sustainability

MDPI

Review

Evidence Commensurate with Claim: Grading Seismic Resilience Studies for Service-Continuity Decisions

Georgios Xekalakis

Independent Researcher, 17124 Nea Smyrni, Greece; gxekalakis@modinosvrahimis.com

Abstract

Seismic resilience studies increasingly inform municipal retrofit, restoration, and emergency-planning decisions, yet their results support claims of very different strength: a hospital that remains standing does not show that residents receive timely care. This critical review grades earthquake engineering and infrastructure studies, using an explicit codebook, by the strongest service-continuity claim their evidence supports. Most evidence stops at delivered service: requirements are rarely declared, rank stability substitutes for robustness, and outcomes are not observed after action. The weakness lies less in reporting than in reuse for decisions the studies were not designed to inform. We propose SCALE (Service-Continuity Assurance Ladder for Earthquake decisions), which ties continuity to non-compensatory requirements for each service, user group and time window; grades recommendations on an eight-level claim ladder; records residual doubts in a defeater register adapted from safety assurance cases; and links commitments to monitored triggers. Re-analysis of two published studies shows that 98.8% of the variation in modelled recovery time stems from workforce mobilisation rather than the ranked option, and that the preferred restoration strategy changes with the declared requirement. A municipal application in Cyprus identifies the evidence needed before a sheltering plan can be called adequate. SCALE allows Sendai Target D to be reported as service adequacy rather than disruption counts.

Keywords: seismic resilience; essential-service continuity; decision robustness; assurance case; non-compensatory requirements; municipal implementation; adaptive pathways; Sendai Framework; sustainable cities

1. Introduction

An earthquake can leave a hospital standing and still cut a neighbourhood off from emergency care. Between structural damage and the care a resident actually receives lie electricity and water supply, usable floor space, staff, road access and queues, and each link can fail on its own (Figure 1). In a simulated community, power outages alone caused functionality loss in up to 75% of buildings that were physically operable, for as long as 14 days [1]. The economic stakes are large even before services are counted: the 2020 European Seismic Risk Model estimates an average annual economic loss of about €7 billion from earthquakes in Europe, a figure that covers damage to buildings and their occupants, not the interruption of the services those buildings host [2]. International policy has absorbed this lesson. Target D of the Sendai Framework for Disaster Risk Reduction seeks to substantially reduce not only damage to critical infrastructure but also the disruption of basic services, among them health and educational facilities [3]. Its monitoring indicators, however, count disruptions [4]. They do not establish whether the service that remained was

Check for updates

Academic Editors: Qiang Zhang, Lei Xu, Biao Hu and Jiann-Wen Woody Ju

Received: 7 October 2026
Revised: 5 November 2026
Accepted: 6 November 2026
Published: 8 November 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

adequate, for whom, or for how long, and these are precisely the questions a municipality must answer when it decides which seismic interventions to fund.

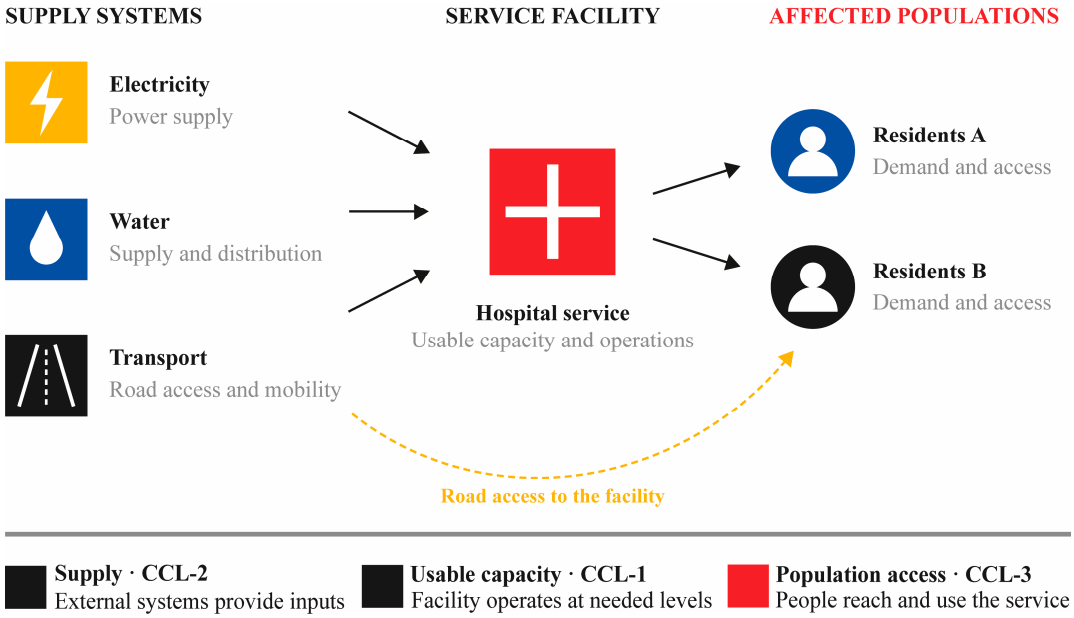

Figure 1. Conceptual dependencies of an essential service. Continuity requires supply from external systems (CCL-2), usable facility capacity (CCL-1), and population access (CCL-3); transport acts as both a supply input and a direct access constraint. Claim levels (CCL) are defined in Section 6. Not a mapped municipality.

Earthquake engineering has moved steadily towards those questions. Bruneau et al. framed seismic resilience in terms of robustness, redundancy, resourcefulness, and rapidity across technical, organisational, social, and economic dimensions [5], and Cimellaro et al. quantified it as the loss and recovery of functionality over time [6]; later reviews catalogue a wide range of resilience definitions and metrics [7]. Performance goals are shifting from life safety towards reoccupancy and functional recovery [8], supported by building-level downtime and recovery models [9,10], by frameworks that link community-level resilience goals to performance targets for individual components [11], and by guidance for designing lifeline systems for functional recovery [12,13].

Service-oriented models now trace consequences beyond the damaged asset. Community analyses couple building damage with power, water, and road serviceability [1]; healthcare studies estimate regional accessibility to hospitals across damaged bridges [14] and simulate emergency-department queues under patient surge [15]; coupled transport–healthcare models evaluate restoration strategies [16]; and open modelling environments chain hazard, physical and social models while propagating uncertainty [17]. Recovery models represent workforce shortages [18] and coordination between infrastructure operators [19], while decision-support studies rank urban-planning policies [20], risk-reduction strategies [21] and retrofit portfolios [22,23]. Recent reviews map this growing field by hazard, sector, method family and resilience phase [24–26], and others survey decision-support systems for community resilience [27].

This progress has created a new problem. The outputs that reach municipal decision makers (damage-state probabilities, functionality curves, accessibility indices, recovery times, rankings, and plans) support claims of very different strength, yet they are routinely read as if they answered the same question. Four conflations recur. First, endpoints are not equivalent: a hospital that can be reached is not necessarily a hospital that can treat the patients who arrive [14,15]. Second, aggregate indices can conceal failure: recovery

trajectories with very different timing can yield the same resilience value [28], different summary metrics can recommend different strategies [29], and a higher community score can coexist with persistent deprivation for particular groups [30,31]. Third, a stable ranking does not guarantee adequacy: robustness metrics disagree about which option is best [32], and an option can remain first in every scenario while its consequences vary several-fold. Fourth, a model-feasible intervention is not a delivered one: in the rebuild of Christchurch's pipes and roads after the 2010–2011 earthquakes, auditors found that public entities initially lacked a common understanding of the levels of service the repaired networks should deliver, and funding arrangements later took up to 19 months to confirm [33]. In each case, a claim climbs further than its evidence.

Existing reviews do not resolve this problem because they classify methods rather than grade claims. In disaster risk management, the practical effectiveness of risk, vulnerability and resilience assessment methods is seldom evaluated [34], and calls for evidence-based practice remain largely programmatic [35]. Other fields that take high-stakes decisions under uncertainty have built explicit instruments for exactly this purpose. Health and environmental policy use evidence-to-decision frameworks that separate the certainty of evidence from feasibility and acceptability [36,37]. Safety-critical engineering uses assurance cases, structured arguments that link a top-level claim to evidence through explicit reasoning [38,39], and eliminative argumentation, which builds confidence by identifying and resolving the doubts that could defeat a claim [40], while warning that numerical confidence scores are not yet trustworthy [41]. Adaptive policy pathways specify in advance the signals that should trigger a change of course [42]. To our knowledge, none of these instruments has been adapted to claims about post-earthquake service continuity: a targeted search found assurance cases applied to resilient control systems and robotic autonomous systems [43,44] but not to community services exposed to earthquakes.

Three gaps follow from this situation.

- **G1. Claims are not graded.** Reviews classify resilience methods by hazard, sector, method family, and phase [24–27], but none asks which decision a given output can support. Results of very different strength therefore circulate as if they were equivalent.
- **G2. Continuity is not tested against requirements.** Composite indices and stable rankings stand in for adequacy [28–32]. Requirements that are explicit, non-compensatory, and specific to a user group and time window are called for in recovery guidance [12,13] but are seldom used to test a recommendation.
- **G3. The step from recommendation to delivered service is not evidenced.** Model feasibility is read as executability, commitment is documented mainly outside journals [33], and whether service was adequate after action is rarely observed. No instrument links these stages to the evidence each one needs.

This paper addresses these gaps through three research questions:

- **RQ1.** What claims about essential-service continuity can published seismic assessment, decision, and implementation studies support, and where do the inferences between them break?
- **RQ2.** How can the strength of such claims be graded consistently, transparently, and reproducibly?
- **RQ3.** Does grading the claims change the conclusions that would otherwise be drawn from published results?

It makes four contributions:

1. **A claim-centred critical synthesis** of a 76-record corpus spanning earthquake engineering, infrastructure modelling, decision science and governance, in which 29 sources are coded by the claim each can support rather than by hazard or method.

2. **SCALE, a Service-Continuity Assurance Ladder for Earthquake decisions.** SCALE places any recommendation on eight claim levels, from physical damage to verified service attainment; ties continuity to explicit, non-compensatory requirements for each service, user group and time window; records the remaining doubts in a defeater register with qualitative certainty grading; and links committed actions to monitored triggers. Its output, a Claim Card, states how far a recommendation can be trusted and what evidence would raise it.
3. **A demonstration on published results and a real municipality**. Re-analysing two recovery studies shows that 98.8% of the variation in modelled recovery time arises from workforce mobilisation rather than from the ranked recovery option, and that the preferred restoration strategy changes with the indicator and the declared service requirement. An application to the municipality of Agios Dometios, Cyprus [45], shows how SCALE identifies the next evidence a plan needs.
4. **A reproducible protocol**: coding rules and decision criteria, worked examples and five testable propositions that other researchers and practitioners can apply, contest and extend.

Contribution 1 addresses G1 by grading published claims. Contribution 2 addresses all three gaps: the ladder and coding rules for G1, the requirement module for G2, and the defeater register and adaptive commitment for G3. Contribution 3 tests whether closing G2 changes published conclusions, and contribution 4 makes the grading reproducible. Section 8.1 returns to the three gaps and states what remains open (Table 8).

