Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Evidence-Based Risk Governance Framework for Public-Private Partnerships in Saudi Arabia: An Uncertainty-Aware Prioritization Framework

Musab Hamed ALshammari

musab.alshammri@gmail.com

Mansoura University

Fikry Abdou Salem

Mansoura University

Mohamed Osami Rageh

Mansoura University

Dina M. Mansour

Future University in Egypt

Research Article

Keywords: Evidence-based governance, Public–private partnerships, Risk prioritization, Saudi Arabia, Uncertainty-aware analysis

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11284489/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Research article

An Evidence-Based Risk Governance Framework for Public-Private Partnerships in Saudi Arabia: An Uncertainty-Aware Prioritization Framework

Musab Hamed ALshammari [1,*], Fikry Abdou Salem [1], Mohamed Osami Rageh [1] and Dina M. Mansour [2]

[1] Structural Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt; musab.alshammri@gmail.com (MHA), drfikry_salem@yahoo.com (FAS), eng.m_osami@mans.edu.eg (MOR)

[2] Structural Engineering and Construction Management Department, Faculty of Engineering & Technology, Future University in Egypt, 90th St., Fifth Settlement, New Cairo, 11835, Egypt; dmohamed@fue.edu.eg (DMM)

* Correspondence: musab.alshammri@gmail.com

Abstract

Saudi Arabia's expanding use of public–private partnerships (PPPs) requires risk priorities that are locally grounded, uncertainty-aware, and operationally actionable. This study developed an auditable evidence-to-governance approach combining literature normalization, stakeholder assessment, quantitative uncertainty analysis, and qualitative contextualization. A 43-source corpus was classified; a 15-source item archive yielded 519 statements, of which 495 were factor-level, 14 contextual, and 10 non-factor-level. Fourteen recurrence-eligible sources supported normalization into 41 factors and selection of 15 risks for a probability–impact questionnaire completed by 205 Saudi PPP stakeholders. Respondent-level probability–impact products were analyzed using Friedman and Holm-adjusted Wilcoxon tests, participant-vector bootstrap resampling, dominance probabilities, tied-rank literature–market alignment, subgroup tests, and sensitivity scenarios. Political and security instability ranked first (CI = 0.5243; 95% bootstrap confidence interval: 0.4835–0.5647), followed by cost escalation and overrun (CI = 0.5173) and coordination and communication failure (CI = 0.5151). Fourteen risks formed the Tier 1 set, while regulatory and legal uncertainty remained Tier 2. Literature recurrence and market criticality showed positive but non-significant alignment (Spearman's $\rho = 0.400$; permutation $p = 0.1403$). Rankings were stable after excluding straightliners and remained highly correlated across prespecified sensitivity scenarios. Open-ended responses yielded 12 themes, including three emergent candidates, which informed proposed controls and key risk indicators without altering quantitative ranks. The resulting Proposed Evidence-Based Saudi PPP Risk Governance Framework links the full risk register to lifecycle ownership, contractual allocation, response controls, monitoring triggers, escalation, and residual-risk review, providing a transparent basis for project-specific governance.

Keywords: Evidence-based governance; Public–private partnerships; Risk prioritization; Saudi Arabia; Uncertainty-aware analysis**.**

1. Introduction

Public–private partnerships (PPPs) leverage private financing and specialized delivery capability to provide public infrastructure. However, they also give rise to long-term contracting systems in which financial, political, regulatory, market, technical and managerial risks interact across project phases and organizational boundaries. PPP literature reviews consistently demonstrate that risk identification and allocation are central to project performance . But they also show significant variation in terminology, factor boundaries, sectors and analytical practice . Such heterogeneity means that the much-cited risk is neither automatically comparable across studies, nor necessarily relevant in a particular national market.

Empirical evidence supports the importance of context. Water-sector PPPs in Ghana, studies of infrastructure delivery methods, stakeholder analyzes of sustainable PPP outcomes, and collaborative operations research suggest that risk salience is sector-, delivery arrangement-, institutional setting-, project stage-, and respondent position dependant [5–8]. A credible system of priorities must therefore keep track of the source of each risk expression, combine only substantively equivalent mechanisms, and separate literature recurrence from the probability and impact assessed for the field.

The problem with governance is not just to produce a ranked checklist. Integrated risk-management systems, performance frameworks, allocation models and operational studies are indicative of ongoing ownership, monitoring, response and reallocation throughout the project life cycles [9-12]. Political uncertainty further alters the contractual confidence and conditions of implementation. Evidence from independent water and power plant projects in Saudi Arabia identifies material concerns of institutional capability, informal implementation, and weak risk-management practice [13–15]. These findings imply that analytical priorities are only useful if they are translated into stage-specific accountability, controls, early-warning indicators, escalation rules and post-control review.

The latest PPP research has involved more comprehensive managerial measures, improved success indicators, simulation based on sustainability, knowledge-driven qualitative assessment and structural-equation modeling [16–20]. Such advances improve particular aspects of assessment or decision support, but do not themselves address how an inventory derived from the literature should be normalized, tested against local stake-holder evidence, ranked under uncertainty, and transformed into a transparent governance register. The question is therefore not just what risks exist but what risks are critical in the target market, how stable their ordering is and what governance response follows from that evidence.

This is particularly the case in Saudi Arabia, where there is a growing application of PPP in the social and economic infrastructure. Current Saudi studies provide useful yet sector-specific evidence on sustainable housing, school infrastructure and PPP success factors [21–23]. Similarly, studies of PPPs for transport, housing and water in other

developing contexts, along with a recent synthesis of PPPs in construction, expand the comparative base [24–27], while research on privatization in Saudi Arabia highlights tendering, liabilities and post-transfer governance as ongoing implementation issues [28]. Collectively, this literature supports contextualization, but does not provide a unified uncertainty aware ranking and governance framework for the Saudi PPP market.

Therefore, research on adjacent construction and infrastructure was used as background evidence rather than as a substitute for the recurrence of PPP-specific factors. Process and capability interpretation is informed by studies of BIM-enabled risk assessment [29], Saudi sustainable-construction capability [30], and cross-country building barriers and operation-and-maintenance information requirements [31,32]. Evidence on high-value contracting, process mining, infrastructure health and safety, road-project financing and housing cost and resource control inform the design, financing and control lenses [33–37]. Additional work on BIM-based safety, data-driven construction risk, Saudi e-procurement, dynamic safety resilience, and oil-and-gas construction further clarifies monitoring and implementation conditions [38–42], while early research on large build-operate-transfer projects establishes the continued relevance of holistic life-cycle risk planning [43-44]. These sources supplement interpretation and framework design, but are not permitted to inflate the study-level frequency of the 14 recurrence-eligible normalization sources.

There is still a compound evidence-to-governance gap. First, literature reviews tend to lump together the whole body of evidence, the part with lists of factors at the individual factor level and the repetition of wordings in individual studies without distinguishing the unit of analysis. Second, probability and impact are often summarized with point estimates or a product of means, which can lose respondent-level pairing, and hide rank uncertainty. Third, the comparison of local priorities with recurrence in the literature by tied-rank and resampling inference is rarely performed, while subgroup stability and open-ended contextual evidence are seldom analyzed. Finally, rankings are rarely translated into an operational register that relates each retained risk to life-cycle stage, accountable owner, contractual allocation, prevention, mitigation and contingency actions, key risk indicators (KRIs), trigger thresholds, escalation and residual-risk review.

This study addresses this gap using a sequential, multi-stage design that is primarily quantitative and with an embedded qualitative strand. An audit and classification of a final 43-source evidence corpus was conducted. In it, a 15-source item-level archive had 519 verbatim statements; record-level re-adjudication classified 495 as factor-level risk expressions, 14 as contextual indicators, and 10 as non-factor-level labels. Fourteen sources had at least one factor-level mechanism that could be recurrentized as one occurrence per study (i.e., defined the normalization denominator). Semantic normalization aggregated the factor-level evidence into 41 risk factors, from which 15 conceptually distinct and Saudi relevant risks were retained in the locked probability-impact questionnaire completed by 205 Saudi PPP stakeholders. The analysis preserves respondent-level probability-impact pairing, uses repeated-measures inference, employs participant-vector bootstrap resampling, establishes dominance-based priority tiers, compares literature and market evidence, tests subgroup and scoring robustness, and maps open-ended themes without allowing qualitative frequency to alter the quantitative rank. The integrated results are translated into a Proposed Evidence-Based Saudi PPP Risk Governance Framework with the governance fields reported as proposed and project-adaptable specifications.

The study thus asks which of the literature-derived risks are the most critical for Saudi PPP projects when considering probability and impact together and how stable their ranks and priority tiers are? Does study-level recurrence match criticality as evaluated by stakeholders? Are there differences in priorities between stakeholder and experience groups or with different samples and scoring rules? What contextual extensions or emergent concerns arise from the open-ended responses? How can combined evidence be translated into stage-specific governance? These questions link evidence synthesis, local market assessment, uncertainty, contextual interpretation, and implementation into one traceable design.

The study provides an auditable evidence-reduction pathway that separates the 43-source corpus, the 15-source item archive, the 14-source recurrence-eligible subset, the 519 raw statements and their final dispositions, the 41-factor register, and the 15-factor questionnaire shortlist. It advances PPP risk prioritization by preserving within-respondent probability-impact dependence, and by reporting rank stability and dominance, rather than imposing an arbitrary score cut-off. It also differentiates literature evidence from Saudi market evidence to avoid double counting, subjects ordering to pre-specified subgroup and sensitivity analyzes and uses open-ended findings to enhance interpretations, controls and KRIs without changing quantitative tiers. Most of all it translates the retained evidence into a proposed six-stage governance register with explicit ownership, response, monitoring, escalation and residual-risk logic, whilst retaining the uncertainty and analytical role of every evidence layer in sight.

2. Methodology

2.1. Research Design

This study employs a sequential multi-stage design, predominantly quantitative with an embedded qualitative strand (Figure 1). Each output feeds into the next stage in the sequence. First, we build a corpus of evidence from 43 sources. A 15-source item-level archive contributed 519 verbatim statements, 495 of which were adjudicated as factor-level, 14 as contextual indicators, and 10 as non-factor-level labels. 14 recurrence-eligible sources formed a denominator for normalization into 41 factors. The locked questionnaire consisted of fifteen literature-informed,

Saudi-relevant and conceptually distinct factors. Second, the selected factors have been evaluated by a probability-impact questionnaire responded by 205 Saudi PPP stakeholders. An open-ended item enabled the collection of additional risks and contextual observations not fully captured by the closed questions.

The quantitative phase involves data-quality screening, descriptive analysis, probability and impact relative importance indices, respondent-level (P × I) criticality assessment, response consistency testing, matched-factor comparisons, and boot-strap-based evaluation of ranking uncertainty. Literature frequency is then compared to field-derived criticality using tied-rank Spearman correlation, permutation inference, and an evidence-alignment matrix. Robustness of the resulting priorities is examined through subgroup and sensitivity analyzes, while the open-ended responses are systematically coded and thematically integrated with the quantitative findings. Finally, the integrated evidence is translated into a Proposed Evidence-Based Saudi PPP Risk Governance Framework linking each retained risk to its lifecycle stage, accountable owner, supporting actors, contractual allocation, three-layer response actions, key risk indicator, trigger threshold, escalation rule, and residual-risk review.



Fig. 1 Integrated research design for evidence-based Saudi PPP risk governance. Integrated multi-stage research design for synthesizing literature and Saudi field evidence, prioritizing PPP risks under uncertainty, integrating open-ended findings, and translating the combined evidence into a proposed Saudi PPP risk-governance framework.

