
- Vulnerable systems: ~5% of the NHS estate on unsupported Windows XP
- Warning period: **58 days** between patch release (March 14) and attack (May 12, 2017)
- **No formal mechanism to ensure patch compliance across trusts**

Coordinated Emergency Patching Cost Estimate

Resource requirements:

- Per-trust emergency patching (80 trusts): 220 hours each = 17,600 hours
- Central NHS Digital coordination: 1,150 hours
- Total: 18,750 hours @ £50/hour = £937,500

Additional costs:

- Emergency vendor support: £150,000
- Testing infrastructure: £50,000
- Communications: £30,000
- Contingency (10%): £116,750

Total preventive cost: £1.28M (conservatively rounded to **£2M** in paper)

Table 4. Cost-Benefit Ratio

Category	Cost	Ratio
Preventive (emergency patching)	£2M	—
Reactive (actual impact)	£92M	—
Reactive:Preventive multiplier	—	**≈46:1**

This **46:1 ratio** represents a conservative estimate — actual preventive costs could have been lower (optimistic scenario: 115:1 ratio), but we maintain the conservative figure throughout the paper.

Applicability of Ponemon Data

The Ponemon study included UK respondents (12.5% of the sample) and public sector organisations (10% of respondents) [3], making baseline data directly applicable.

Quantifying Cognitive Bias Impact

Anchoring bias cost: NHS remained committed to the 2014 migration timeline despite a 58-day warning period. Cost of anchoring: £92M - £2M = **£90M.**

Overconfidence bias cost: Officials cited 88 on-site assessments (37% coverage) as evidence of preparedness [20]. Confidence, divorced from actual coverage, left 63% of trusts unassessed.

Temporal mismatch cost: Annual governance cycles prevented emergency reallocation of the £2M required within the 58-day window.

The **46:1 cost multiplier** directly quantifies the financial consequence of framework-embedded cognitive biases. This ratio resists metric gaming because it compares audited incident costs (£92M in documented losses, independently verified by NAO) against counterfactual prevention costs (£2M coordination estimate based on industry data). Unlike self-reported compliance metrics or process documentation counts, incident costs cannot be obscured through selective reporting or definitional manipulation. The 46:1 ratio represents a measurable governance failure — a figure that no amount of DSPT compliance status or risk register updates could explain away.

Conclusion

The 46:1 ratio is methodologically grounded in documented NAO recovery costs [19] and industry-benchmark prevention estimates [2]. It serves as a governance failure indicator that quantifies the impact of decision latency, not as a business case for specific implementation approaches.

The ratio's robustness across sensitivity scenarios (30-90x range depending on prevention cost assumptions) supports the cognitive bias amplification hypothesis: even if emergency coordination costs were 2x higher than estimated, governance architecture failure still created

a 30x cost multiplier. This magnitude justifies serious consideration of governance architecture reforms whilst acknowledging that specific implementation costs require NHS-internal data unavailable to external analysis.

Synnovis 2024 (£32.7M additional cost [6]) demonstrates pattern recurrence, suggesting the 46:1 historical ratio may be conservative if governance architecture vulnerabilities persist. However, this does NOT predict that the proposed mechanisms will achieve equivalent cost savings — that determination requires progressive pilots that measure actual implementation costs, organisational adoption challenges, and prevention effectiveness under real operational constraints.

Appendix D: Board Capability Development Through Cognitive Bias Workshops

Purpose: Successful implementation requires board-level recognition of cognitive bias patterns in their own governance decisions. Generic training is insufficient—boards must analyse documented governance history to identify anchoring, overconfidence, and temporal mismatch in trust-specific contexts.

Why This Matters: The Kanban-based governance mechanisms (Section 5) only work if boards recognise the cognitive biases they're designed to counter. Without metacognitive capability, this becomes just another compliance ritual rather than a bias-correction tool.

Three-Phase Development Pathway

Phase 1: Historical Pattern Recognition (Half-day facilitated session). Board reviews 3 years of documented decisions (risk registers, DSPT assessments, decision timelines) with an external facilitator guiding the identification of anchoring, overconfidence, and temporal-mismatch patterns. Critical success factor: psychological safety, enabling boards to acknowledge patterns as structural insight, not personnel criticism.

Phase 2: Mechanism Piloting (Quarterly sessions over 12 months). Boards apply Section 5 mechanisms to real decisions through structured reflection: Q1-2 WIP limits reveal anchoring; Q3-4 CISA KEV integration challenges portfolio assumptions; Q5-6 threshold-based delegation counters loss aversion; Q7-8 chaos engineering replaces compliance metrics with capability evidence.

Phase 3: Autonomous Self-Assessment (Ongoing). Boards demonstrate independent bias recognition without external facilitation, embedding self-correction as an ongoing governance practice. At this stage, boards should possess sufficient metacognitive capability to identify and counter cognitive biases in real-time during decision-making, requesting external facilitation only for novel patterns, major governance restructuring, or post-incident analysis.

Validation Through Outcomes, Not Activity

Inappropriate metrics (gameable):

- Workshop attendance.
- Training completion rates.

Appropriate metrics (capability evidence):

- Decision latency for CRITICAL alerts decreased from 100+ days to <48h.
- Portfolio composition changed quarterly: 40% WIP turnover vs. 10% baseline.
- Chaos tests revealed gaps; remediated and retested with documented outcomes.

Facilitator Requirements for Initial Development

Effective facilitation requires a rare capability combination:

1. **Behavioural economics application expertise**: Demonstrated ability to identify and articulate cognitive bias patterns (anchoring, loss aversion, overconfidence, bounded rationality) in organisational decision-making contexts, with practical experience applying bias-correction mechanisms in governance settings.
2. **Enterprise architecture perspective**: Ability to analyse how governance structures systematically amplify cognitive biases through architectural design choices, connecting decision-making patterns to organisational frameworks and compliance regimes.
3. **Cross-domain critical infrastructure experience**: Practical security and governance leadership across multiple regulated sectors (telecommunications, financial services, mining, or industrial environments), demonstrating understanding of governance patterns that transcend sector-specific contexts.
4. **International and cross-cultural change leadership**: Evidence of successfully influencing organisational transformation across diverse cultural contexts, particularly in situations requiring influence without direct hierarchical authority.

Time-limited engagement (12 months maximum) — the goal is to transfer self-assessment capability, not create facilitator dependency. The ultimate objective: boards internalise cognitive bias recognition, enabling non-executive directors with appropriate expertise to

provide ongoing challenge and oversight as part of normal governance duties rather than requiring external facilitation.

Integration with Governance Architecture

This workshop provides the metacognitive capability that enables Section 5 mechanisms to function: WIP limits counter anchoring, external intelligence counters overconfidence, flow optimisation counters temporal mismatch, and capability validation counters compliance complacency. Without this capability, architectural mechanisms become new compliance rituals rather than bias-correction tools.

A detailed implementation guide with session templates, reflection exercises, and facilitator materials is available by request.

Vitae

Vsevolod Shabad, CISSP, CCSP, is a Principal Enterprise Architect at BT Group and a Fellow of BCS, The Chartered Institute for IT. He has over two decades of international experience in cybersecurity and enterprise architecture, including senior leadership roles in the financial services and critical infrastructure sectors. He holds an MEng in Applied Mathematics and a PGDip in Information Security and is currently pursuing an MSc in Cybersecurity at the University of Liverpool. His research focuses on how cognitive biases and temporal misalignments affect governance frameworks across cybersecurity, operational technology, and AI. He has published preprints on SSRN and is preparing journal submissions whilst contributing to UK and EU policy discussions on AI and security governance.