The remainder of the paper is organised as follows. Section 2 describes the review design and coding protocol. Sections 3–5 synthesise the evidence along the ladder: from physical consequence to delivered service (Section 3), from adequacy to robust choice (Section 4) and from commitment to verified attainment (Section 5). Section 6 formalises SCALE, and Section 7 demonstrates it. Section 8 discusses implications for sustainable development and the Sendai Framework, together with limitations and a research agenda, and Section 9 concludes.

2. Review Design and Coding Protocol

2.1. Design

The study is a critical integrative review [46,47] combined with claim-level coding of a purposive corpus. A purposive design was chosen deliberately. The aim is analytical: to test whether a grading instrument can separate claims of different strength across heterogeneous studies, not to estimate how often particular methods occur in the field [48]. The corpus therefore emphasises contrast across hazard model, endpoint, scale and evidence form. Reporting follows the applicable items of the PRISMA extension for scoping reviews [49]. The corpus was not extended by database searching or citation snowballing [50]; Section 8.5 sets out that extension as a test of the findings.

The unit of analysis is the claim, not the study. For each included study, we record the strongest conclusion its evidence supports and, separately, the strongest conclusion its authors state. This lets the review locate where inferences break (RQ1) without treating a study's legitimate scope limits as flaws.

2.2. Corpus and Screening

The corpus comprises 76 records assembled by the author during preliminary work, organised in 14 thematic sets spanning seismic resilience and recovery; interdependent infrastructure and emergency accessibility; seismic prioritisation and resource allocation; translation of damage into recovery decisions; earthquake decision support; municipal governance; emergency preparedness and institutional authority; evidence-to-decision

methods; structured argumentation; adaptive planning under uncertainty; knowledge-to-action; and risk communication. Full texts were available for all records; one was an image-only file without extractable text and was therefore excluded from coding.

Each record was screened against the criteria in Table 1. Twenty-seven records met all criteria and were coded; seven were reviews, used for context but not coded; two were seismic studies whose outcomes did not meet the outcome criterion; and 39 were methodological sources from other fields or hazards, which inform the instrument but are not treated as evidence about earthquakes (Figure 2). Because implementation evidence (CCL-6 and CCL-7) is reported mainly outside journals, a targeted search of institutional sources added two documents that were also coded: the NIST guidance on designing lifelines for functional recovery [12] and the follow-up audit of Christchurch's pipe and road rebuild [33]. The coded set therefore contains 29 sources, published between 2006 and 2027; 23 of them appeared in 2023 or later.

Table 1. Eligibility criteria for coded evidence.

Criterion	Included	Excluded
Hazard	Earthquake, alone or within a multi-hazard study that reports seismic results	Studies with no seismic results
Object	Buildings, lifelines or services at facility, network, community or municipal scale	Single structural components with no link to function, service or decision
Outcome	At least one of the following: functional or service outcome; intervention, prioritisation or recovery choice; implementation or delivery evidence	Hazard, ground-motion or collapse-risk characterisation only; communication products only
Source	Peer-reviewed article; authoritative guidance; official post-event evaluation	Reviews (used for context only); opinion pieces without data or method

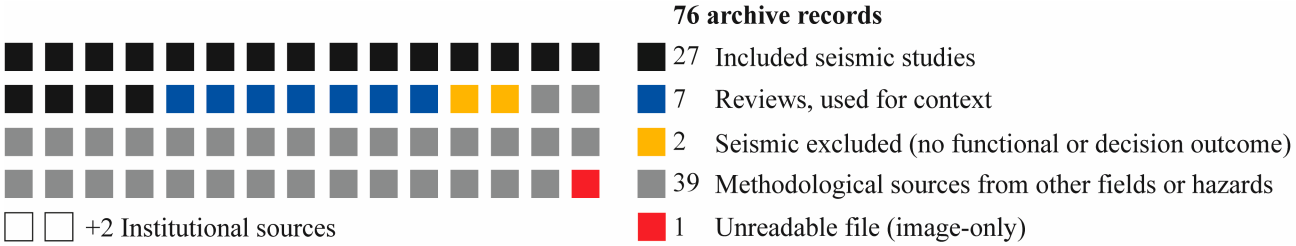

Figure 2. Screening of the 76-record corpus. Each square is one record. The two hatched squares are institutional sources added by targeted search.

2.3. Claim-Level Coding

Each coded source was assessed with the instrument summarised in Table 2 and defined fully in Section 6. Every level from CCL-0 to CCL-7 received one of three codes: F, a focal endpoint whose warrant the study's own results meet; P, a focal endpoint whose warrant is only partly met; or S, supporting context, such as a model input. Two derived quantities follow.

- **Supported level:** the highest level coded F or P.
- **Stated level**: the strongest explicit statement in the abstract or conclusions about what the results show or justify, coded on the same scale. Hedged statements of potential use ("may inform", "could support") are not coded as claims.

The difference between the two, the claim gap, measures how far stated implications exceed demonstrated evidence. A gap of zero is the expected outcome for most studies. We report its distribution and do not use it to score individual authors.

Table 2. Coding scheme.

Field	Codes
Continuity Claim Level	CCL-0 physical; CCL-1 functional; CCL-2 serviceable; CCL-3 delivered; CCL-4 adequate; CCL-5 robustly adequate; CCL-6 committed; CCL-7 attained
Level code	F focal, warrant met; P focal, warrant partly met; S supporting context
Ranks options	Focal; supporting; no
Evidence form	Model or simulation; guidance; documented observation; observation combined with model; expert elicitation
Service	Healthcare; water and power; transport; shelter and housing; emergency response; education; buildings (generic); multi-service
Service requirement	None; explicit, with its basis (guidance, agreed levels of service, stakeholder deliberation)
Aggregation	Single service; compensatory composite; disaggregated by service, group or place; not applicable
Uncertainty treatment	None; parametric sensitivity; scenario set; probabilistic; not applicable
Defeaters	D1 endpoint non-equivalence; D2 compensatory aggregation; D3 rank-only robustness; D4 uncovered uncertainty; D5 demand misrepresentation; D6 model-feasible taken as executable; D7 outcome not observed or not attributable

2.4. Coding Procedure and Transparency

Full texts were converted to machine-readable form and coded by the author against the codebook. The screening register and coded dataset are available from the author on request. Because a single coder cannot establish reliability, independent re-coding by a second analyst, with agreement measured by quadratically weighted kappa [51] and interpreted against conventional benchmarks [52], is set out as an explicit validation test (proposition P4, Section 6).

2.5. Synthesis and Re-Analysis

The synthesis has three layers. First, descriptive mapping reports the distribution of supported and stated levels, aggregation choices, uncertainty treatments and defeaters across the coded set. Second, a critical narrative synthesis (Sections 3–5) traces where inferences break between adjacent levels and distinguishes genuine contradiction from endpoint, condition or evidence-stage non-equivalence. Third, two published studies were re-analysed to test RQ3 (Section 7). Re-analysis used only values reported in the original tables or text, without digitising figures, and was limited to transformations that do not alter the original estimates: variance decomposition, requirement-based admissibility tests and comparison of indicators. Every value was checked against the source document.

3. From Physical Consequence to Delivered Service (CCL-0 to CCL-3)

3.1. Profile of the Coded Evidence

The coded evidence is concentrated on the lower half of the ladder (Figure 3). Of the 29 sources, seven support their strongest claim at CCL-0, five at CCL-1, four at CCL-2 and nine at CCL-3. One study reaches CCL-4 partially, three sources reach CCL-6 (one partially), and none reach CCL-5 or CCL-7. Among the 27 journal studies, 25 stop at or below CCL-3. Only three sources declare an explicit service requirement, and only one of the 14 studies whose focal output is a ranking of options tests that ranking against such a requirement. Thirteen sources aggregate outcomes into compensatory composites, and eight report no treatment of uncertainty.

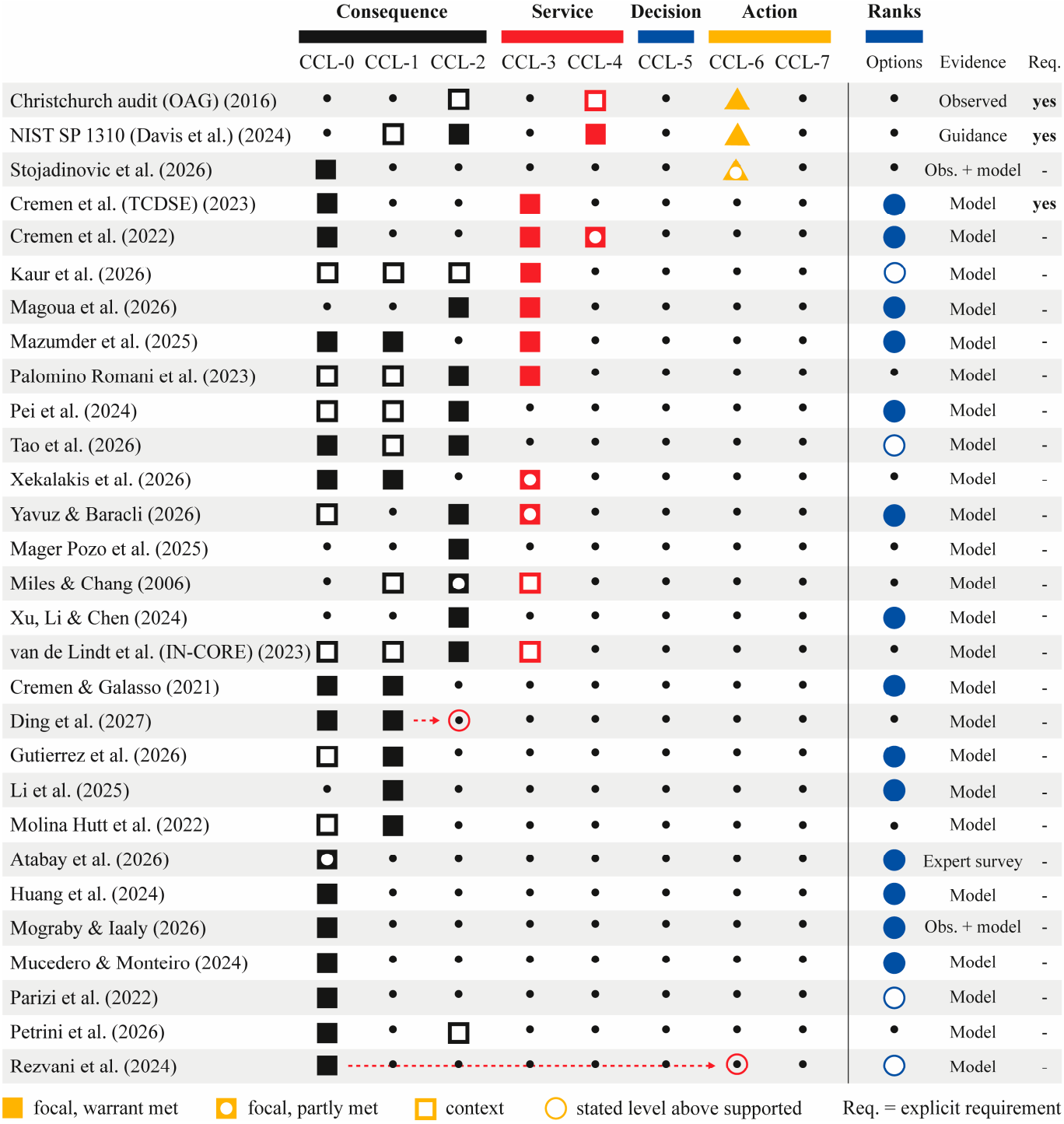

Figure 3. Claim levels addressed by the 29 coded sources, ordered by supported level. Filled marks: focal endpoint whose warrant is met; marks with a white centre: warrant partly met; open marks: supporting context. Red rings and dashed arrows show where the stated level exceeds the supported level. Req. = explicit service requirement declared. Rows, from top to bottom, correspond to references [9,10,12,14–21,33,45,53–68].