2.2. Literature Evidence Base, Risk Extraction, and Normalization

This phase used a nested evidence architecture, starting with the earliest layer for which a full record level inventory is maintained. The auditable starting point was a corpus of 43 archived full-text sources, together with a stable code and mutually exclusive analytical role for each. In this corpus, 15 sources generated 519 item-level statements, of which 495 were categorized as factor-level, 14 as contextual indica-tors, and 10 as non-factor-level. Fourteen sources had at least one factor-level mechanism eligible for recurrence. Semantic consolidation resulted in 41 normalized factors, 15 of which were kept in the locked questionnaire. These layers and denominators are differentiated in Table 1, Figure 2 and Supplementary Figure S1.

The 43-source corpus was informed by earlier searches of Scopus, Web of Science, ScienceDirect, the MDPI publishing platform, and Google Scholar for English-language PPP risk evidence published during 2010–2025. The search concepts retained were public private partnerships, operational risk or risk management, infrastructure or construction projects and lifecycle operation or sustainability. These details are reported as historical search provenance, not as a reconstructed present-study search flow. The core Boolean logic, tailored to the syntax of the various platforms, was:

Core Boolean logic: ("public-private partnership" OR PPP OR P3) AND ("operational risk" OR "risk factor" OR "risk management" OR "risk allocation") AND (construction OR infrastructure OR project) AND (operation OR lifecycle OR sustainab).

Eligibility time frame: 1 January 2010 – 31 December 2025. For eligibility, publication year was defined as year of earliest verifiable public availability. For journal articles published online ahead of assignment to a later volume or issue, eligibility was based on the official publisher's first online date, but the bibliographic reference retained the final issue year. Based on this rule, Alghuried [30] qualified as its first online publication date was 30 January 2025, although it was later assigned to a 2026 issue. One source predating the windows, Yaworsky [43] (1994),

was kept as an explicitly declared foundational/contextual exception, as it provides a holistic lifecycle framing for large and BOT projects. This exception was not applicable to item-level extraction, normalized-factor recurrence, F i, LP i, literature ranking, or questionnaire-shortlist selection. The auditable corpus therefore comprised 42 records conforming to the 2010–2025 rule and one disclosed pre-window contextual exception.

The eligibility window under the earliest-public-availability rule closed on 31 December 2025. The archived full-text inventory was consolidated and frozen on 15 August 2026. There is no complete list of all upstream candidates at the record level , database - specific query exports , duplicate - removal decisions , or intermediate screening counts in the project archive . The present study reproducible source audit begins with the 43 actually retained and re-evaluated full-text records. Record uniqueness was evaluated by stable study codes, source file names, and short titles. All 43 records and their final fates are reported in Supplementary Table S1A and the source-flow decision trail is reported in Supplementary Data S1H.

Table 1 Composition and analytical roles of the literature evidence layers.

Evidence layer / status	Unit and n	Eligibility or definition	Analytical role
Auditable full-text corpus	43 sources	42 records meeting the 2010–2025 rule plus one declared pre-window contextual exception	Present-study record-level starting point: 14 Normalization, 13 Supporting, 16 Contextual
Window-eligible records	42 sources	Earliest verifiable public availability occurred from 1 January 2010 to 31 December 2025	Eligible for role assignment; only Normalization sources can enter recurrence
Online-first eligibility case	1 source (E30[1])	Alghuried: first online 30 January 2025; final issue assignment 2026	Within window; Contextual role; no F_i/LP_i contribution
Declared pre-window exception	1 source (E43[1])	Yaworsky (1994), retained only as foundational/contextual BOT lifecycle framing	No item extraction, F_i, LP_i, rank, or shortlist contribution
Item-level extraction archive	15 sources / 519 statements	Traceable lists, indicators, groups, or appraisal items	Preserves every verbatim item; includes E12 Supporting appraisal criteria
Recurrence-eligible subset	14 sources	At least one specific factor-level mechanism suitable for binary source–factor coding	Only denominator used for F_i and LP_i
Normalized register / questionnaire shortlist	41 / 15 factors	Distinct mechanisms / literature-led Saudi PPP field shortlist	Full register / locked field instrument

The 43 records in the archive were reassessed at the title/abstract and full-text levels in accordance with the purpose of the present study. A record was eligible for detailed assessment if it met the 2010–2025 earliest-public-availability rule or the disclosed pre-window con-textual-exception rule, was published in English, covered PPPs or directly transferable infrastructure/construction risk, governance, allocation, or operational-performance issues, and included traceable evidence associated with a stable source code. We excluded from pooling records that had mismatched content, were not about infrastructure, had editorial or methodological content without extractable risk evidence, or that had no usable factor-level material. The measures, critical success factors, latent constructs, and non-PPP contextual barriers were not silently mixed with direct risk inventories. They were kept as supporting or contextual lenses.

The auditable starting corpus was 43 archived full-text sources. Forty-two records met the 2010–2025 temporal rule, and one was designated as a pre-window foundational/contextual exception. One final analytical role was assigned to each source that was mutually exclusive – 14 Normalization, 13 Supporting, or 16 Contextual. Only the 14 Normalization sources were entered into the study-level recurrence denominator. Interpretation, triangulation, or conceptual framing drew on supporting and contextual sources, but these did not contribute to F_i or LP_i. Exclusion reasons for 16 records located in legacy working files outside this corpus are separately retained in Supplementary Table S1E. Their upstream search-stage positions are not preserved, they are not identified as the former 19, they are not added to the 43 records to reconstruct a larger source total, and they do not enter any analytical denominator. A numerical risk-of-bias score was not used because evidence was not pooled to estimate a causal effect. Alternatively, relevance and extractability were assessed based on PPP specificity, risk evidence directness, recoverability at the factor level, source traceability, and applicability for coding of one occurrence per source. Supplementary Data S1A, S1E, S1H and S1I report final roles, the separate legacy log and the temporal decisions.

A traceable list, table, appendix, indicator set, group or allocation-driver set from which verbatim wording could be retrieved and linked to a stable source record was required to enter the item-level archive. Fifteen sources met this archival criterion and contributed exactly 519 statements. The criteria for eligibility for recurrence were more

strict. A source was included in S only if at least one extracted statement described a sufficiently specific risk mechanism that was amenable for binary source-factor coding after consolidation. Four of the E12 records were allocation/appraisal criteria, not factor-level mechanisms. E12 is still traceable in the item archive but reclassified as Supporting. Thus, the normalization subset eligible for recurrence contained 14 sources . Source-specific contributions and attributions are shown in Supplementary Table S1B.

Each extracted statement was maintained as a separate record with a stable expression identifier, source code, verbatim wording, source title, item identifier if available, source category or domain, and extraction note. The corrected crosswalk includes the original assignment and final disposition, including the assigned factor code, normalized domain, lifecycle indication, recurrence-eligibility flag, and record-level rationale. The full 519-row crosswalk is provided in Supplementary Data S1C and all 134 former Other/paper specific records are reiterated in the focused audit log in Supplementary Data S1G.

Only expressions representing the same underlying risk mechanism were combined. The coding rules distinguished a cause from its consequence, financial from contractual risk, delay as an event from its causal drivers, and political/governmental from legal/regulatory mechanisms. All 134 records previously assigned to Other/paper-specific factor were re-reviewed (record-level re-adjudication). Of these, 110 were reallocated as factor-level expressions within the existing 41-factor register, 14 were mapped as contextual indicators but excluded from recurrence, and 10 were classified as non-factor-level umbrella, appraisal, consequence-only or analytic labels. The final dispositions across the entire archive were thus 495 factor-level expressions, 14 contextual indicators, and 10 non-factor-level statements. The analytical register has no residual category of Other left. Supplementary Data S1C and S1G are unchanged in their final decisions and rationale.

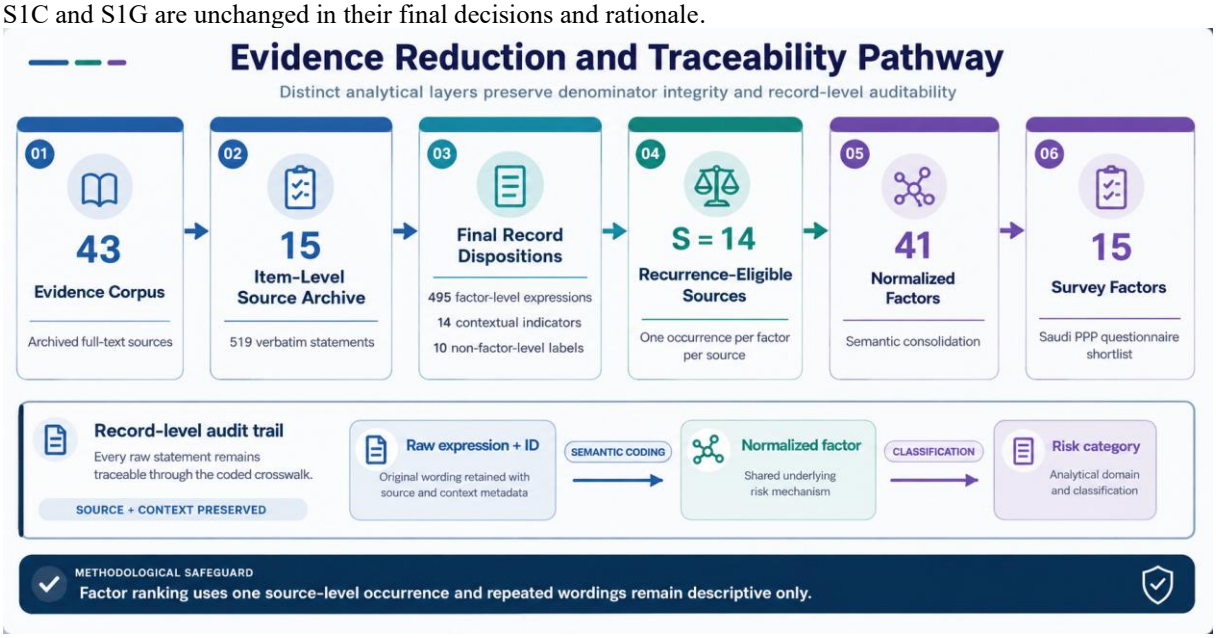


Fig. 2 Auditable literature-evidence pathway and questionnaire shortlisting. Auditable pathway from the 43-source evidence corpus and 15-source item-level archive to the final record dispositions, the 14-source recurrence-eligible subset, the 41-factor normalized register, and the 15-factor questionnaire shortlist.

Semantic assignment and re-adjudication at the record level were done by a single coder. There was no independent 20–25% recoding exercise documented on the 519-record mapping, and Cohen's kappa is not reported, no inter-coder reliability claim is made. Instead, reproducibility is enabled by the comprehensive source-to-statements-to-factors crosswalk, transparent eligibility and coding rules, the targeted 134-record decision log, and version-controlled analysis files. Methodological limitation: absence of independent semantic recoding.

The unit of recurrence was the independent normalization source, not the number of repeated words or items within a source. Occurrence was coded once for normalized factor i and normalization source s through the indicator function:

$$z_{si} = \mathbf{1}\{\text{factor } i \text{ occurs in study } s\} \tag{1}$$

Here, $\mathbf{1}\{\cdot\} = 1$ when the stated condition is satisfied and 0 otherwise. The unique-study frequency was:

$$F_i = \sum_{s=1}^{S} z_{si} \tag{2}$$

Literature prevalence was then expressed as:

$$\mathrm{LP}_i = \frac{F_i}{S} \times 100, \qquad S = 14 \tag{3}$$

Here, F_i is the number of recurrence-eligible normalization sources containing factor i, and LP_i is its literature prevalence among S = 14 sources. Raw-expression counts, contextual-indicator counts, and non-factor-level labels were retained as descriptive audit metadata only; they did not influence ordering because one source may use several statements for the same mechanism and non-factor records cannot create a factor occurrence.