The claim gap is small. Twenty-seven sources state conclusions at exactly the level their evidence supports. Of the two exceptions, one describes building-portfolio recovery weighted by functional importance (CCL-1) as an evaluation of urban function (CCL-2) [62]. The other states that a municipal composite index enhances resilience and supports implementation (CCL-6), while its results characterise risk and capacity (CCL-0) [68]. Several

studies are explicitly careful about scope, for example by stating that expert rankings are perceptions rather than measurements, or that a screening queue supports inspection and not vulnerability prediction.

This pattern is the first substantive finding of the review. Inferential overreach in this literature arises less from how authors describe their own results than from how results are combined and reused. A study that legitimately reports accessibility, or a ranking under stated weights, becomes the basis of a claim about adequate, robust or deliverable service only when it is carried into a decision. That is where an instrument for grading claims is needed, and it is where the following sections locate the breakpoints.

3.2. Physical Consequence to Function (CCL-0 to CCL-1)

Seven sources stop at physical consequence. They include multi-scale loss assessment across building stock and utility networks [67], composite indices of urban physical resilience [66], joint seismic and energy-retrofit prioritisation [65], loss-based multicriteria decision support [21], and an expert survey ranking damage factors and mitigation measures [63]. These studies answer legitimate questions about damage, loss and priority. Their indicators, however, describe the condition of assets, not their use.

Building-level recovery models make the step to function explicit. Molina Hutt et al. separate stability, shelter-in-place, reoccupancy, functional recovery, and full recovery, and model impeding factors and repair sequencing [9]. Gutiérrez et al. quantify functionality as the share of non-usable floor area, so that partial loss can be represented [10]. The defeater at this transition is D1, endpoint non-equivalence. Usable floor area is not the same as an operating service: a hospital may retain most of its area and lose its emergency department, and a building can be physically sound but have no water or power. It is the most frequent defeater in the coded set, present in 16 of 29 sources.

3.3. Function to Serviceable Systems (CCL-1 to CCL-2)

The step from a usable asset to a system that can provide a service runs through interdependence. In a simulated community, power outages caused functionality loss in up to 75% of buildings that were physically operable, for as long as 14 days [1]. Open modelling environments chain building, network and social models for this reason [17], and restoration models optimise repair sequences across coupled power and water systems [19,60].

Two studies show how strongly interdependence can change a service conclusion. A road-network model found that, under an earthquake scenario for Tehran–Karaj, one-fifth of fire stations lose more than half of their access roads and four become completely inaccessible [58]. A simulation of post-earthquake fire response found a firefighting success rate of 53.0% in a strong-earthquake scenario, with 62.8% of failures caused by insufficient water supply. Under a magnitude 8.0 scenario, success fell to 20.8%, and 67.7% of failures were caused by transport inaccessibility [56]. The dominant mechanism of service failure thus changed with hazard intensity. A study that represented only one dependency would have identified the wrong intervention.

The lesson for claims is that CCL-2 evidence is conditional on which dependencies are modelled. Physical, geographic, network and organisational dependence are distinct mechanisms, and a model that represents one says nothing about the others. A strong CCL-2 claim must therefore name the mechanisms it covers.

3.4. Serviceable Systems to Delivered Service (CCL-2 to CCL-3)

Nine sources reach CCL-3, the level at which service reaches users. They do so through three different endpoints, which are not interchangeable.

- **Access**. Kaur et al. combine hospital and bridge damage with population geography in a floating-catchment accessibility index; targeted hospital retrofit raised average accessibility by 77%, 40%, and 78% in three municipalities [14]. Pei et al. evaluate coupled transport–healthcare accessibility during restoration [16].
- **Throughput**. Palomino Romani et al. simulate emergency-department queues under damage and patient surge, reporting waiting times and lengths of stay [15].
- **Unmet need**. Displacement and unmet demand measure the people who do not receive a service: displaced residents in urban-planning and risk-reduction studies [20,54,55], unmet power and water demand during restoration [19], and displaced persons in a municipal case [45].

An accessibility index does not show that arriving patients are treated, and a queue model does not show that patients can reach the hospital. The healthcare studies above are therefore complementary measurements of one service, not competing estimates of it. Distribution matters as well: municipal averages can conceal groups that remain without service [30]. A district-level allocation model goes some way towards this by weighting social vulnerability and accessibility, but it plans delivery rather than demonstrating it [57].

The main threat at this transition is demand. Delivered service is the ratio of what is provided to what is needed, and demand after an earthquake changes with displacement, injury and surge. Two coded sources rely on simplified demand assumptions (D5), including the household-size proxy in the municipal case [45], which is examined in Section 7.3.

4. From Delivered Service to Adequate and Robust Choices (CCL-4 and CCL-5)

4.1. The Missing Requirement

The step from delivered to adequate service requires something no model can supply by itself: a statement of how much service is enough, for whom and when. Only three of the 29 coded sources declare such a requirement. NIST guidance specifies recovery objectives for lifeline systems [12,13]. The Christchurch audit records that levels of service were eventually agreed between the public entities [33]. The Tomorrow's Cities decision-support environment uses a stakeholder visioning step to fix the highest acceptable value of a displacement-based risk metric and checks candidate policy bundles against it [54]. The Tomorrow's Cities environment is the only study in the set that tests a ranking against a declared requirement. It reaches CCL-4 only partially because its threshold applies to an aggregate risk metric rather than a named service, user group, and time window.

The idea is not new. Mieler et al. proposed linking community-level resilience goals to performance targets for individual components a decade ago [11]. The coded set shows that studies recommending interventions rarely make this link. Without a requirement, the strongest defensible claim about an intervention is comparative ("better than the alternatives tested"), not absolute ("good enough").

4.2. Aggregation Can Hide Service Failure

Thirteen coded sources combine outcomes into compensatory composites, so a gain in one component can offset a loss in another. Two published results show why this matters for continuity. First, recovery trajectories with very different timing can produce the same resilience value [28], and different summary metrics of the same curves can recommend different strategies [29]. Figure 4A illustrates the first point with a constructed example: two trajectories with the same time-averaged service level (0.60) leave users below a requirement of 0.5 for very different durations, giving continuity deficits of 0.11 and 0.41.

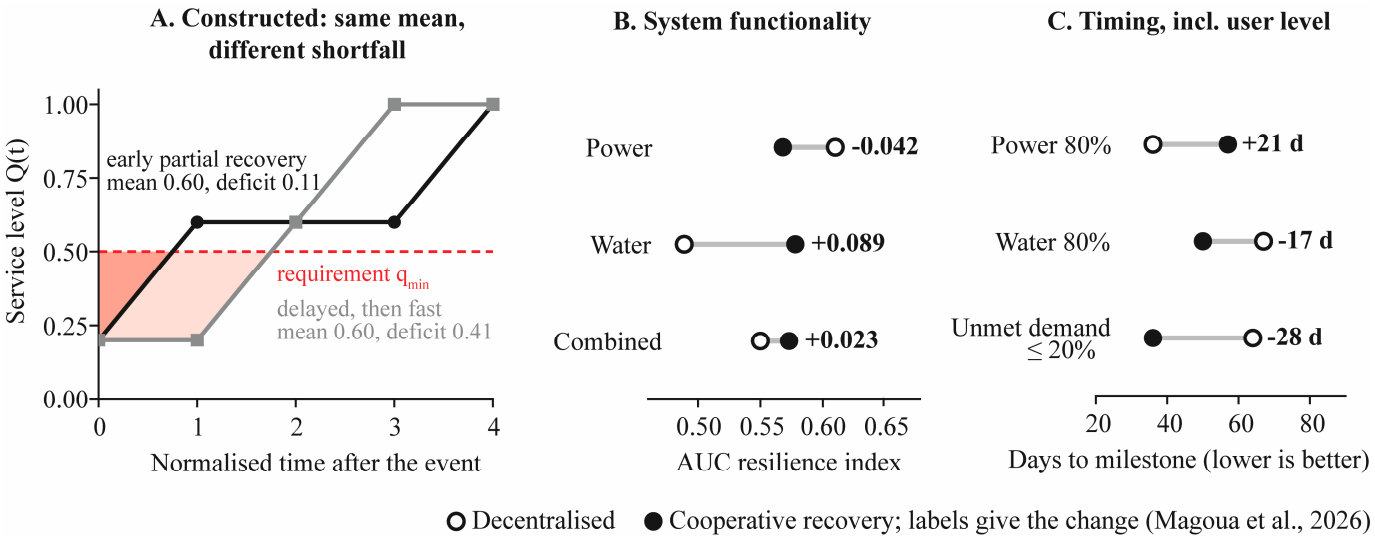

Figure 4. Aggregate indicators and service continuity. (**A**) Constructed example: two recovery trajectories with the same mean service level (0.60) have continuity deficits below the requirement q_min of 0.11 and 0.41. (**B**) AUC resilience indices under decentralised and cooperative recovery, as reported in Table 3 of Magoua et al. [19]. (**C**) Days to reach milestones in the same study; cooperative recovery delays power by 21 days but reduces unmet demand to 20% or less 28 days earlier.

Second, in a published power–water restoration study, cooperative decision-making raised the combined resilience index from 0.550 to 0.573, while the power index fell from 0.610 to 0.568 and the water index rose from 0.489 to 0.578 [19]. The same study reports a user-level indicator, unmet demand. On that indicator, the cooperative schedule is clearly better: unmet demand fell to 20% or less after 36 days rather than 64 (Figure 4B,C). Read by functionality, the strategy trades power for water. Read by delivered service, it helps users sooner. Neither reading establishes adequacy, because no requirement is declared. The point is not that composites are wrong, but that a composite answers a weighted question, and a continuity claim asks a non-compensatory one.

4.3. A Stable Ranking Is Not a Robust Choice

Fourteen coded sources produce a ranking of options as their focal output. Rankings are informative, but their stability is often mistaken for robustness. The difference can be stated in one line. Suppose the losses of two interventions A and B in scenario ω are

$$L_A(\omega) = h(\omega) + c_A, \qquad L_B(\omega) = h(\omega) + c_B, \qquad c_A < c_B.$$

Then A is preferred in every scenario, yet nothing bounds $h(\omega)$. The ranking is perfectly stable while the outcome of the preferred option can vary without limit, and it can fail any requirement in some scenarios or in all of them. Section 7.1 shows that this is not hypothetical.

The coded studies display three distinct kinds of sensitivity, which are often reported as one. **Preference sensitivity:** the preferred urban-planning policy changes with the weight given to social consequences [20], and the optimal early-warning action changes with stakeholder priorities [61]. **Parametric sensitivity:** varying model inputs within ranges leaves the preferred option unchanged [20]. **Outcome sensitivity:** the consequences of the preferred option vary widely across scenarios [18]. Only the third bears on whether the preferred option delivers adequate service.

Decision science distinguishes these properties explicitly. Robustness metrics based on absolute performance, regret or satisficing can rank the same options differently [32], and distributionally robust allocation protects against mis-specified loss distributions at the price

of conservatism [23]. Robust Decision Making stress-tests options against requirements across many plausible futures [69], and adaptive pathways specify when a decision should be revisited [42]. Of the 29 coded sources, 13 treat uncertainty probabilistically, five use a small scenario set, one reports parametric sensitivity and eight report no treatment. None applies deep-uncertainty methods or tests whether a recommended option remains admissible under a declared requirement. CCL-5 is therefore empty.

5. From Choice to Committed Action and Verified Attainment (CCL-6 and CCL-7)

5.1. Model-Feasible Is Not Executable

In optimisation and recovery models, a feasible solution is one that satisfies the constraints written into the model: budgets, crews, repair durations, and precedence. Six coded sources rely on such feasibility to support action-oriented conclusions (D6). Restoration schedules assume that operators will follow the optimised sequence [19,60], allocation models assume that resources can be moved as planned [57], and recovery experiments assume mobilisation levels that no institution has committed to [18]. These assumptions are legitimate for the question each model asks. They do not show that any organisation has the authority, funding, staff, or agreements to carry out the plan.

Two coded studies address intermediate steps towards execution with unusual care. Stojadinović et al. translate observed damage into executable repair quantities while keeping this step separate from safety assessment and loss modelling [53]. This establishes the scope of work but not who will fund and deliver it, so it is coded as a partial CCL-6. Mograby and Iaaly test spatial input layers for provenance and applicability before combining them for municipal screening. Their procedure admitted only one of six candidate layers, and the resulting inspection queue (400 buildings) differed from that of the unscreened composite (303) [64]. They explicitly limit the claim to inspection triage. Both studies show what bounded, action-relevant claims look like.

5.2. Commitment and the Institutions That Carry It

Commitment is an institutional fact, so its evidence lies mostly outside journals. NIST guidance on lifeline functional recovery couples recovery objectives with the organisational actions needed to meet them, including dependencies, staff access, and mutual assistance [12,13]. It specifies what a CCL-6 claim should contain, but as guidance it cannot show that any municipality has done so.

The Christchurch rebuild shows commitment as a process that changes over time. The 2013 audit of the alliance repairing the city's pipes and roads found that the public entities did not share a common understanding of the levels of service the repaired networks should deliver. By the 2016 follow-up, levels of service had been agreed, and governance improved, but funding arrangements had taken up to 19 months to confirm, leaving about 30 wastewater and stormwater projects uncertain for more than eight months [33]. A CCL-6 claim is therefore perishable: requirements, owners, and funding must be documented, and they must be re-verified as delivery proceeds.

Studies from other hazards explain why technically sound evidence loses force at this stage. In local climate adaptation, centrally produced risk assessments lose decision relevance as responsibility passes from planning to budgeting and implementation [70]. In local disaster risk reduction, formal mandates diverge from the institutions that operate in practice [71]. These are mechanisms, not seismic outcome estimates, but they predict where a seismic recommendation will stall.

5.3. Verified Attainment: The Empty Rung

No coded source reaches CCL-7, the observation that service met a declared requirement after works or an event, with the outcome attributable to the intervention. This is unsurprising for ex ante studies, which cannot observe a future earthquake. Post-event evaluations rarely start from a declared requirement, so they can document what was delivered but not whether it was enough. The practical remedy is to make ex ante recommendations testable: name the requirement, the users, the indicators, and the monitoring arrangements before the event, so that a post-event evaluation has something to test.

5.4. Where Inferences Break: Synthesis

Table 3 brings Sections 3–5 together and answers RQ1. The ladder has seven transitions. At each, a specific assumption tends to carry a claim further than its evidence, and a specific kind of evidence is needed to make the step legitimate. The three transitions with the weakest support in the coded set are 3 → 4 (no declared requirement), 4 → 5 (rank stability taken as robustness), and 6 → 7 (no observed attainment).

Table 3. Breakpoints between adjacent claim levels in the coded evidence.

Transition	Assumption That Carries Claims Too Far	Typical Defeater	Signal in the Coded Set (n = 29)	Evidence Needed to Climb
CCL-0 → 1	An intact asset is a usable asset	D1	7 sources stop at CCL-0	Non-structural damage, contents, impeding factors
CCL-1 → 2	A usable asset is a functioning system	D1, D4	5 stop at CCL-1	Named dependency mechanisms and their failure rules
CCL-2 → 3	Available capacity is delivered service	D1, D5	4 stop at CCL-2; D1 in 16 sources	Demand model; access and throughput together; disaggregation
CCL-3 → 4	Better is enough	D2	9 stop at CCL-3; 26 declare no requirement; 13 compensatory	Requirement R with a stated basis; non-compensatory test
CCL-4 → 5	A stable ranking is a robust choice	D3, D4	14 rank options, 1 tests a requirement; 0 reach CCL-5	Admissibility across a declared stress set
CCL-5 → 6	A feasible plan is an executable one	D6	D6 in 6 sources	Owner, authority, funding, agreements, triggers
CCL-6 → 7	Delivered works are attained service	D7	0 reach CCL-7	Baseline, monitoring, attribution design

6. The SCALE Framework

SCALE (Service-Continuity Assurance Ladder for Earthquake decisions) turns the breakpoints in Table 3 into an instrument. It grades the claim attached to a seismic recommendation, not the model that produced it. It has four modules: a requirement, a ladder, a register of doubts, and an adaptive commitment. Its output is a Claim Card. Figure 5 shows the four modules acting together on a single decision.

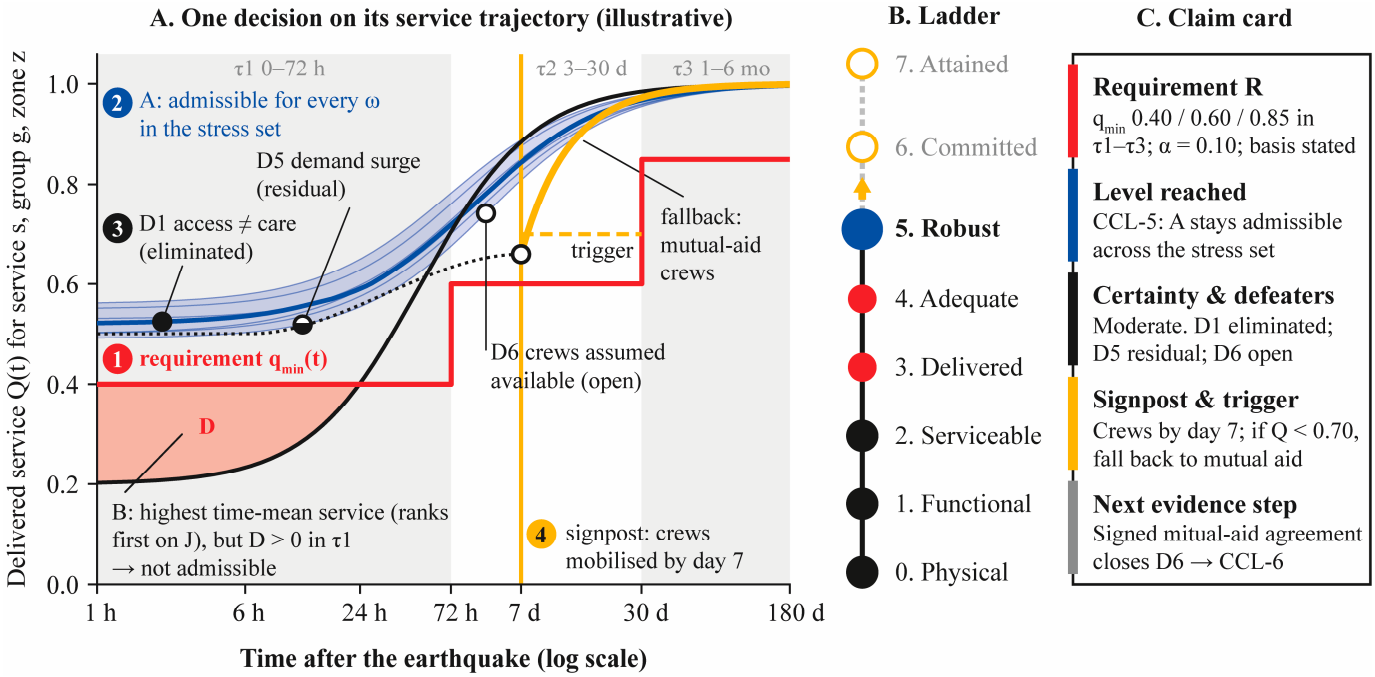

Figure 5. Anatomy of a SCALE assessment for one decision (illustrative values). (**A**) Delivered service for one service, user group, and zone across three time windows. Module 1 sets the requirement envelope q_min(t) (red line). Alternative B has the higher time-mean service and would rank first on a compensatory objective J, but its shortfall D during the first 72 h makes it inadmissible. Alternative A meets the requirement in every scenario of the declared stress set (blue band). Module 3 pins each defeater where it acts (filled: eliminated; half-filled: residual; open: open). Module 4 monitors a signpost at day 7: if observed service (dotted line) is below the trigger, the fallback pathway (yellow) is activated. (**B**) Module 2 places the claim at CCL-5; closing the open defeater D6 is the step to CCL-6. (**C**) The Claim Card that reports the result (Section 6.5).

6.1. Module 1: The Service-Continuity Requirement

Every continuity claim is tied to an explicit requirement for a service s, a user group g, a zone z and a time window τ:

$$R = \langle s,\ g,\ z,\ \tau,\ q_{\min}(t),\ \alpha,\ \text{basis} \rangle$$

Here q_min(t) is the minimum acceptable service level over the window, α is the tolerated probability of failing it, and the basis records where the requirement comes from: statute, an emergency plan, a recovery objective [12,13], or a documented deliberation with those responsible [11]. Typical windows are 0–72 h, 3–30 days and 1–6 months. The framework measures only the shortfall below the requirement, for each service and group:

$$D_{s,g}(a,\omega) = \int_{\tau} \max\left(0,\ q_{\min}(t) - Q_{s,g}(t;\ a,\omega)\right) dt$$

where Q is the delivered service level under alternative a and scenario ω. A deficit of zero means the requirement is met throughout the window. Unlike the area under a recovery curve, the deficit distinguishes trajectories with the same mean (Figure 4A).

The decision rule is to satisfice first and optimise second. An alternative is admissible only if it meets every critical requirement in the set C with the required reliability. Only admissible alternatives are then ranked on the usual objective J, such as cost, loss or a composite index:

$$F = \left\{ a : \Pr_{\omega}\left[D_{s,g}(a,\omega) \le D^{\text{tol}}_{s,g} \ \ \forall (s,g) \in C \right] \ge 1 - \alpha \right\}, \qquad a^{*} = \underset{a \in F}{\operatorname{argmax}} J(a)$$

where scenarios have no defensible probabilities, the probability is replaced by "for every ω in the declared stress set". If F is empty, the correct output is that no tested alternative meets the requirement. That is a finding, not a failure: it calls for new alternatives or a deliberate, documented revision of R.