Questionnaire shortlisting followed the documented sequential rule: study-level recurrence first, followed by relevance to Saudi PPP conditions, conceptual non-overlap, and balanced domain coverage. After record-level adjudication, 12 factors had $F_i \geq 9$ and four factors were tied at the boundary with $F_i = 8$ ($LP_i = 57.1\%$; average rank = 14.5): R09 Demand and market risk, R10 Approval and permitting delays, R14 Tax and tariff changes, and N17 Financing and liquidity constraints. Three positions were available. R09, R10, and R14 were retained because they preserve distinct market-demand, institutional-approval, and fiscal/tariff mechanisms; N17 remained in the 41-factor register because its financing/liquidity mechanism partly overlapped the selected financial and contractual coverage. The literature-only rule was reapplied without survey results or raw-expression counts and reproduced the locked 15-factor instrument. Because this verification followed questionnaire administration, the result is reported as post-audit verification rather than as a newly prospective selection exercise.

2.3. Questionnaire Design, Sampling, and Data Screening

In the second phase, the 15 literature-based risk factors were transformed into a probability-impact questionnaire, and data were collected from a non-probability sample of Saudi PPP stakeholders. The instrument included participant characteristics, separate probability and impact ratings, and an open-ended response. The main analysis was conducted on the 205 valid records, with experience-based subsets and a dataset that excluded straight lining used for benchmarking and sensitivity analysis.

Data were collected in Saudi Arabia through an administered questionnaire using Google Forms (Google LLC, web application; no release-version identifier was record-ed). The timestamps of submissions that were kept are from 17 May 2026 at 22:49:56 to 3 June 2026 at 19:58:55 (GMT+3). Recruitment was by non-probability purposive electronic route to a heterogeneous Saudi PPP stakeholder population. Direct PPP participation was not mandatory; the profile block captured organization type, primary field of expertise, years of professional experience, direct PPP exposure, PPP-infrastructure experience, and self-reported PPP knowledge.

For the purposes of this analysis, a record was considered valid if all six profile fields and all 30 required probability-impact ratings were present and all closed ratings were within the prespecified 1-5 range. This resulted in 205 valid completed questionnaires. Because no defensible denominator of eligible invitees reached was available in the audit archive, a response rate could not be calculated; N = 205 is therefore presented as the number of valid completed questionnaires rather than as a response rate.

The questionnaire included participant and professional-experience, 15 risk factors, a probability and impact rating for each of the risk factors, and an open-ended item for other risks or comments. For K=15 factors and D=2 rating dimensions the number of closed rating items was:

$$N_{\text{rating}} = K \times D = 15 \times 2 = 30 \quad (4)$$

The questionnaire components and their analytical roles are reported in Table 2; the open-ended item and profile fields were not included in the 30-item rating count.

Table 2 Questionnaire structure and measurement scheme.

Component	Content	Items	Response format / role
Participant profile	Professional background and experience	Profile fields	Sample description and subgroup definition
Probability block	Occurrence likelihood for 15 risk factors	15	Separate ordinal scale from 1 to 5
Impact block	Consequence severity for the same 15 factors	15	Separate ordinal scale from 1 to 5
Open-ended item	Additional risks or comments	1	Free-text qualitative response
Closed-item total	15 factors x 2 rating dimensions	30	Primary quantitative dataset

Probability and impact were measured independently on ordinal scales from 1 (lowest) to 5 (highest). For risk factor i and participant j:

$$P_{ij},\ I_{ij} \in \{1,2,3,4,5\}, \quad i = 1,\dots,15;\ j = 1,\dots,205 \quad (5)$$

Here, P_{ij} and I_{ij} denote the probability and impact ratings, respectively. Supplementary File S2 reproduces the exact wording, response anchors, and variable mapping verified against the original survey export, which served as the authoritative source.

Participants were classified as a Saudi PPP stakeholder sample, reflecting heterogeneous levels of direct PPP and infrastructure experience. The 205 valid responses formed the primary analytical dataset. Progressively restrictive experience-based subsets were used to assess whether conclusions remained stable among practitioners with more direct PPP and infrastructure experience; a separate response-pattern subset was used only for data-quality sensitivity testing. The definitions and uses of all analytical sets are summarized in Table 3 and their relationship is illustrated in Figure 3.

Table 3 Analytical samples and their prespecified uses.

Analytical set	n	Definition	Primary use
All valid responses	205	Saudi PPP stakeholder sample passing basic validity checks	Primary analysis
Direct PPP experience	78	Stakeholders reporting direct experience in PPP projects	Practitioner benchmark
PPP + infrastructure	64	PPP-experienced stakeholders with infrastructure experience	Experience-based sensitivity
Highest-confidence subset	39	Stakeholders meeting the most restrictive confidence criteria	Conservative sensitivity
Excluding straightliners	184	Primary sample after excluding 21 uniform-response records	Response-quality sensitivity

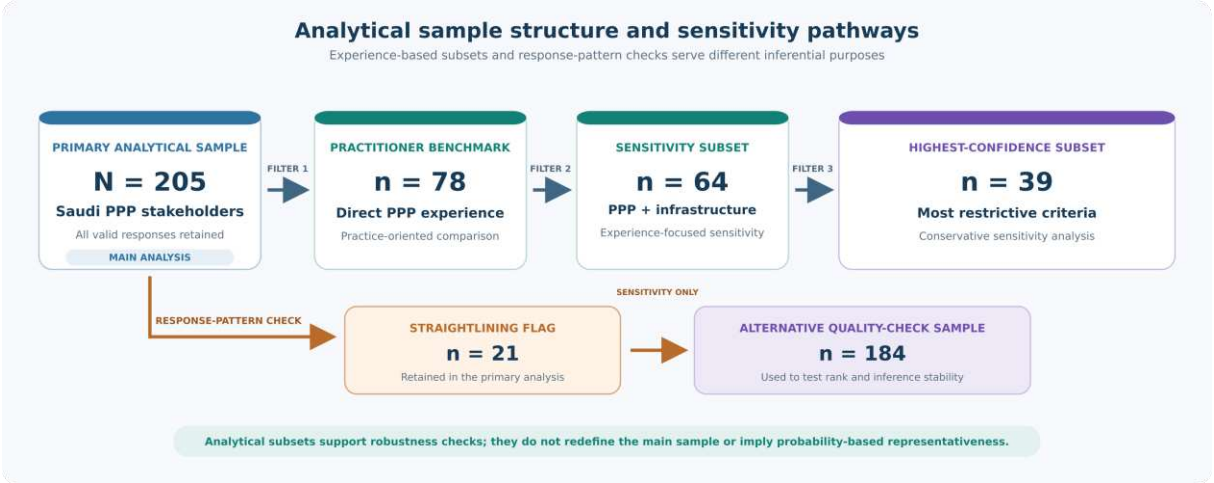


Fig. 3 Analytical samples and sensitivity-analysis pathways. Analytical sample structure and prespecified sensitivity-analysis pathways.

Because recruitment was non-probability based, sample adequacy was assessed only through a descriptive precision benchmark. The conventional proportion-based expression was:

$$n_0 = \frac{z^2_{1-\alpha/2}p(1-p)}{e^2} \quad (6)$$

where z is the standard normal critical value, p is the assumed proportion, and e is the target margin. Using z = 1.96, p = 0.50, and e = 0.07 produced:

$$n_0 = \frac{(1.96)^2(0.50)(1-0.50)}{(0.07)^2} = 196.0 < 205 \quad (7)$$

So the 205 valid records surpassed the descriptive benchmark of 196. This calculation was not intended to serve as a claim of full statistical representativeness, population coverage, or inference from a probability sample.

Missing values, ratings outside of the 1-5 range, duplicate timestamps, identical demographic and rating fields, identical response vectors across participants, and straight lining across the 30 closed items were screened for in the original export of the survey. A repeated timestamp was viewed as a screening flag rather than evidence of duplication and corroborating evidence from the full record was required for exclusion.

For participant j, straight lining was operationalized as a zero-response range across the 30 closed items:

$$\mathrm{SL}_j = \mathbf{1}\{\max_m X_{jm} - \min_m X_{jm} = 0\}, \quad m = 1, \ldots, 30 \quad (8)$$

where Xjm is participant j's response to closed item m. Twenty-one records met this flag. They were retained in the primary N = 205 analysis because a uniform evaluation of all risks may represent a genuine judgment rather than an invalid response. The analysis was therefore repeated after excluding these records (n = 184), and any material change in risk ordering or inference was to be reported as a sensitivity finding rather than used to redefine the main sample. Complete screening outcomes, anonymous profile summaries, and respondent-level cohort flags are provided in Supplementary Data S3.

2.4. Survey Analysis and Uncertainty-Aware Risk Prioritization

The third stage transforms the 205 valid paired probability-impact records into uncertainty-aware risk priorities. All primary calculations are performed at the participant level, to preserve the pairing of probabilities and impacts for each respondent, and the dependence structure across the 15 risks. The analytical sequence consists of ordinal-aware descriptive statistics, separate probability and impact indices, respondent-level criticality scoring, response-pattern consistency diagnostics, repeated-measures inference, participant-level bootstrap resampling, and dominance-based priority-tier construction. The complete workflow is shown in Figure 4, and the prespecified factor-level outputs are summarized in Table 4.

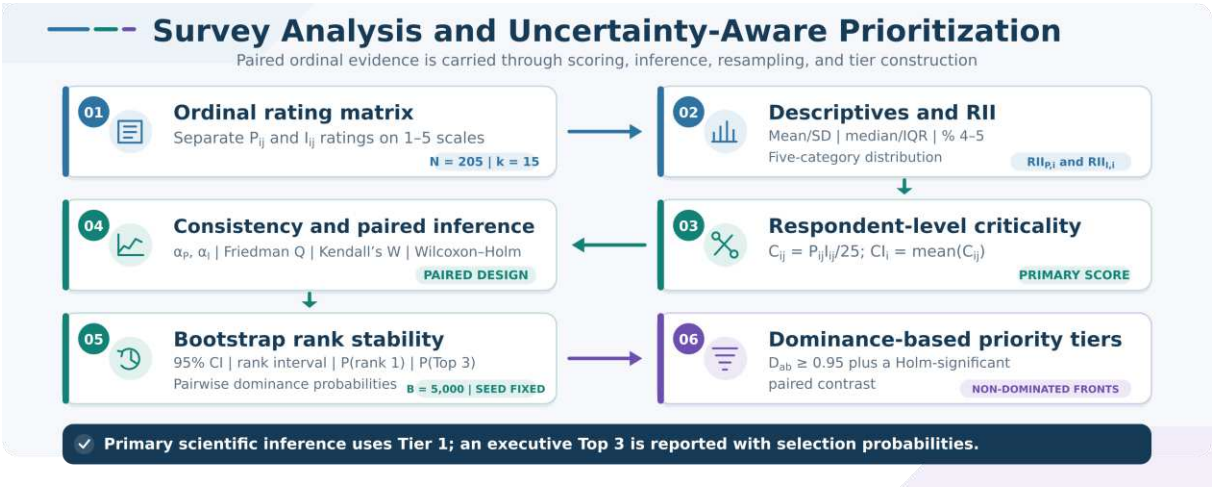


Fig. 4 Survey analysis and uncertainty-aware risk prioritization. Survey-analysis workflow from paired ordinal ratings to descriptive indices, respondent-level criticality, repeated-measures inference, bootstrap rank stability, and dominance-based priority tiers.

Table 4 Prespecified factor-level outputs and analytical roles.