6.2. Module 2: The Continuity Claim Ladder

A recommendation is placed at the highest level its evidence supports. The levels are cumulative. Table 4 gives the definition and the warrant needed for each level; Table 2 gives the corresponding codes. Figure 6 places the coded evidence on the ladder.

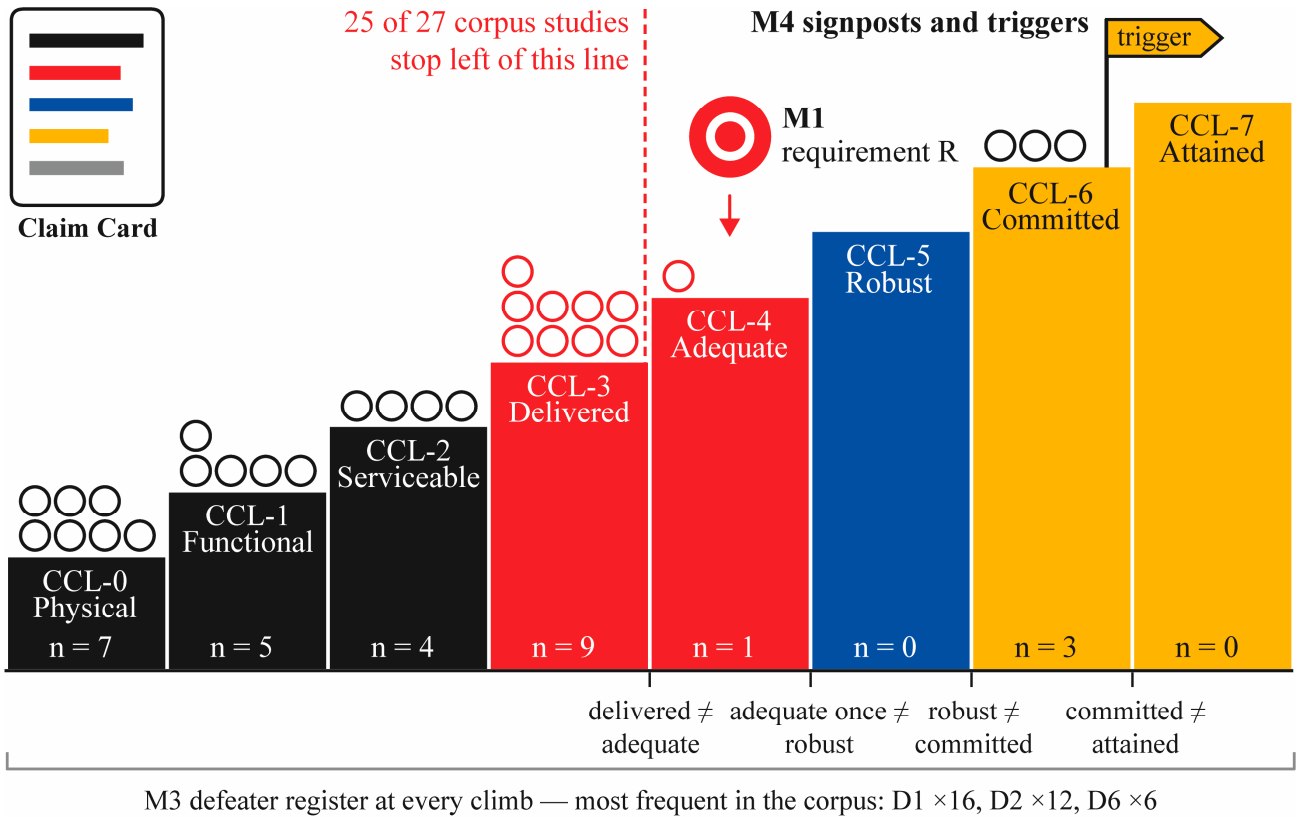

Figure 6. The Continuity Claim Ladder (Module 2) drawn over the evidence it was derived from. Each step is a claim level; its height orders the claims, not their value. Circles on each step are the coded sources whose supported claim stops there (n = 29, including two institutional sources). Module 1 (requirement R) enters at CCL-4; Module 3 (defeater register) applies at every climb; Module 4 (signposts and triggers) makes CCL-6 operational; the Claim Card reports the result. The coloured lines in the Claim Card icon represent its fields, using the colour code of Figure 5; grey marks the next evidence step.

Table 4. The Continuity Claim Ladder.

Level	Claim the Evidence Supports	Warrant Needed
CCL-0 Physical	Damage states, losses, safety	Validated hazard, exposure and fragility
CCL-1 Functional	Asset usability, reoccupancy, recovery time	Non-structural damage, contents, impeding factors

Table 4. *Cont.*

Level	Claim the Evidence Supports	Warrant Needed
CCL-2 Serviceable	The system or facility can provide the service	Named dependency mechanisms
CCL-3 Delivered	Users obtain the service, by group, place and time	Demand model; access and throughput; disaggregation
CCL-4 Adequate	Delivered service meets requirement R	Declared R with basis; non-compensatory test
CCL-5 Robustly adequate	The recommended option stays admissible across the declared uncertainty set	Stress set; admissibility test
CCL-6 Committed	Owner, authority, funding, agreements and triggers are documented	Decision record endorsed by responsible actors
CCL-7 Attained	Observed service met R after works or an event	Baseline, monitoring, attribution

6.3. Module 3: The Defeater Register and Certainty Grading

SCALE borrows its treatment of doubt from assurance cases in safety-critical engineering [38,39]. In eliminative argumentation, confidence in a claim grows as the doubts that could defeat it are identified and resolved [40]. Three kinds of defeater are distinguished:

- **Rebutting**: evidence that the claim is false, such as the power index falling under the "better" strategy.
- **Undermining**: doubt about the evidence, such as an assumed household size that differs from the census.
- **Undercutting**: doubt about the inference, such as using an accessibility index to claim clinical continuity.

The seven defeater types D1–D7 (Table 2) are the recurrent defeaters found in the coded evidence. Each is marked eliminated (with the evidence that eliminates it), residual (accepted and stated) or open. Certainty in the claim is then graded qualitatively (high, moderate, low, or very low) on five domains adapted from evidence-to-decision frameworks [36,37]: model credibility, indirectness, imprecision, inconsistency, and scenario selection. Indirectness corresponds to D1, and imprecision to the spread of outcomes across the stress set. SCALE deliberately avoids a numerical confidence score, because proposed techniques for quantifying confidence in assurance arguments can give implausible results and have not been validated [41].

6.4. Module 4: Adaptive Commitment

A CCL-6 claim names, for each requirement, the responsible operator, a signpost to monitor, the trigger value at which the recommended option would leave the admissible set, and the fallback action. The terms follow dynamic adaptive policy pathways [42]. In Section 7.1, for example, the signpost is the level of workforce mobilisation.

6.5. Output: The Claim Card

The Claim Card records, for one recommendation and each affected service: the requirement R and its basis; the highest level reached; the certainty grade; the open and residual defeaters; the signposts, triggers, and fallback; and the next evidence step needed to climb one level. It is short enough to accompany a council decision and structured enough to be audited later. It is designed as a shared object between engineers, operators, and municipal decision makers, in the sense used in knowledge-to-action research, where usable

evidence must be credible, salient, and legitimate to all parties [72], and boundary objects help groups with different expertise reason together [73].

6.6. Relation to Existing Frameworks

SCALE does not replace performance-based engineering, recovery guidance, or reviews of methods; it grades what their outputs can support. Table 5 summarises the distinction.

Table 5. SCALE and the nearest frameworks.

Framework	Primary Object	Non-Compensatory Service Requirement	Explicit Register of Doubts	Implementation and Monitoring Stages
Rapid screening framework [25]	Infrastructure screening under constrained data	Not a stated aim	Not a stated aim	Evidence-to-decision bridge
Review of resilience methods [26]	Method families and their applicability	Not a stated aim	Not a stated aim	Not a stated aim
NIST lifeline functional recovery [12,13]	Recovery objectives and design	Yes (recovery objectives)	Not explicit	Yes (planning)
GRADE evidence-to-decision [36,37]	Health and environmental recommendations	Not seismic	Yes (certainty domains)	Partly
SCALE (this study)	**The claim attached to a seismic recommendation**	**Yes, per service, group and window**	**Yes (defeater register)**	**Yes (CCL-6, CCL-7, triggers)**

Note: The framework proposed in this study is shown in bold.

6.7. Testable Propositions

SCALE implies five propositions that can be tested independently.

1. **P1.** Most published seismic decision studies support claims at CCL-3 or below. In the present corpus, 25 of 27 journal studies do so; a systematic sample is needed to test the proposition beyond it.
2. **P2.** Replacing compensatory aggregation with the satisfice-then-optimise rule changes the recommended option in a non-trivial share of published cases (tested on one case in Section 7.2).
3. **P3.** Implementation conditions explain more of the variation in modelled outcomes than the choice among ranked technical options (tested on one case in Section 7.1).
4. **P4.** Independent coders assign claim levels reproducibly, with weighted kappa above 0.6 [51,52].
5. **P5.** Decision makers given a Claim Card identify the unsupported part of a recommendation more often than those given a composite resilience score.

7. Demonstration

This section applies SCALE to two published studies and one municipality. The two re-analyses use only values reported in the original papers. All requirements below are illustrative unless a source declares them; their purpose is to show how the verdict depends on R, not to propose values.

7.1. Rank Stability Versus Outcome: Building Recovery Under Workforce Constraints

Li et al. modelled portfolio-level building recovery for three recovery sequencing options under three workforce-mobilisation scenarios [18]. Option 3 ranks first in every scenario. A two-way decomposition of the nine reported recovery times attributes variation to the scenario, the option and their interaction:

$$SS_{total} = SS_{scenario} + SS_{option} + SS_{interaction}$$

The scenario accounts for 98.8% of the variation, the option for 0.8%, and the interaction for 0.5% (Figure 7B). Within a scenario, choosing the best option rather than the worst saves 104 to 629 days. Holding the option fixed and changing the scenario changes recovery by 3228 to 3820 days. The ranking is also less stable than first rank suggests: Option 1 beats Option 2 in the baseline and pessimistic scenarios but not in the optimistic one.

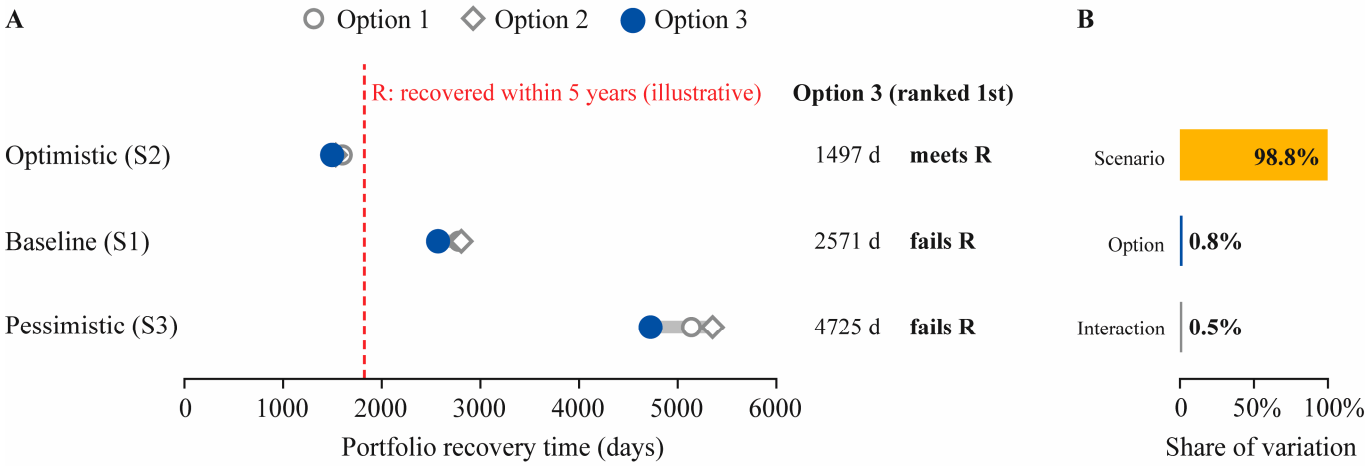

Figure 7. Re-analysis of the values reported in Table 6 of Li et al. [18]. (**A**) Portfolio recovery time for three options under three workforce-mobilisation scenarios; Option 3 ranks first in every scenario but meets an illustrative five-year requirement only under optimistic mobilisation. Markers overlap where options differ by less than about 100 days, as in the optimistic scenario (1497–1601 d). (**B**) Share of the variation across the nine results attributable to scenario, option, and interaction. Colours follow the SCALE bands: yellow for implementation conditions (scenario), blue for the decision (option) and grey for interaction.

Table 6. Preferred strategy under alternative decision rules (values from Magoua et al. [19]; the two deadline rules are hypothetical).

Decision Rule	Indicator	Decentralised	Cooperative	Preferred
Compensatory	Combined AUC index (higher is better)	0.550	0.573	Cooperative
Single service	Power AUC index	0.610	0.568	Decentralised
Worst service	Longer of the two t80 values (days)	67	57	Cooperative
Service priority	Power t80 within 45 days	36 (meets)	57 (fails)	Decentralised only
Joint requirement	Power within 45 days and water within 60 days	Water 67 (fails)	Power 57 (fails)	None admissible
User level	Days until unmet demand ≤ 20%	64	36	Cooperative

Applying Module 1 changes the conclusion. Against an illustrative requirement that the portfolio recover within five years (1825 days), Option 3 is admissible only under optimistic mobilisation (1497 days); it fails under the baseline (2571 days) and pessimistic (4725 days) scenarios (Figure 7A). Against a 3000-day requirement, no option is admissible under pessimistic mobilisation, so the admissible set F is empty. Option 3 is robustly adequate (CCL-5) only for requirements of at least 4725 days.

The SCALE reading is therefore that Option 3 is the better sequence (a CCL-1 comparative claim) but that adequacy depends almost entirely on mobilisation, which is an institutional condition rather than a technical choice. Module 4 makes this operational:

workforce mobilisation is the signpost, and its trigger lies between the optimistic and baseline assumptions for a five-year requirement. The case supports proposition P3. (Li et al. report slightly different values for Option 1 in their text and in their Table 6; we use Table 6 throughout.)

7.2. The Verdict Depends on the Indicator and the Requirement: Power–Water Restoration

Magoua et al. compared decentralised, information-sharing, cooperative, and centralised restoration of interdependent power and water systems in Mianzhu after the 2008 Wenchuan earthquake [19]. Table 6 compares decentralised and cooperative recovery under six decision rules, using the published indices and days to reach 80% functionality (t80). Figure 4B,C shows the same values graphically.

Six decision rules applied to one published experiment yield three different answers. Figure 8 generalises the table to every pair of deadlines: cooperative recovery is admissible when the power deadline is at least 57 days and the water deadline at least 50 days; decentralised recovery when they are at least 36 and 67 days. In the region where both qualify, the composite objective decides. In the region where neither qualifies, which includes the joint requirement in Table 6, the correct output is an empty admissible set. The SCALE reading is that "cooperative recovery is better" is warranted at CCL-3 on the user-level indicator, contestable at CCL-2 on system functionality, and undetermined at CCL-4 until a requirement is declared. The case illustrates proposition P2.

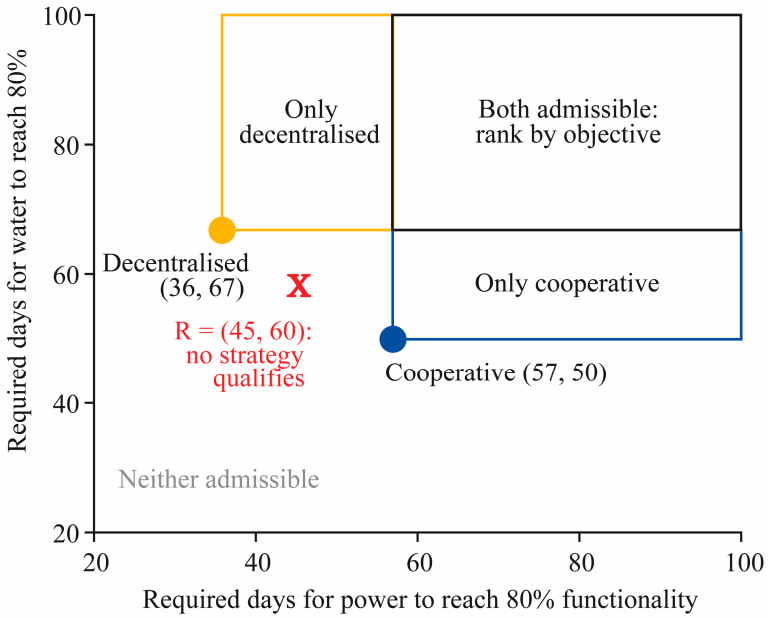

Figure 8. Admissible restoration strategy as a function of the deadlines set for power and water to reach 80% functionality, using t80 values from Magoua et al. [19]. Deadlines are hypothetical. At (45, 60) days, no tested strategy qualifies.

7.3. A Municipal Claim Card: Agios Dometios, Cyprus

The third application uses a fragility-based assessment of Agios Dometios, a municipality of Lefkosia [45]. Under the reference scenario (0.20 g, 475-year return period), the study estimates times to 90% functionality of 53.99 days for adobe and 36.93 days for stone masonry. It estimates about 615 displaced persons out of about 774 exposed residents in 172 adobe and 86 stone buildings, between 27,500 and 37,200 displaced person-days, and a temporary accommodation cost of EUR 0.29–0.37 million. It also maps candidate

refuge sites and primary and secondary walking routes. Table 7 shows the Claim Card for emergency sheltering.

Table 7. Claim Card: emergency sheltering, Agios Dometios.

Item	Entry
Recommendation	Plan temporary refuge for residents of the vulnerable masonry stock
Requirement R (illustrative)	Every displaced resident of the masonry stock has a refuge place within walking distance during 0–72 h; tolerated shortfall probability 10%; basis to be agreed with the civil-protection authority
Highest level reached	CCL-1 (functional loss and recovery time); CCL-3 partly, on the demand side (displaced persons)
Certainty	Moderate for CCL-1; low for CCL-3 (indirectness: routes stand in for access; imprecision: demand and duration ranges)
Residual defeaters	D5: three persons per building assumed versus 2.51 in the census, overstating demand by about 16% (about 515 rather than 615 persons). Only four vulnerable typologies modelled
Open defeaters	D1: route geometry is not access under debris or road blockage. Site capacities not assessed, so CCL-4 cannot be tested
Signpost and trigger	Share of school grounds available under signed agreements; trigger when available capacity falls below the lower demand bound
Next evidence step	Inventory 72-h capacity of each candidate site; compare with displaced demand by zone under both household-size assumptions; add road-blockage probability; agree R with the responsible authority

The card shows what the municipality can and cannot rely on today. The study is sufficient to locate the problem and size it to within about 16%. It is not yet sufficient to claim that sheltering will be adequate, and the card names the three pieces of evidence that would change that.

8. Discussion

8.1. What the Evidence Establishes

The three research questions can now be answered. RQ1: In the coded evidence, claims break at identifiable transitions (Table 3). They break most often between delivered and adequate service, where requirements are almost never declared; between adequate service and robust choice, where rank stability stands in for robustness; and between committed action and attained service, where no source observes outcomes against a requirement. Authors rarely overstate their own results, as shown by a claim gap of zero in 27 of 29 sources. The weak point is the reuse of results for decisions they were not designed to support. RQ2: a small set of explicit rules (warrants for each level, F/P/S codes, a stated-versus-supported comparison, and seven recurrent defeaters) was sufficient to code all 29 sources and to separate claims that are usually reported as one. Whether these rules are applied consistently by different coders remains to be tested (P4). RQ3: yes. Grading the claims changed the conclusion in both re-analyses. A recovery option that ranks first everywhere becomes adequate only under one mobilisation scenario, and a restoration strategy that improves a composite index is preferred, rejected, or inadmissible depending on the indicator and the requirement.

The contribution relative to existing work is specific. Reviews of resilience methods map methods by hazard, sector, and phase [24–26]; recovery guidance specifies objectives and design [8,12,13]; and decision-support studies rank options [20–23]. SCALE adds

an explicit, reproducible way to state how far any of these outputs can be trusted for a given decision and what evidence would extend that trust. It does so by transferring two mature ideas into earthquake engineering: certainty grading from evidence-to-decision frameworks [36,37] and defeater-based assurance from safety-critical engineering [38–41].

Table 8 closes the loop with the three gaps stated in the Introduction. For each, it summarises what this review establishes, what SCALE contributes, and what remains open.

Table 8. Research gaps, contributions of this study and remaining work.

Gap	What the Review Establishes	What This Study Contributes	What Remains Open
G1. Claims are not graded	Most coded studies stop at delivered service (CCL-3) or below; authors rarely overstate, so overreach arises when results are reused for adequacy decisions (Section 3)	Continuity Claim Ladder with warrants, F/P/S coding rules and the claim gap (Tables 2 and 4); codebook and coded data (available on request)	Independent second coding (P4); a systematic sample to test P1
G2. Continuity is not tested against requirements	26 of 29 sources declare no service requirement; 13 aggregate compensatorily; 1 of 14 ranking studies tests its preferred option against a requirement (Section 4)	Requirement R, continuity deficit and the satisfice-then-optimise rule (Module 1); re-analyses in which a declared requirement changes the conclusion (Sections 7.1 and 7.2)	Requirements set with responsible authorities (Section 8.5); further re-analyses to test P2 and P3
G3. The step from recommendation to delivered service is not evidenced	6 sources treat model feasibility as executability (D6); CCL-6 rests on guidance, one audit and one partial study; no source reaches CCL-7 (Section 5)	Defeater register and certainty grading (Module 3); signposts, triggers and fallbacks (Module 4); Claim Card with a next-evidence step (Sections 6.3–6.5 and Section 7.3)	Workshop test with municipal staff (P5); post-event evaluation against declared requirements

8.2. Implications for Practice and Sustainable Development

For municipalities, the immediate implication is to commission and receive seismic analyses as claims with stated levels. A request for "the most resilient option" invites a ranking. A request for "the option that keeps emergency healthcare accessible to every district within 72 h, with the evidence level stated" invites a CCL-4 answer and makes the remaining gap visible. The Claim Card (Section 6.5) gives that request a standard form, and its next-evidence-step field turns a limitation into a work order.