Output family	Reported statistics	Scale or range	Analytical role
Probability and impact descriptives	Mean and SD; median and IQR; percentage rated 4–5; frequencies and percentages for categories 1–5	Reported separately for probability and impact	Provides an ordinal-aware distribution summary while retaining comparability with risk-management studies that report means
Relative importance	$RII_{P,i}$ and $RII_{I,i}$	0.20–1.00	Separates probability-driven from impact-driven importance
Primary criticality	C_{ij}, CI_i, and $RI_i = 25CI_i$	C_{ij} and CI_i: 0.04–1.00; RI_i: 1–25	Preserves within-respondent probability–impact pairing and provides the primary prioritization score
Response-pattern consistency	Separate alpha diagnostics for probability and impact blocks; corrected item–total correlation; alpha if item deleted	Block-level diagnostics	Describes response-pattern consistency without assuming that the 15 risks form one latent construct
Paired inference	Tie-corrected Friedman test; Kendall's W; 105 two-sided asymptotic Wilcoxon signed-rank contrasts (zero_method="wilcox"; zero differences removed; average ranks for tied absolute nonzero differences; correction=False); Holm family-wise adjustment; matched rank-biserial effects.	Omnibus and paired comparisons	Tests whether the same participants assign systematically different criticality to the 15 risks
Bootstrap rank stability	95% score interval; median rank; 95% rank interval; P(Rank 1); P(Top 3); dominance probability	Ranks 1–15; probabilities 0–1	Quantifies uncertainty in factor scores, ordering, and selection
Priority tiers	Iterative non-dominated fronts supported by bootstrap dominance and Holm-adjusted paired evidence	Tier 1, Tier 2, …	Avoids imposing arbitrary numerical cut-offs or an unstable fixed Top-3 rule

For each risk factor, probability and impact ratings are summarized separately by mean and standard deviation, median and interquartile range (IQR), percentage of ratings in high categories 4–5, and the full five-category frequency distribution. The median and IQR are reported as the responses are ordinal; means and standard deviations are also provided for comparison with existing probability-impact studies. Let P ij and I ij be the probability and impact scores assigned by participant j to factor i, where N=205 and k=15. The relative importance index of likelihood is:

$$RII_{P,i} = \frac{\sum_{j=1}^{N} P_{ij}}{5N}, \qquad i = 1, \ldots, k. \tag{9}$$

The corresponding impact index is:

$$RII_{I,i} = \frac{\sum_{j=1}^{N} I_{ij}}{5N}, \qquad i = 1, \ldots, k. \tag{10}$$

Both indices vary between 0.20 and 1.00. They are reported separately, so that a factor with high perceived consequence but lower probability of occurrence is not treated as equivalent to a probability driven risk.

The first criticality calculation is based on the product of probability and impact for each participant before aggregation. The respondent level normalized score is:

$$C_{ij} = \frac{P_{ij} I_{ij}}{25}, \qquad 0.04 \le C_{ij} \le 1.00. \tag{11}$$

The factor-level Criticality Index is the mean of the paired respondent-level products:

$$\mathrm{CI}_i = \frac{1}{N}\sum_{j=1}^{N} C_{ij} = \frac{1}{25N}\sum_{j=1}^{N} P_{ij}\, I_{ij}. \tag{12}$$

For interpretability on the original 1–25 probability–impact scale, the raw Risk Index is:

$$\mathrm{RI}_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij}\, I_{ij} = 25\mathrm{CI}_i. \tag{13}$$

The higher the value, the more critical the stakeholder sees it. The main score is the product at the respondent level, as it maintains the association between probability and impact for each participant. We also compute a product-of-means score as a sensitivity analysis of the scoring estimator:

$$\mathrm{CI}_i^{\mathrm{PM}} = = \frac{\bar{P}_i \bar{I}_i}{25}, \bar{P}_i = \frac{1}{N}\sum_{j=1}^{N} P_{ij}, \bar{I}_i = = \frac{1}{N}\sum_{j=1}^{N} I_{ij} \tag{14}$$

The difference between the two estimators is determined by the within-factor covariance between probability and impact:

$$\overline{PI}_i = \bar{P}_i \bar{I}_i + \mathrm{Cov}_N(P_i, I_i), \mathrm{CI}_i - \mathrm{CI}_i^{\mathrm{PM}} = \frac{\mathrm{Cov}_N(P_i, I_i)}{25} \tag{15}$$

Here, Cov_N uses denominator N, consistent with the identity for the observed participant records. The product-of-means score is therefore treated only as a sensitivity estimator and does not replace CI_i in the primary ranking or tier construction.

Response-pattern consistency is evaluated separately for the 15 probability items and the 15 impact items using Cronbach's alpha:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{\ell=1}^{k} s_\ell^2}{s_T^2}\right). \tag{16}$$

For each block, the analysis reports alpha, the corrected item–total correlation for every factor, and alpha if that item is deleted. These statistics are interpreted only as diagnostics of response-pattern consistency. They are not treated as evidence that the 15 substantively distinct risks form a single reflective latent construct, and no risk is removed solely to increase alpha.

Because every participant evaluates all 15 factors, differences in criticality are tested using a repeated-measures rank procedure. Within each participant, the C_{ij} values are ranked across the $k = 15$ factors using average ranks for ties. Let R_i denote the sum of within-participant ranks for factor i, and let t_{jg} denote the size of tied group g for participant j. The Friedman statistic and its tie correction are:

$$Q_0 = \frac{12}{Nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3N(k+1), C_F = 1 - \frac{\sum_{j=1}^{N}\sum_g \left(t_{jg}^3 - t_{jg}\right)}{N(k^3-k)}, Q = \frac{Q_0}{C_F} \tag{17}$$

The tie-corrected Q is evaluated against a chi-square distribution with $k - 1 = 14$ degrees of freedom. Kendall's coefficient of concordance is reported as the omnibus effect size:

$$W = \frac{Q}{N(k-1)}, \qquad 0 \le W \le 1. \tag{18}$$

A significant Friedman result is followed by all pairwise Wilcoxon signed-rank comparisons between the 15 factors. The prespecified number of contrasts is:

$$\binom{k}{2} = \binom{15}{2} = \frac{15\times14}{2} = 105. \tag{19}$$

For factors a and b, the paired differences were defined as Δj,ab = Cja − Cjb. Zero differences were removed before the signed-rank calculation (zero_method="wilcox"), and tied absolute nonzero differences received average ranks. Two-sided asymptotic p-values were computed in SciPy version 1.17.0 (method="asymptotic"), with continuity correction disabled (correction=False). The resulting 105 p-values were adjusted simultaneously using Holm's step-down procedure. The direction and magnitude of each paired difference were summarized using the matched-pairs rank-biserial coefficient defined in Equation (20).

$$r_{rb,ab} = \frac{W_{ab}^{+} - W_{ab}^{-}}{W_{ab}^{+} + W_{ab}^{-}}. \tag{20}$$

Here, W_{ab}^{+} and W_{ab}^{-} are the positive and negative signed-rank sums after zero differences have been removed. Positive values indicate higher criticality for factor a, negative values indicate higher criticality for factor b, and the absolute value describes the paired effect magnitude.

The uncertainty in the criticality scores and rankings is assessed by resampling the participants as full response vectors. For each of B=5000 bootstrap replicates, sample N=205 rows of participants with replacement. Recalculate all 15 CI_i values at once, and assign rank 1 to the highest score (and average ranks for exact ties). This allows for dependence between probability and impact, between factors and between all items provided by the same respondent. The seed is fixed at 20260809. The percentile score interval and the empirical rank interval are given by:

$$\mathcal{C}_i^{0.95} = \left[Q_{0.025}\left\{\mathrm{CI}_i^{*(b)}\right\}, Q_{0.975}\left\{\mathrm{CI}_i^{*(b)}\right\}\right], \mathcal{R}_i^{0.95} = \left[Q_{0.025}\left\{r_i^{*(b)}\right\}, Q_{0.975}\left\{r_i^{*(b)}\right\}\right] \tag{21}$$

For each factor, the median bootstrap rank, the probability of occupying rank 1, and the probability of appearing within the top three are calculated as:

$$P_i(\text{Rank } 1) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(r_i^{*(b)} = 1\right), P_i(\text{Top } 3) = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(r_i^{*(b)} \le 3\right) \quad (22)$$

Pairwise bootstrap dominance indicates the consistency of factor a dominating factor b, with half weight being given to exact bootstrap ties:

$$A_{ab}^{(b\prime)} = \mathbf{1}\left(\mathrm{CI}_a^{*(b\prime)} > \mathrm{CI}_b^{*(b\prime)}\right), E_{ab}^{(b\prime)} = \mathbf{1}\left(\mathrm{CI}_a^{*(b\prime)} = \mathrm{CI}_b^{*(b\prime)}\right), D_{ab} = \frac{1}{B}\sum_{b\prime=1}^{B}\left[A_{ab}^{(b\prime)} + 1/2\, E_{ab}^{(b\prime)}\right] \quad (23)$$

Factor a is considered to dominate factor b only when three conditions are satisfied simultaneously: $D_{ab} \ge 0.95$; the paired score direction favors factor a; and the Wilcoxon comparison remains significant after Holm correction. These joint conditions prevent a narrow point-estimate difference or an unstable bootstrap ordering from being interpreted as a scientifically meaningful separation. Priority tiers are then generated as iterative non-dominated fronts according to Table 5.

Table 5 Construction and interpretation of dominance-based priority tiers.

Analytical step[1]	Prespecified rule	Interpretation and required reporting
1. Lock the reference analysis	Use full-sample respondent-level CI_i values; rank higher criticality as better priority, with average ranks for exact ties	Retain the baseline score, point rank, 95% score interval, median bootstrap rank, and 95% rank interval for every factor
2. Establish pairwise dominance	Define $a \succ b$ only when $D_{ab} \ge 0.95$, the paired direction favors a, and the Holm-adjusted Wilcoxon p-value is ≤ 0.05	Report the dominance probability, adjusted p-value, and matched rank-biserial effect; otherwise treat the pair as not clearly ordered
3. Construct Tier 1	Assign to Tier 1 every factor not dominated by any remaining factor	Tier 1 is the primary scientific priority set; it may contain more or fewer than three factors
4. Construct subsequent tiers	Remove Tier 1, identify the non-dominated factors among the remainder as Tier 2, and repeat until every factor is assigned	Produces transparent non-dominated fronts without arbitrary criticality cut-offs
5. Preserve incomparable factors	Keep factors on the same front when the evidence does not support a directional dominance relation	Overlapping or uncertain ranks are not forced into an artificial strict ordering
6. Report an executive Top 3	Display the three highest point-estimate factors for operational communication, together with P(Top 3), P(Rank 1), and rank intervals	The executive Top 3 is descriptive and does not replace Tier 1 as the primary scientific result
7. Test stability	Compare the reference tiers and ranks with prespecified alternative samples and the product-of-means estimator in Phase 5	Report factor-level shifts, Tier 1 retention, rank concordance, maximum displacement, and Top-3 similarity

1 All paired Wilcoxon contrasts used the same explicit implementation in Supplementary Data S4, Supplementary Table S5, and Supplementary Data S6: two-sided asymptotic inference, zero_method="wilcox", average ranks for tied absolute nonzero differences, and no continuity correction.

All factor level statistics are reported with unrounded values kept in the analysis files and rounded values presented with appropriate rounding in the manuscript. The primary ranking is derived from respondent-level criticality; RII values, the raw 1–25 risk index, product-of-means scores and the executive Top 3 are supplementary interpretations and not competing primary rankings. All analyzes were conducted using version-controlled scripts in Python 3.12.13, NumPy 2.3.5 and SciPy 1.17.0. Supplementary Data S4, Supplementary Table S5, and Supplementary Data S6 were re-calculated using a single explicit Wilcoxon implementation. S12 confirms identical W statistics, raw p-values, and Holm-adjusted p-values for all 105 pairs.