For sustainable development, SCALE offers a way to report Sendai Target D in terms of adequacy rather than counts. The global indicators record the number of disruptions to basic services [4]. A requirement-based deficit records who went without which service, for how long, and by how much. It therefore links Target D to SDG 11.5, which seeks to reduce the number of people affected by disasters with a focus on the poor and people in vulnerable situations, and to SDG 9.1, which seeks resilient infrastructure that supports human well-being with equitable access for all [74]. Non-compensatory requirements prevent a gain for one group or service from hiding a loss for another, which is the distributional concern raised by equity-focused resilience research [30,31]. The finding that institutional conditions can dominate technical choices (Section 7.1) also bears on the allocation of scarce resources. Investment in workforce mobilisation, agreements, and funding readiness may protect services more than refining the ranking of technical options.

8.3. Implications for Research

Four reporting practices would make published results usable at higher claim levels without changing their methods. Studies could state the claim level of their main conclusion in the abstract; report service outcomes disaggregated by service, group, and time window alongside any composite; report the spread of outcomes for the preferred option, not only its rank; and state any service requirement they test, with its basis. None of these practices requires new models, and each would let a reader see at once which decisions a result can support.

8.4. Limitations

The findings have five limitations. First, the corpus is purposive. The 76 records were assembled to span contrasting methods and evidence forms, so the frequencies reported here describe this corpus and cannot be generalised to the field. Second, the first coding pass was AI-assisted and has not yet been replicated by an independent human coder; all codes are available from the author for checking, and reliability is posed as a test (P4) rather than assumed. Third, implementation evidence is thin. Only two institutional sources were added, so CCL-6 and CCL-7 conclusions rest on few cases. Fourth, the re-analyses use two studies and the values they report; their requirements are illustrative and do not represent real service standards. Fifth, the municipal demonstration uses a study by the author, which made full data access possible but also requires readers to judge the card on its stated evidence alone.

8.5. Research Agenda

The propositions in Section 6.7 define the next steps. A systematic database search, extended by backward and forward citation searching [50], would test P1 on a representative sample. Independent re-coding of the present corpus would test P4. Re-analysis of further published factorial and multi-strategy studies would test P2 and P3 beyond single cases. A workshop experiment with municipal staff, comparing decisions made with Claim Cards and with composite scores, would test P5. Finally, applying Module 1 with a civil-protection authority to set a real sheltering requirement for Agios Dometios would carry the demonstration from CCL-3 towards CCL-4.

9. Conclusions

Earthquake engineering can now estimate not only what an earthquake will damage but how long services will be interrupted and which interventions shorten the interruption. The outputs of this work, however, support claims of very different strength. In the coded evidence, most claims stop at delivered service; almost none are tested against a declared requirement; rank stability is routinely the only robustness reported; and no source observes whether service was adequate after action was taken. The weak point is less how authors describe their results than how those results are reused for decisions they were not built to support.

This paper proposes SCALE, a Service-Continuity Assurance Ladder for Earthquake decisions, to grade those claims. SCALE ties continuity to explicit, non-compensatory requirements; places any recommendation on eight claim levels; records the doubts that remain in a defeater register; and links committed actions to monitored triggers. Applied to published results, it changes conclusions. In one study, 98.8% of the variation in modelled recovery time came from workforce mobilisation rather than the ranked option. In another, the preferred restoration strategy changed with the indicator and the service requirement. Applied to a Cypriot municipality, it identified the three pieces of evidence needed before a sheltering plan could be called adequate.

SCALE does not replace engineering models. It makes explicit what its results can support, so that municipalities can act on them with known confidence and researchers can see which evidence would matter most. Its codebook, coded data, and analysis scripts are available from the author on request, so others can test, contest, and extend it.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study (screening register, codebook, coded dataset, values extracted from published sources, and analysis scripts) are available from the corresponding author upon reasonable request.

Conflicts of Interest: The author declares no conflicts of interest. Section 7.3 applies the framework to a study co-authored by the author [45].

References

1. Mazumder, R.K.; Sutley, E.J. A multi-step framework for measuring post-earthquake recovery: Integrating essential infrastructure system's serviceability in building functionality. *Int. J. Disaster Risk Reduct.* **2024**, *114*, 104929. [CrossRef]
2. Crowley, H.; Dabbeek, J.; Despotaki, V.; Rodrigues, D.; Martins, L.; Silva, V.; Romão, X.; Pereira, N.; Weatherill, G.; Danciu, L. *European Seismic Risk Model (ESRM20)*; EFEHR Technical Report 002, V1.0.1; EFEHR: Zurich, Switzerland, 2021. [CrossRef]
3. United Nations Office for Disaster Risk Reduction (UNISDR). *Sendai Framework for Disaster Risk Reduction 2015–2030*; UNISDR: Geneva, Switzerland, 2015.
4. United Nations Office for Disaster Risk Reduction (UNISDR). *Technical Guidance for Monitoring and Reporting on Progress in Achieving the Global Targets of the Sendai Framework for Disaster Risk Reduction: Collection of Technical Notes on Data and Methodology*; UNISDR: Geneva, Switzerland, 2017.
5. Bruneau, M.; Chang, S.E.; Eguchi, R.T.; Lee, G.C.; O'Rourke, T.D.; Reinhorn, A.M.; Shinozuka, M.; Tierney, K.; Wallace, W.A.; von Winterfeldt, D. A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthq. Spectra* **2003**, *19*, 733–752. [CrossRef]
6. Cimellaro, G.P.; Reinhorn, A.M.; Bruneau, M. Framework for analytical quantification of disaster resilience. *Eng. Struct.* **2010**, *32*, 3639–3649. [CrossRef]
7. Hosseini, S.; Barker, K.; Ramirez-Marquez, J.E. A review of definitions and measures of system resilience. *Reliab. Eng. Syst. Saf.* **2016**, *145*, 47–61. [CrossRef]
8. *FEMA P-2090/NIST SP-1254*; Federal Emergency Management Agency (FEMA); National Institute of Standards and Technology (NIST). Recommended Options for Improving the Built Environment for Post-Earthquake Reoccupancy and Functional Recovery Time. NIST: Gaithersburg, MD, USA, 2021. [CrossRef]
9. Molina Hutt, C.; Vahanvaty, T.; Kourehpaz, P. An analytical framework to assess earthquake-induced downtime and model recovery of buildings. *Earthq. Spectra* **2022**, *38*, 1283–1320. [CrossRef]
10. Gutiérrez, J.; Ayala, A.G.; Rangel-Ramirez, J.G.; Amezcua, H.R.; González-Calva, C. A probabilistic framework for assessing seismic functionality loss and recovery in buildings using a spatially explicit functionality metric. *Bull. Earthq. Eng.* **2026**, *24*, 7203–7236. [CrossRef]
11. Mieler, M.; Stojadinovic, B.; Budnitz, R.; Comerio, M.; Mahin, S. A framework for linking community-resilience goals to specific performance targets for the built environment. *Earthq. Spectra* **2015**, *31*, 1267–1283. [CrossRef]
12. Davis, C.A.; Johnson, L.A.; Kiremidjian, A.; Kwasinski, A.; O'Rourke, T.D.; Stanley, E.; Yu, K.; Zareian, F.; Johnson, K.; Hortacsu, A. *Initial Framework to Design Lifeline Infrastructure for Post-Earthquake Functional Recovery*; NIST SP 1310; NIST: Gaithersburg, MD, USA, 2024; Volume 1. [CrossRef]
13. Davis, C.A.; Johnson, L.A.; Kiremidjian, A.; Kwasinski, A.; O'Rourke, T.D.; Stanley, E.; Yu, K.; Zareian, F.; Johnson, K.; Hortacsu, A. *Initial Framework to Design Lifeline Infrastructure for Post-Earthquake Functional Recovery*; NIST SP 1311; NIST: Gaithersburg, MD, USA, 2024; Volume 2. [CrossRef]
14. Kaur, K.; Molina Hutt, C.; Kim, A.M. Assessing post-earthquake emergency healthcare accessibility considering damage to hospital buildings and transportation infrastructure. *Earthq. Spectra* **2026**, *42*, e70024. [CrossRef]
15. Palomino Romani, G.; Blowes, K.; Molina Hutt, C. Evaluating post-earthquake functionality and surge capacity of hospital emergency departments using discrete event simulation. *Earthq. Spectra* **2023**, *39*, 402–433. [CrossRef]

16. Pei, S.-S.; Zhai, C.-H.; Hu, J.; Wang, Z.-Q.; Xie, L.-L. Resilience assessment and enhancement of interdependent transportation-healthcare system: A spatial accessibility approach. *Transp. Res. Part D Transp. Environ.* **2024**, *128*, 104090. [CrossRef]
17. van de Lindt, J.W.; Kruse, J.; Cox, D.T.; Gardoni, P.; Lee, J.S.; Padgett, J.; McAllister, T.P.; Barbosa, A.; Cutler, H.; Van Zandt, S.; et al. The interdependent networked community resilience modeling environment (IN-CORE). *Resilient Cities Struct.* **2023**, *2*, 57–66. [CrossRef]
18. Li, L.; Chang-Richards, A.; Boston, M.; Elwood, K.; Molina Hutt, C. Modelling post-earthquake building recovery under human resource constraints. *Int. J. Disaster Risk Reduct.* **2025**, *122*, 105389. [CrossRef]
19. Magoua, J.J.; Huang, J.; Kwak, Y.H.; Li, N. Collaborative decision making for natural disaster recovery of interdependent infrastructure systems. *Reliab. Eng. Syst. Saf.* **2026**, *272*, 112472. [CrossRef]
20. Cremen, G.; Galasso, C.; McCloskey, J. A simulation-based framework for earthquake risk-informed and people-centered decision making on future urban planning. *Earth's Future* **2022**, *10*, e2021EF002388. [CrossRef]
21. Huang, C.; Molina Palacios, S.; Meslem, A. Development of a new tool for seismic risk assessment and multi-criteria decision making. *Int. J. Disaster Risk Reduct.* **2024**, *106*, 104261. [CrossRef]
22. Sediek, O.A.; Roohi, M.; van de Lindt, J.W. A genetic algorithm framework for seismic retrofit of building portfolios to enhance community resilience. *Int. J. Disaster Risk Reduct.* **2024**, *108*, 104570. [CrossRef]
23. Kavvada, I.; Horvath, A.; Moura, S. Distributionally robust budget allocation for earthquake risk mitigation in buildings. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part A Civ. Eng.* **2024**, *10*, 04023050. [CrossRef]
24. Du, A.; Wang, X.; Xie, Y.; Dong, Y. Regional seismic risk and resilience assessment: Methodological development, applicability, and future research needs—An earthquake engineering perspective. *Reliab. Eng. Syst. Saf.* **2023**, *233*, 109104. [CrossRef]
25. Zhang, X.; Sun, L.; Chang, L.; Sun, Z. Natural hazards, critical infrastructure systems, and resilience enhancement: A structured state-of-the-art review and a synthesis-based rapid assessment framework. *Nat. Hazards Rev.* **2027**, *28*, 03126011. [CrossRef]
26. Chen, Q.; Tan, Y.; Zhang, G.; Hou, L. Critical infrastructure resilience under natural disasters: Current methods, challenges and future directions. *Int. J. Crit. Infrastruct. Prot.* **2026**, *55*, 100887. [CrossRef]
27. Elkady, S.; Hernantes, J.; Labaka, L. Decision-making for community resilience: A review of decision support systems and their applications. *Heliyon* **2024**, *10*, e33116. [CrossRef] [PubMed]
28. Cremen, G. A new end-user-oriented and dynamic approach to post-disaster resilience quantification for individual facilities. *Risk Anal.* **2025**, *45*, 701–709. [CrossRef] [PubMed]
29. Poulin, C.; Kane, M.B. Infrastructure resilience curves: Performance measures and summary metrics. *Reliab. Eng. Syst. Saf.* **2021**, *216*, 107926. [CrossRef]
30. Logan, T.M.; Guikema, S.D. Reframing resilience: Equitable access to essential services. *Risk Anal.* **2020**, *40*, 1538–1553. [CrossRef] [PubMed]
31. Coleman, N.; Li, X.; Comes, T.; Mostafavi, A. Weaving equity into infrastructure resilience research: A decadal review and future directions. *npj Nat. Hazards* **2024**, *1*, 25. [CrossRef]
32. McPhail, C.; Maier, H.R.; Kwakkel, J.H.; Giuliani, M.; Castelletti, A.; Westra, S. Robustness metrics: How are they calculated, when should they be used and why do they give different results? *Earth's Future* **2018**, *6*, 169–191. [CrossRef]
33. Controller; Auditor-General. *Effectiveness and Efficiency of Arrangements to Repair Pipes and Roads in Christchurch–Follow-Up Audit*; B.29[16h]; Office of the Auditor-General: Wellington, New Zealand, 2016; Available online: https://oag.parliament.nz/2016/scirt/docs/scirt-follow-up.pdf (accessed on 30 October 2026).
34. Tehler, H.; Cedergren, A.; de Goër de Herve, M.; Gustavsson, J.; Hassel, H.; Lindbom, H.; Nyberg, L.; Wester, M. Evidence-based disaster risk management: A scoping review focusing on risk, resilience and vulnerability assessment. *Prog. Disaster Sci.* **2024**, *23*, 100335. [CrossRef]
35. Alexander, D.E. On evidence-based practice in disaster risk reduction. *Int. J. Disaster Risk Sci.* **2021**, *12*, 919–927. [CrossRef]
36. Alonso-Coello, P.; Schünemann, H.J.; Moberg, J.; Brignardello-Petersen, R.; Akl, E.A.; Davoli, M.; Treweek, S.; Mustafa, R.A.; Rada, G.; Rosenbaum, S.; et al. GRADE Evidence to Decision (EtD) frameworks: A systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ* **2016**, *353*, i2016. [CrossRef] [PubMed]
37. Senerth, E.; Whaley, P.; Akl, E.; Beverly, B.; Alonso-Coello, P.; Rooney, A.; Schünemann, H.J.; Thayer, K.A.; Tsaioun, K.; Morgan, R.L. GRADE guidance 40: The GRADE evidence-to-decision framework for environmental and occupational health. *Environ. Int.* **2025**, *197*, 109314. [CrossRef] [PubMed]
38. Kelly, T.; Weaver, R. The Goal Structuring Notation—A safety argument notation. In Proceedings of the Dependable Systems and Networks 2004 Workshop on Assurance Cases, Florence, Italy, 1 July 2004.
39. Assurance Case Working Group (ACWG). *Goal Structuring Notation Community Standard, Version 3*; SCSC-141C; Safety-Critical Systems Club: York, UK, 2021.
40. Diemert, S.; Goodenough, J.B.; Joyce, J.; Weinstock, C.B. Incremental assurance through eliminative argumentation. *J. Syst. Saf.* **2023**, *58*, 7–15. [CrossRef]

41. Graydon, P.J.; Holloway, C.M. An investigation of proposed techniques for quantifying confidence in assurance arguments. *Saf. Sci.* **2017**, *92*, 53–65. [CrossRef]
42. Haasnoot, M.; Kwakkel, J.H.; Walker, W.E.; Ter Maat, J. Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Glob. Environ. Change* **2013**, *23*, 485–498. [CrossRef]
43. Weimer, J.; Sokolsky, O.; Bezzo, N.; Lee, I. Towards assurance cases for resilient control systems. In Proceedings of the 2nd IEEE International Conference on Cyber-Physical Systems, Networks, and Applications (CPSNA), Hong Kong, China, 25–26 August 2014. [CrossRef]
44. Campean, F.; Kabir, S.; Dao, C.; Zhang, Q.; Eckert, C. Towards a resilience assurance model for robotic autonomous systems. In Proceedings of the International Conference on Engineering Design (ICED21), Gothenburg, Sweden, 16–20 August 2021. [CrossRef]
45. Xekalakis, G.; Nikolaieva, O.; Kalnis, G.; Christou, P. Fragility-based seismic resilience and recovery assessment for municipal emergency planning in Mediterranean urban environments. *Front. Earth Sci.* **2026**, *14*, 1872159. [CrossRef]
46. Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
47. Whittemore, R.; Knafl, K. The integrative review: Updated methodology. *J. Adv. Nurs.* **2005**, *52*, 546–553. [CrossRef] [PubMed]
48. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]
49. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef] [PubMed]
50. Wohlin, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14), London, UK, 13–14 May 2014*; ACM: New York, NY, USA, 2014; p. 38. [CrossRef]
51. Cohen, J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* **1968**, *70*, 213–220. [CrossRef] [PubMed]
52. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef]
53. Stojadinović, Z.; Marinković, M.; Nedeljković, Đ.; Ivanović, M.; Stojadinović, B.; Marinković, D.; Simić, N. From observed damage to executable repair quantities: The RELA Repair Module. *Earthq. Spectra* **2026**, *42*, e70107. [CrossRef]
54. Cremen, G.; Galasso, C.; McCloskey, J.; Barcena, A.; Creed, M.; Filippi, M.E.; Gentile, R.; Jenkins, L.T.; Kalaycioglu, M.; Mentese, E.Y.; et al. A state-of-the-art decision-support environment for risk-sensitive and pro-poor urban planning and design in Tomorrow's cities. *Int. J. Disaster Risk Reduct.* **2023**, *85*, 103400. [CrossRef]
55. Mazumder, R.K.; Rana, S.; Wang, X.; Ansary, M.A.; Li, Y. Community-level seismic risk and retrofitting portfolios in developing countries: Cost-effective assessment framework and application. *Nat. Hazards Rev.* **2025**, *26*, 05025003. [CrossRef]
56. Tao, Y.; Xu, Z.-D.; Liu, X.-Y.; Wei, Y.; Peng, H. Uncertainty analysis and resilience enhancement for post-earthquake fire emergency response considering cascading effects of interdependent infrastructures. *Sustain. Cities Soc.* **2026**, *149*, 107859. [CrossRef]
57. Yavuz, H.I.; Baraçlı, H. A dynamic multi-objective model for district-level post-earthquake resource allocation integrating social vulnerability, occupational safety, and Markov-based updating: An Istanbul case study. *Appl. Sci.* **2026**, *16*, 4425. [CrossRef]
58. Mager Pozo, A.M.; Priesmeier, P.; Fekete, A. Measuring spatial accessibility to critical infrastructure: The Access Road Identification model. *Int. J. Crit. Infrastruct. Prot.* **2025**, *49*, 100760. [CrossRef]
59. Miles, S.B.; Chang, S.E. Modeling community recovery from earthquakes. *Earthq. Spectra* **2006**, *22*, 439–458. [CrossRef]
60. Xu, M.; Li, G.; Chen, A. Resilience-driven post-disaster restoration of interdependent infrastructure systems under different decision-making environments. *Reliab. Eng. Syst. Saf.* **2024**, *241*, 109599. [CrossRef]
61. Cremen, G.; Galasso, C. A decision-making methodology for risk-informed earthquake early warning. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *36*, 747–761. [CrossRef]
62. Ding, J.-W.; Cheng, Q.-L.; Tian, Y.; Xu, M.-Y.; Cao, Z.-G.; Lu, D.-G. A comprehensive seismic resilience assessment framework for urban building portfolios considering building functional importance. *Reliab. Eng. Syst. Saf.* **2027**, *277*, 113215. [CrossRef]
63. Atabay, S.; Ozay, R.; Yilmaz, D.; Yilmaz, I.C. A statistical prioritization framework for earthquake-induced urban infrastructure damage factors and mitigation measures. *Buildings* **2026**, *16*, 2323. [CrossRef]
64. Mograby, M.; Iaaly, A. Evidence admission in spatial decision models: Gating layers before fusion in municipal seismic screening. *ISPRS Int. J. Geo-Inf.* **2026**, *15*, 425. [CrossRef]
65. Mucedero, G.; Monteiro, R. An integrated regional prioritisation framework for seismic and energy-efficiency performance upgrading of residential buildings. *Int. J. Disaster Risk Reduct.* **2024**, *103*, 104341. [CrossRef]
66. Parizi, S.M.; Taleai, M.; Sharifi, A. A GIS-based multi-criteria analysis framework to evaluate urban physical resilience against earthquakes. *Sustainability* **2022**, *14*, 5034. [CrossRef]
67. Petrini, F.; Pedone, L.; Francioli, M.; Matteoni, M.; Pampanin, S. Seismic risk assessment of urban areas: An integrated multi-scale and multi-refinement framework. *Int. J. Disaster Risk Reduct.* **2026**, *138*, 106101. [CrossRef]

68. Rezvani, S.M.H.S.; Falcão Silva, M.J.; Marques de Almeida, N. The Risk-Informed Asset-Centric (RIACT) urban resilience enhancement process: An outline and pilot-case demonstrator for earthquake risk mitigation in Portuguese municipalities. *Appl. Sci.* **2024**, *14*, 634. [CrossRef]
69. Lempert, R.J. Robust Decision Making (RDM). In *Decision Making Under Deep Uncertainty: From Theory to Practice*; Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M., Popper, S.W., Eds.; Springer: Cham, Switzerland, 2019; pp. 23–51. [CrossRef]
70. Baek, S.; Kim, J.; Park, C.; Lee, T. Evidence lost in transition: Climate-information continuity failures in South Korea's local adaptation planning and implementation. *Clim. Serv.* **2026**, *42*, 100649. [CrossRef]
71. Alcántara-Ayala, I.; Velásquez-Espinoza, G.; Montes de Jesús, A. From mandates to mechanisms: Institutional vulnerability, decentralized governance, and the challenges of local disaster risk reduction implementation. *Int. J. Disaster Risk Sci.* **2025**, *16*, 709–723. [CrossRef]
72. Cash, D.W.; Clark, W.C.; Alcock, F.; Dickson, N.M.; Eckley, N.; Guston, D.H.; Jäger, J.; Mitchell, R.B. Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8086–8091. [CrossRef] [PubMed]
73. Becker, A. Using boundary objects to stimulate transformational thinking: Storm resilience for the Port of Providence, Rhode Island (USA). *Sustain. Sci.* **2017**, *12*, 477–501. [CrossRef]
74. United Nations General Assembly. *Transforming Our World: The 2030 Agenda for Sustainable Development*; A/RES/70/1; United Nations: New York, NY, USA, 2015.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.