2.5. Literature-Market Evidence Integration

The fourth phase compared the uniqueness of study literature recurrence with Saudi stake-holder criticality for the same 15 risk factors. It estimated the tied-rank concordance and its resampling uncertainty and then retained the two sources of evidence as separate percentile coordinates in an evidence-alignment matrix. The 15 survey factors were shortlisted from the literature without the use of a literature-market composite as a basis for the primary ranking; a weighted integrated score was limited to sensitivity analy-sis. The whole integration pathway is summarized in Figure 5.

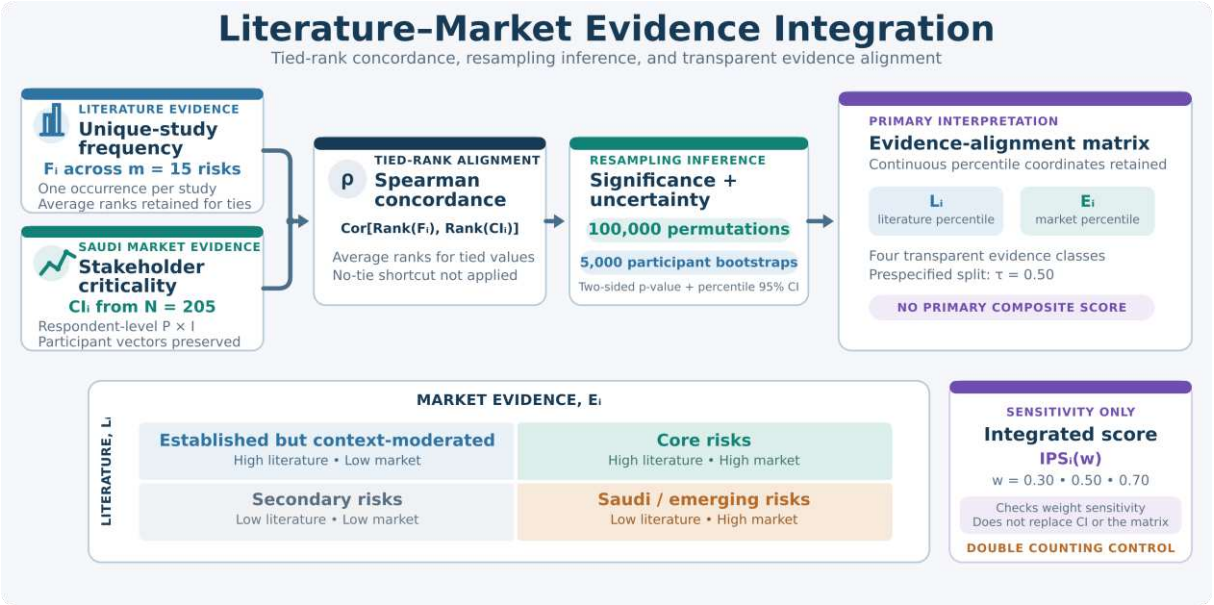


Fig. 5 Literature–market evidence integration and alignment**.** Literature-market integration pathway from tied-rank concordance to evidence-alignment classes and sensitivity-only composite scoring.

For factor i, F(i) denotes the number of recurrence-eligible normalization sources in which the factor occurred, whereas CI(i) is the respondent-level market Criticality Index defined in Phase 3. Because tied frequencies and tied criticality estimates are possible, Spearman's coefficient will be calculated as the Pearson correlation between average-rank vectors:

$$\rho_s = \mathrm{Cor}[\mathrm{Rank}_{\mathrm{avg}}(F_i), \mathrm{Rank}_{\mathrm{avg}}(\mathrm{CI}_i)] \quad \textbf{(24)}$$

Therefore, the coefficient would be obtained directly from tied ranks m = 15 matched factors. We will not use the shortcut based on squared rank differences without ties. Positive values mean that factors more often found in the literature are also more likely to be of higher criticality in the Saudi stakeholder sample. The coefficient is interpreted as evidence of alignment, not causation or evidence of implementation effectiveness.

A two-sided Monte Carlo permutation test will be used to test the statistical evidence of rank alignment. The market rank vector will be randomly re-assigned to the 15 literature ranks M = 100,000 times and Spearman's coefficient will be recomputed after each re-assignment. A plus-one correction prevents zero estimated p-values:

$$p_{\mathrm{perm}} = \frac{1+\sum_{r=1}^{M} 1(|\rho_s^{(r)}| \geq |\rho_{s,\mathrm{obs}}|)}{M+1}, M = 100{,}000 \quad \textbf{(25)}$$

The sampling uncertainty in the market evidence will be independent quantified, by resampling the N = 205 participants as complete response vectors of 30 items . For each of B = 5,000 bootstrap samples, all CI(i) values and the tied-rank correlation will be recomputed with F(i) held constant. The percentile interval is given:

$$\mathrm{CI}_{95\%}(\rho_s) = [Q_{0.025}\{\rho_s^{*(h)}\}, Q_{0.975}\{\rho_s^{*(h)}\}], h = 1, \ldots, B \quad \textbf{(26)}$$

We will report the observed coefficient, permutation p-value, and 95% participant-bootstrap interval together. This bootstrap reflects uncertainty from the sample of stakeholders, not uncertainty from literature retrieval, normalization, or prior selection of the 15 factors. The latter may restrict the range of literature frequencies and will be mentioned when talking about the magnitude of the coefficient.

The literature and market evidence will be transformed into normalized ascending percentile ranks so that we can compare the two sources on the same descriptive scale without losing their distinct meaning. Where rank 1 is the lowest and rank m is the highest. Ties are awarded their average rank. The coordinate in the literature is:

$$L_i = \frac{\mathrm{Rank}^{\uparrow}_{\mathrm{avg}}(F_i)-1}{m-1} \quad \textbf{(27)}$$

The corresponding market coordinate is:

$$E_i = \frac{\mathrm{Rank}^{\uparrow}_{\mathrm{avg}}(\mathrm{CI}_i)-1}{m-1} \quad \textbf{(28)}$$

For m=15 the denominator is 14. L(i) and E(i) keep their relative position in the shortlist and they will be exhibited as continuous coordinates for all the factors. They are not interpreted as probabilities or measures of absolute risk magnitude.

For clear descriptive classification a pre-specified midpoint τ= 0.50 will separate relatively high from relatively low evidence on each axis. The high group will get the values at the center point. The two binary indicators are:

$$H_{L,i} = 1(L_i \geq \tau), H_{E,i} = 1(E_i \geq \tau), \tau = 0.50 \quad \textbf{(29)}$$

The resulting four evidence classes and their intended interpretation are defined in Table 6. Continuous coordinates and the original F(i) and CI(i) values will accompany every quadrant label so that the midpoint split does not conceal near-threshold differences.

Table 6 Operational rules for literature–market evidence alignment.

Evidence class	Indicator rule	Coordinate pattern	Interpretive use
Core risks	$H_{L,i} = 1;\ H_{E,i} = 1$	High literature; high market	Consistently prominent risks supported by both recurrence and Saudi stakeholder criticality
Saudi / emerging risks	$H_{L,i} = 0;\ H_{E,i} = 1$	Low literature; high market	Context-sensitive or emerging risks that warrant specific attention in Saudi PPP practice
Established but context-moderated risks	$H_{L,i} = 1;\ H_{E,i} = 0$	High literature; low market	Established concerns whose perceived Saudi criticality is comparatively moderated
Secondary risks	$H_{L,i} = 0;\ H_{E,i} = 0$	Low literature; low market	Lower relative priority within the 15-factor shortlist, without implying absence of risk

Since the literature recurrence was the source of the factor selection, using it again in the primary ranking could double count the same evidence. Therefore an Integrated Priority Score based on percentiles will be calculated only as a sensitivity analysis:

$$\mathrm{IPS}_i(w) = wE_i + (1-w)L_i, w \in \{0.30, 0.50, 0.70\} \quad \textbf{(30)}$$

Here, w is the market-evidence weight: 0.30 is literature-leaning, 0.50 is balanced, and 0.70 is market-leaning. IPS ranks and shifts relative to the primary CI ranking will be reported for each scenario; IPS will not replace CI, dominance-based tiers, or the evidence-alignment matrix.

The average ranks, the tied-rank Spearman correlation, M = 100,000 permutations, B = 5,000 participant-vector bootstraps, the evidence classification, and the three IPS sensitivity scenarios were implemented in Python 3.12.13 using NumPy 2.3.5 and SciPy 1.17.0 with the fixed seed 20260809. Supplementary Data S7 contains the F(i), CI(i), tied ranks, L(i), E(i), evidence class, the full permutation distribution, the bootstrap correlation distribution and the IPS sensitivity results for each factor.

2.6. Subgroup and Robustness Analyses

The fifth phase will evaluate whether the 15 respondent-level risk-criticality scores differ across specified Saudi PPP stakeholder groups and whether the generated priority order is robust to alternative samples and scoring choices. Group comparisons will be done using rank-based tests with hierarchical multiplicity control. Robustness will be assessed by repeating the full 15-factor ranking under six prespecified scenarios. These are explanatory and diagnostic analyzes; they set the stage but do not substitute for the primary full-sample prioritization and dominance-based tiers established in Phase 3. The analytical path is summarized in Figure 6.

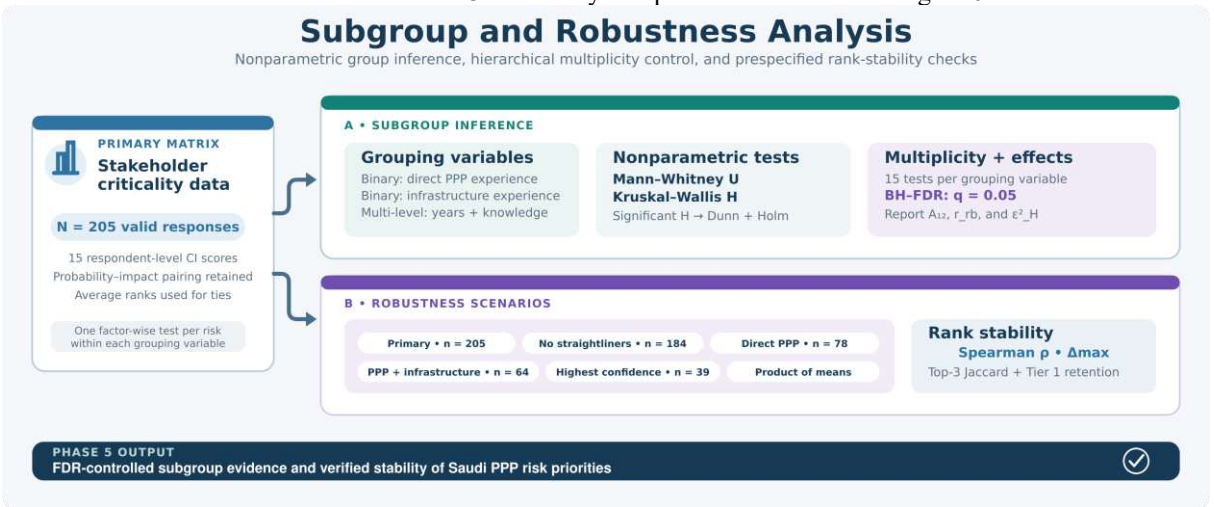


Fig. 6 Subgroup inference and robustness-analysis workflow. Subgroup and robustness-analysis pathway linking nonparametric group inference, hierarchical multiplicity control, prespecified alternative samples and estimators, and rank-stability outputs.

The main outcome for factor i and participant j is the respondent-level criticality score C(ij) = P(ij)I(ij)/25. Therefore, for each grouping variable there will be 15 tests, one per risk, at the factor level. The binary contrasts will be: Direct PPP experience vs. no PPP experience and Infrastructure experience vs. no infrastructure experience. Years of experience and self-reported PPP knowledge will be analyzed within the original pre-specified categories. All 205 valid responses will be included. Each comparison will be based on available cases for the grouping variable. Denominators will be reported for each group. No imputation of missing group membership will be performed. Probability and impact scores could be examined separately as secondary diagnostics but conclusions on subgroups and control of multimultiplicity will be based on C(ij). No collapsing of outcomes will occur after inspection of outcomes; any pre-analysis pooling required by sparse cells will be documented with its substantive justification.

For each factor, a two-sided Mann-Whitney U test will be performed to compare two independent groups. Scores will be combined with their average ranks for ties. If the group sizes are n(1) and n(2) and the rank sum for group 1 is R(1) then the oriented statistic is:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2} \quad \textbf{(31)}$$

Here U(1) counts favorable group-1 comparisons, plus one-half of exact ties. Effect magnitude will be reported as the probability of superiority A(12) and its rank-biserial transformation:

$$A_{12} = \frac{U_1}{n_1 n_2}, r_{\mathrm{rb}} = 2A_{12} - 1 = \frac{2U_1}{n_1 n_2} - 1 \quad \textbf{(32)}$$

Positive r(rb) indicates higher scores in the group named first; negative values indicate the reverse. Two-sided asymptotic p-values with tie correction will be used because the discrete probability–impact product produces many ties. Group medians, interquartile ranges, U, A(12), r(rb), raw p-values, and multiplicity-adjusted p-values will be reported.

Variables with G greater than two categories will be analyzed using the Kruskal–Wallis's test. With N available observations, n(g) observations and rank sum R(g) in category g, the uncorrected statistic is:

$$H = \frac{12}{N(N+1)} \sum_{g=1}^{G} \frac{{R_g}^2}{n_g} - 3(N+1) \quad \textbf{(33)}$$

If t(l) observations share the l-th tied value among L tie groups, the tie factor C and corrected statistic H(c) are:

$$\mathrm{C} = 1 - \frac{\sum_{\mathrm{l}=1}^{\mathrm{L}} {\mathrm{t_l}}^3 - \mathrm{t_l}}{\mathrm{N}^3 - \mathrm{N}}, \mathrm{H_c} = \frac{\mathrm{H}}{\mathrm{C}} \quad \textbf{(34)}$$

The omnibus effect size will be the non-negative epsilon-squared estimate:

$$\varepsilon_H^2 = \max\{0, \frac{H_c - G + 1}{N - G}\} \quad \textbf{(35)}$$

Only a factor whose omnibus test remains significant after the across-factor FDR procedure will proceed to Dunn pairwise comparisons. For categories g and h, the tie-corrected Dunn statistic is:

$$z_{gh} = \frac{\overline{R}_g - \overline{R}_h}{[\frac{N(N+1)C}{12}(\frac{1}{n_g} + \frac{1}{n_h})]^{\frac{1}{2}}} \quad \textbf{(36)}$$

The 15 binary or omnibus p-values for each grouping variable will form one inferential family and will be corrected for false-discovery rate using the Benjamini-Hochberg procedure at q = 0.05. After sorting the m p-values in ascending order, the monotone adjusted value is:

$$p_{(i)}^{\mathrm{BH}} = \min_{j \geq i}\{\min[1, \frac{m}{j} p_{(j)}]\}, m = 15, q = 0.05 \quad \textbf{(37)}$$

A factor will be treated as subgroup-associated only when its BH-adjusted p-value is at most 0.05. For a BH-significant multi-category factor, Holm adjustment will then be applied across all G(G−1)/2 Dunn contrasts for that factor. This hierarchy controls discovery across risks while limiting post-hoc testing to supported omnibus effects. The complete testing and reporting sequence is specified in Table 7.

Table 7 Subgroup testing, multiplicity control, and effect-size reporting.

Analytical stage	Primary procedure	Multiplicity rule	Minimum reported evidence
Binary factor-wise tests	Two-sided Mann–Whitney U; average ranks and tie correction	BH across 15 risks for each grouping variable	Group n, median (IQR), U, A_{12}, r(rb), raw and BH-adjusted p
Multi-category omnibus tests	Tie-corrected Kruskal–Wallis H(c)	BH across 15 risks for each grouping variable	Category n, median (IQR), H(c), df, ε^2(H), raw and BH-adjusted p
Supported post-hoc contrasts	Dunn comparisons only after a BH-significant omnibus test	Holm across all category pairs within that risk	Mean-rank difference, z, direction, and Holm-adjusted p
Interpretation	Effect sizes and uncertainty considered with adjusted evidence	No selection by unadjusted p-values	Subgroup findings contextualize; they do not replace the primary ranking

The 15 risks will be reprioritized across the six scenarios in Table 8. The reference analysis is the full sample and the product at the level of the respondent. We will not remove the 21 straightliners from that analysis as identical ratings can be a genuine judgment and their removal is limited to sensitivity analysis. Subset definitions will be frozen prior to rank comparison. All scenarios will have the same factor coding and the direction of higher scores indicating greater criticality.

Table 8 Prespecified robustness scenarios and analytical purposes.

Scenario	n	Definition	Analytical purpose
Primary reference	205	All valid responses; respondent-level P × I / 25	Reference ranking and dominance-based tiers
No straightliners	184	Exclude 21 invariant 30-item response patterns	Response-pattern sensitivity without presuming invalidity

Scenario	n	Definition	Analytical purpose
Direct PPP experience	78	Direct PPP experience	Practitioner benchmark
PPP + infrastructure experience	64	Both direct PPP and infrastructure experience	Domain-specific sensitivity
Highest-confidence subset	39	Strictest prespecified experience–knowledge profile	Conservative stakeholder benchmark
Product-of-means estimator	205	Mean probability × mean impact / 25	Scoring-estimator sensitivity

For scenario a, S(i,a) denotes its factor score: the within-subset respondent-level CI for sample scenarios, or the product-of-means score for the alternative estimator. Concordance with the baseline ranking will be measured using tied-rank Spearman correlation:

$$\rho_s{}^{(a)} = \mathrm{Cor}[\mathrm{Rank}_{\mathrm{avg}}(\mathrm{CI}_i{}^{\mathrm{base}}), \mathrm{Rank}_{\mathrm{avg}}(S_i{}^{(a)})] \quad \textbf{(38)}$$

The largest absolute displacement of any factor will be:

$$\Delta_{\max}{}^{(a)} = \max_{i=1,\ldots,15} |r_i{}^{(a)} - r_i^{\mathrm{base}}| \quad \textbf{(39)}$$

Top-set similarity will use the Jaccard coefficient; T(k) contains all factors at or above the k-th ordered score, including boundary ties:

$$J_k{}^{(a)} = \frac{|T_k{}^{\mathrm{base}} \cap T_k{}^{(a)}|}{|T_k{}^{\mathrm{base}} \cup T_k{}^{(a)}|}, k = 3 \quad \textbf{(40)}$$

We will report Spearman's rho, maximum rank shift, Top-3 Jaccard sim-ilarity, factor-level shifts and baseline Tier 1 retention for each scenario. Metrics will be interpreted together without any arbitrary cutoff. When ranks are close, tier 1 retention is primary.

We implemented subgroup tests, effect sizes, Dunn statistics defined in Equation (36), multiplicity adjustments, and all six robustness scenarios in Python 3.12.13 with NumPy 2.3.5 and SciPy 1.17.0. Supplementary Data S8 preserves group sizes, all 60 risk-wise tests, BH-FDR results, effect estimates, and eligible Dunn-Holm comparisons; Supplementary Data S9 preserves scenario scores, ranks, stability metrics, and Tier 1 membership.

2.7. Embedded Qualitative Analysis of Open-Ended Responses

In Phase 6 the open-ended item of the questionnaire will be used as an embedded explanatory qualitative strand. A hybrid qualitative content-analysis design will be used. First responses will be inductively open-coded and thematically developed. Then the stabilized themes will be mapped deductively to the 15 risk factors surveyed. This sequence allows for genuinely new Saudi PPP concerns to emerge without being forced into the existing list. Qualitative findings will be used to explain and enrich the quantitative results but will not affect the criticality scores, ranks, or priority tiers. The auditable workflow is presented in Figure 7.

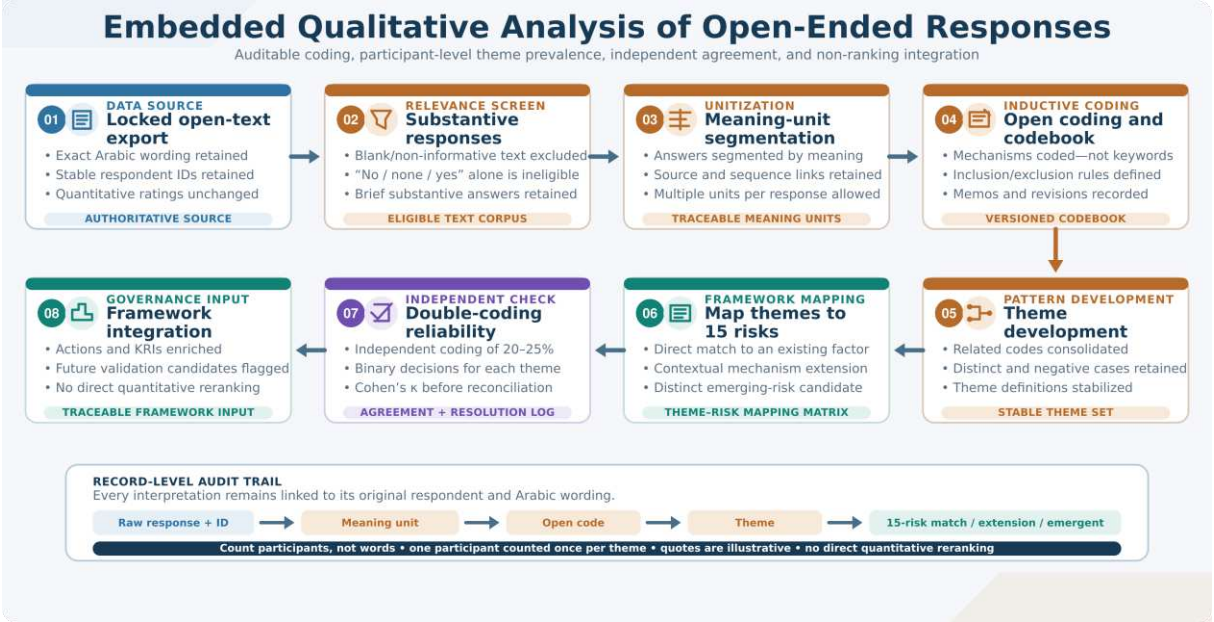


Fig. 7 Embedded qualitative-analysis workflow. Auditable embedded qualitative-analysis workflow linking relevance screening, meaning-unit segmentation, inductive coding, theme development, mapping to the 15 survey risks, participant-level prevalence, and non-ranking framework integration.

The canonical source is the locked survey export. We will provide a stable anonymous respondent identifier and the verbatim text of all submitted open-text responses. Blank entries and answers consisting only of non-informative expressions (e.g. (no), (none), or (yes)) will be excluded if they do not contain any interpretable risk, cause,

consequence, control or contextual observation. Short, meaningful answers will be retained. Decisions on screening and reasons for exclusion will be documented. Any quotation reported will be de-identified of any personal or organizational identifiers.

If a response contains more than one distinct substantive proposition, it will be divided into meaning units. Each unit will be preserved with the source respondent, sequence in the response, and link to the original text. This record-level crosswalk will preserve context and enable individual risk mechanisms within a compound response to be coded independently.

The underlying risk mechanism will be coded inductively as opposed to repeated words. The initial codes will remain close to the participants' meanings and will be captured in a version-controlled codebook. The codebook will include the code name, operational definition, criteria for inclusion and exclusion, example text, analytical memo, and revision history. When codes describe the same mechanism, related codes will be grouped into themes. If the governance implications of causes, risk events and consequences differ, they will remain distinct. Rather than suppressing them, negative or deviant cases will be retained. Table 9 defines the full decision protocol and the required audit outputs.

Table 9 Screening, coding, integration, and audit rules for qualitative analysis.

Analytical step	Unit or input	Prespecified decision rule	Required audit output
Source control	Locked open-text export	Retain exact wording and a stable anonymous respondent identifier; do not modify the quantitative ratings	Immutable source file and de-identified respondent key
Relevance screening	Submitted response	Exclude blank or non-informative text only when no interpretable content is present; retain brief substantive text	Eligibility flag and documented exclusion reason
Meaning-unit segmentation	Substantive response	Split only at a change in proposition or risk mechanism; preserve source, sequence, and original text	Response-to-unit crosswalk
Open coding	Meaning unit	Code the mechanism conveyed, not isolated keywords; permit multiple codes when substantively necessary	Coded-unit table, memos, and versioned codebook
Theme development	Related open codes	Merge codes representing the same mechanism; retain distinct, causal, consequential, and negative cases where warranted	Theme definitions and code-to-theme crosswalk
Risk mapping	Stabilized theme	Classify as direct match, contextual extension, or emergent candidate after—not before—inductive theme development	Theme-to-risk matrix with rationale
Independent check	Random 20–25% of substantive responses	Second researcher independently segments and codes before reconciliation; calculate theme-wise binary agreement	Sampling record, coder matrices, kappa results, and resolution log
Participant-level counting	Respondent × theme presence	Count each participant once per theme, irrespective of repeated words, units, or mentions	Theme-level n and percentage with denominator
Integration	Mapped and emergent themes	Use to explain findings and enrich controls or KRIs; do not directly change quantitative ranks or tiers	Traceable framework input or future-validation flag

The mapping will only start once the theme definitions have stabilized. A direct match will indicate the same mechanism as one of the 15 factors. A contextual extension will add a Saudi-specific condition, actor, stage, or consequence, without being a separate risk mechanism. An emergent candidate will indicate a substantively distinct mechanism that could not be subsumed under any factor's operational definition. Am-biguous classifications will be resolved through documented adjudication and the selected class and reason for selection will be retained in the mapping matrix.

The completed qualitative data set did not employ independent second-coder review. Therefore, no report or claim of any double-coded proportion, Cohen's kappa coefficient or reconciliation matrix is made.

Thus, the final analysis is reported as a single-coder process. An anonymous response-to-unit crosswalk, explicit screening decisions, versioned code definitions, code-to-theme links, and a theme-to-risk mapping matrix enable reproducibility.

This limitation is reported transparently in the Results and Supplementary Data S10; no independent-coding or agreement output is inferred from the retained files.

Theme prevalence will be calculated at participant level rather than from word, code, or meaning-unit frequencies. Let A(jt) equal 1 when substantive respondent j mentions theme t in at least one meaning unit and 0 otherwise; let N(q) denote the number of respondents with substantive open-text content. The participant count and percentage are:

$$n_t = \sum_{j=1}^{N_q} A_{jt}\,, \pi_t = \frac{100 n_t}{N_q} \quad (41)$$

Thus, a participant may contribute to several themes but only once to any given theme. Repeated wording or multiple units from the same participant will be descriptive only and will not inflate prevalence. Reporting will include the numbers of submitted responses, non-substantive exclusions, substantive respondents, meaning units, codes, and final themes, together with n(t), π(t), and the mapping class for every theme.

The mapped themes will be used to interpret the quantitative results and to refine descriptions of risks, prevention and mitigation actions, contingency measures and key risk indicators in the proposed framework. Emergent candidates will be retained as explicitly labeled qualitative findings to be empirically investigated in the future and will not be directly slotted into the CI ranking or the dominance-based priority tiers. Where quotations are used they will be short and illustrative rather than treated as prevalence evidence; the original text will be retained and any English rendering will be translation-checked.

Supplementary Data S10 includes the screening log, de-identified substantive re-sponses, response-to-unit crosswalk, working codebook, code-to-theme crosswalk, theme-to-risk mapping matrix, participant-level prevalence, and the explicit agree- ment-status record. Counts and crosswalk integrity checks were implemented in Python 3.12.13 with version-controlled scripts. No separate manual coding software was used or claimed.

2.8. Proposed Evidence-Based Saudi PPP Risk Governance Framework

Phase 7 uses the evidence from Phases 1-5, ranked and uncertainty-qualified, and also the coded qualitative findings from Phase 6, to develop an operational Saudi PPP risk-governance structure. All 15 risks are still in the master register. Tier 1 defines increased management attention and resource priority rather than inclusion or exclusion. The resultant artifact is named the Proposed Evidence-Based Saudi PPP Risk Governance Framework because its governance fields are study derived, pro-ject-adaptable specifications. The general architecture is shown in Fig.8.

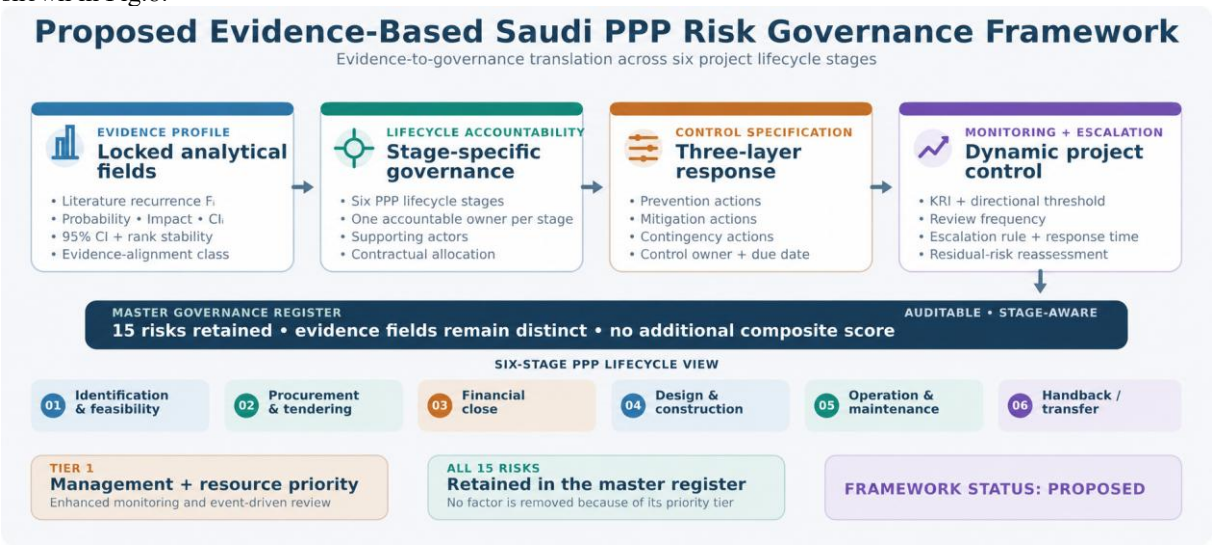


Fig. 8 Translation of risk evidence into lifecycle governance. Translation of ranked evidence into the proposed Saudi PPP risk-governance register, lifecycle responsibilities, controls, monitoring triggers, and tier-based management priorities.

The framework is reported to be PROPOSED. Literature recurrence Probability Impact CI Uncertainty Rank stability Tier Evidence-alignment class The locked analytical outputs are imported without reweighting. Include explicit governance specifications that can be tailored to the project for lifecycle roles, contractual allocation, controls, KRIs, thresholds, review cycles and escalation routes.

A new primary composite score will not be developed. This avoids double counting and preserves the distinct meanings of evidence prevalence, stakeholder criticality, uncertainty and contextual alignment. The register will therefore aid decision making by means of transparent fields and rules, not by an additional opaque index.

The data model will include a 15 row master risk table and will be linked to stage specific governance records. The stable evidence profile of each factor will be held in the master table, whereas changes in ownership, allocation, actions, indicators and residual risk through the lifecycle will be facilitated by linked records without duplicating or overwriting the analytical evidence. The whole field architecture is defined by Table 10.

Table 10 Data architecture and mandatory fields of the proposed risk-governance register.

Register module	Mandatory fields	Construction rule	Governance output
Risk identity and evidence	Risk code and name; literature frequency; probability; impact; CI; 95% interval; rank stability; Tier; evidence-alignment class	Import unchanged from the locked outputs of earlier phases	Auditable evidence profile for each of the 15 risks

Register module	Mandatory fields	Construction rule	Governance output
Lifecycle mapping	Primary stage; all active stages; stage-entry and stage-exit status	One risk may remain active in multiple stages; no forced single-stage assignment	Six-stage lifecycle view and handoff record
Accountability and allocation	Accountable risk owner; supporting actors; contractual allocation; allocation rationale	Exactly one owner for every active risk-stage; support may be multiple; operational ownership and contractual bearing remain separate	Responsibility and allocation map
Control specification	Prevention, mitigation, and contingency actions; action owner; due date; evidence; status	Each action must be specific, attributable, time-bounded, and auditable	Stage-specific control schedule
Monitoring and escalation	KRI; formula or unit; source; adverse direction; threshold; frequency; escalation recipient; response time	Threshold and response are recorded before routine monitoring begins	Early-warning and escalation protocol
Residual risk and audit history	Residual probability; residual impact; residual-risk score; last review; decision; change reason; approver	Reassess after controls, at stage gates, and after a material trigger event	Current risk position and traceable change log

Let R(s) denote the set of risks requiring active governance at lifecycle stage s. For risk i, the activity indicator is:

$$L_{is} = 1(i \in R_s), s = 1, \dots, 6 \quad \textbf{(42)}$$

A factor may have L(is) = 1 in more than one stage. At every stage transition, the outgoing and incoming governance teams will reconfirm the active risks, accountable owner, contractual allocation, open controls, current KRI status, and residual risk. The minimum stage-specific update is defined in Table 11.

Table 11 Lifecycle-specific governance focus and update points.

Lifecycle stage	Primary governance focus	Minimum register update	Required handoff
1. Project identification and feasibility	Strategic fit, affordability, feasibility assumptions, approvals, and early risk ownership	Initial evidence profile; active-stage map; preliminary owner, allocation, controls, and KRIs	Approved feasibility risk brief to procurement team
2. Procurement and tendering	Tender clarity, bankability, allocation logic, bidder interfaces, and evaluation conditions	Tender allocation; supporting actors; bid-stage controls; data sources and trigger definitions	Contract risk baseline at award
3. Financial close	Financing conditions, guarantees, covenants, dependencies, and conditions precedent	Final allocation; financing-related KRIs; unresolved conditions; residual-risk acceptance	Financial-close baseline to delivery team
4. Design and construction	Design interfaces, approvals, cost, schedule, safety, quality, and commissioning readiness	Live KRI values; action status; trigger events; residual risk; escalation decisions	Commissioning and operational-readiness record
5. Operation and maintenance	Service performance, availability, lifecycle maintenance, demand, revenue, and compliance	Operational KRIs; review frequency; contingency status; residual risk and trend	Periodic governance review and handback plan
6. Handback or transfer	Asset condition, contractual compliance, records, knowledge transfer, liabilities, and closure	Condition indicators; outstanding controls; residual liabilities; acceptance decision	Accepted transfer, remediation plan, or formal closeout

Operational accountability will be stage-specific. If O(ias) equals 1 when actor a is accountable for risk i at stage s, the ownership constraint is:

$$\sum_{a=1}^{A} O_{ias} = L_{is}, O_{ias} \in \{0,1\} \quad \textbf{(43)}$$

So each active risk-stage has one responsible owner, but can have multiple supporting actors. The accountable owner is responsible for monitoring and action closure; this role is not necessarily the same as the party who bears the risk under the contract. Contractual allocation will therefore be separately recorded as public, private or shared with a rationale based on control capacity, information access, consequence management and enforceability.

The proposed register will list for each of the active risk-stages prevention actions to reduce likelihood, mitigation actions to reduce consequence or duration and contingency actions triggered after occurrence. Each action will have an owner, due date, evidence of completion and status. Each key risk indicator (KRI) will also have its definition, unit or formula, data source, adverse direction, threshold, review frequency, escalation recipient and response time required.

To accommodate indicators for which either rising or falling values are adverse, let d(i) = +1 when higher values indicate deterioration and d(i) = −1 when lower values indicate deterioration. The trigger rule is:

$$T_{it} = 1[d_i(\mathrm{KRI}_{it} - \tau_i) \geq 0], d_i \in \{-1, +1\} \quad \textbf{(44)}$$

Here, τ_i is a factor-specific project threshold. Categorical indicators will use an equivalent prespecified breach condition. Threshold values and review frequencies represent proposed operational starting points and should be calibrated to project-specific contractual requirements, data availability, and risk appetite before implementation; they will not be fabricated from the survey scale.

Residual risk will be reassessed from project-specific post-control ratings rather than inferred from the cross-sectional survey. At review time t:

$$\mathrm{RR}_{it} = \frac{P_{it}^{\mathrm{res}} I_{it}^{\mathrm{res}}}{25} \quad \textbf{(45)}$$

Escalation will occur when the KRI trigger is breached, residual risk exceeds its approved tolerance, or a critical control action is overdue:

$$E_{it} = 1[T_{it} = 1 \vee \mathrm{RR}_{it} > \lambda_i \vee D_{it}^{\mathrm{overdue}} = 1] \quad \textbf{(46)}$$

In Equation (46), the overdue-control indicator equals 1 when a critical action is late. All 15 factors remain monitored; Tier 1 receives enhanced attention, priority resources, and more frequent or event-driven review. Other tiers retain standard monitoring but are escalated whenever the condition in Equation (46) is satisfied. Any change requires a dated reason, source, approver, and version entry.

The version-controlled register will use fixed risk codes and a linked data dictionary. No new scoring model is introduced; imported analytical values and proposed governance fields will remain traceable to their source, rule, and development status.

2.9. Use of Artificial Intelligence

During the preparation of this manuscript and its Supplementary Materials, the authors used ChatGPT Solo 5.6 to assist with improving the quality and readability of scientific figures, developing and refining analysis code, drafting the Introduction, and supporting the organization and presentation of supplementary tables, manifests, and reproducibility materials. The tool was not used to formulate the research idea, define the scientific objectives, design the conceptual framework, generate data, or interpret the final results. The authors independently reviewed, verified, and edited all AI-assisted content and take full responsibility for the accuracy, originality, and integrity of the published work.

3. Results

3.1. Literature Evidence and Questionnaire Risk Shortlist

The 519-entry audit crosswalk was re-adjudicated at the record level. The final archive contained 495 factor-level expressions (95.4%), 14 contextual indicators (2.7%) mapped for traceability but excluded from recurrence, and 10 non-factor-level statements (1.9%). The previous 134-record Other/paper-specific bucket was eliminated: 110 of these were reassigned to existing factors, 14 were retained as contextual indicators, and 10 were not counted as factors. The factor-level evidence was still consolidated in the same 41-factor register, with raw-expression counts from 1 to 35 per factor. The ranking considered one occurrence per recurrence-eligible source, not expression multiplicity.

F_i ranged from 1 to 11 and LP_i from 7.1% to 78.6% among the 14 recurrence-eligible sources. Cost escalation and overrun (R01), Coordination and communication failure (R03), and Capability and management-system weakness (R04) together had the highest recurrence (F_i = 11; LP_i = 78.6%; average tied rank = 2.0). Four factors had F_i = 10 and five had F_i = 9. These 12 factors were found outside the questionnaire boundary.

For three available positions at the corrected boundary, R09, R10, R14 and N17 were tied at F_i = 8 (LP_i = 57.1%; average rank = 14.5). Thus, R09, R10, and R14 were retained using the documented literature-only relevance, non-overlap, and domain-coverage criteria, and N17 was retained in the full 41-factor register. So the locked questionnaire stayed the same. The corrected shortlist values and the boundary decision are shown in Table 12. The recurrence for all 41 factors is shown in Figure 9. The complete record level audit is provided in Supplementary Data S1.

Table 12 Study-level recurrence and questionnaire risk-factor selection.

Code	Risk factor	Domain	F_i	LP_i (%)	Rank	Boundary rationale
R01	Cost escalation and overrun	Financial	11	78.6	2.0	—
R03	Coordination and communication failure	Managerial	11	78.6	2.0	—
R04	Capability and management-system weakness	Managerial/Governance	11	78.6	2.0	—
R02	Contractual and risk-allocation issues	Contractual	10	71.4	5.5	—

Code	Risk factor	Domain	F_i	LP_i (%)	Rank	Boundary rationale
R05	Resource, equipment, and maintenance constraints	Technical/Operational	10	71.4	5.5	—
R06	Land, site, and environmental constraints	External/Technical	10	71.4	5.5	—
R13	Regulatory and legal uncertainty	Institutional/Regulatory	10	71.4	5.5	—
R07	Planning, scope, and schedule deficiencies	Managerial	9	64.3	10.0	—
R08	Political and security instability	Institutional/External	9	64.3	10.0	—
R11	Design and technical deficiencies	Technical	9	64.3	10.0	—
R12	Government intervention and commitment risk	Institutional/Governance	9	64.3	10.0	—
R15	Corruption and unethical practices	Institutional/Governance	9	64.3	10.0	—
R09	Demand and market risk	Financial/Market	8	57.1	14.5	Distinct market-demand mechanism retained at the four-way boundary.
R10	Approval and permitting delays	Institutional/Regulatory	8	57.1	14.5	Distinct institutional approval/permitting mechanism retained at the boundary.
R14	Tax and tariff changes	Financial/Regulatory	8	57.1	14.5	Distinct fiscal/tariff mechanism retained at the boundary.

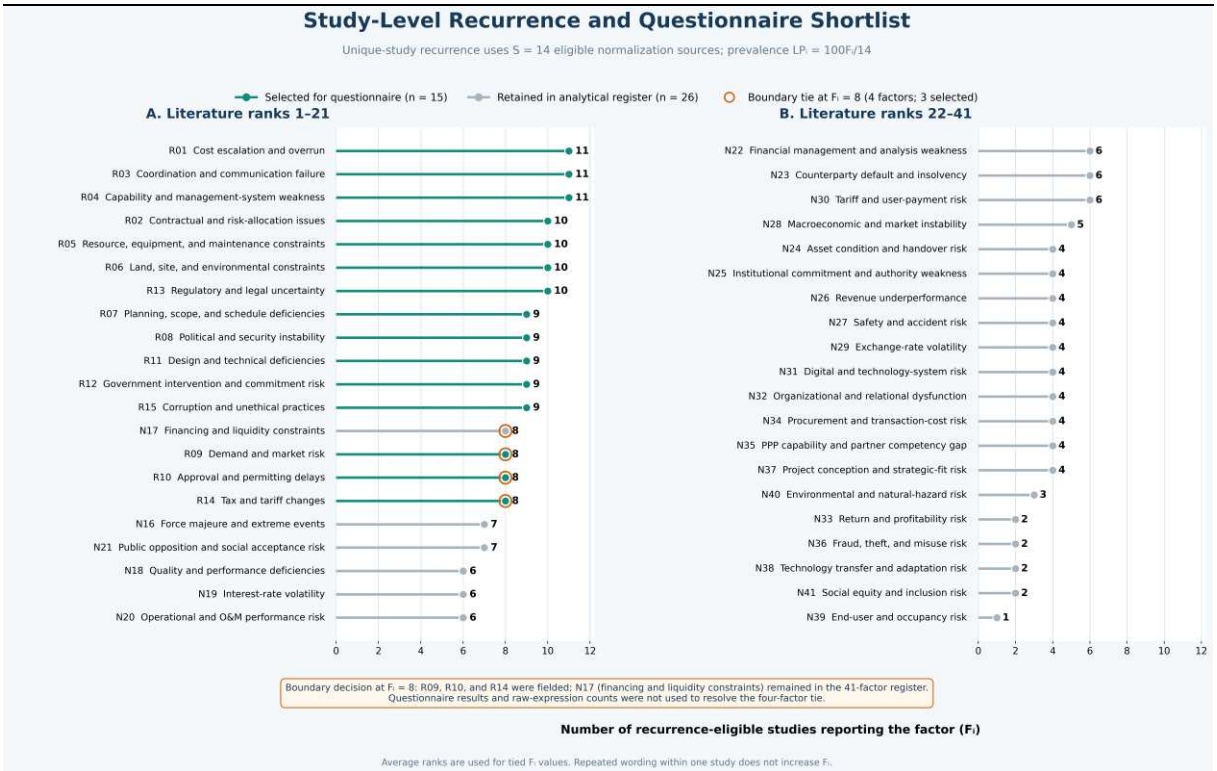


Fig. 9 Study-level recurrence and the questionnaire risk shortlist. Corrected study-level recurrence of the 41 normalized PPP risk factors and identification of the questionnaire shortlist. Recurrence uses one occurrence per factor in each of S = 14 eligible normalization sources. The four orange-outlined factors share $F_i = 8$; R09, R10, and R14 were fielded, while N17 remained in the analytical register. Raw-expression counts and questionnaire results were not used to resolve the tie.

3.2. Response Screening and Stakeholder Profile

The locked export included 205 valid stakeholder records, and a completed 6,150-cell probability-impact rating matrix (205 respondents × 30 closed items). Screening showed that no required profile and rating data were missing,

no values were outside the 1–5 scale, no repeated timestamps were present, and no exact demographic–rating duplicates existed. There were 24 records involved in six recurring 30-item response patterns, but none was confirmed to be a duplicate record. Twenty-one (10.2%) records were straight lined according to the pre-specified zero-range rule. Their exclusion defined the response-quality sensitivity sample (n = 184; 89.8%), whereas they stayed in the primary analysis.

The sample was diverse and not a single expert panel. There were 99 responses (48.3%) from government entities and 57 (27.8%) from private-sector companies. Most common were project management (69; 33.7%). Of the respondents, 104 (50.7%) had 10 or more years of professional experience. Direct PPP participation was reported by 78 respondents (38.0%) and 64 (31.2%) reported both direct PPP and PPP-infrastructure experience. The highest-confidence cohort was 39 (19.0%) respondents who met the latter two experience criteria and who reported good or excellent PPP knowledge. Table 13 provides the full screening and profile distributions and Figure 10 shows the sample and prespecified cohorts as parallel overlapping analytical sets, rather than a sequential exclusion funnel. Complete screening, profile and cohort-flag records are available in Supplementary Data S3.

Table 13 Data-screening outcomes, analytical samples, and stakeholder characteristics.

Characteristic / screening item	Category / criterion	n	% of N	Result / analytical use
Panel A. Data quality and analytical samples				
Locked analytical sample	All valid records	205	100.0	Primary analysis
Missing required data	Participant-profile and 30 closed-rating fields	0	0.0	No imputation required
Out-of-range ratings	Any closed rating outside 1–5	0	0.0	No invalid rating detected
Duplicate timestamps	Repeated timestamp value	0	0.0	No timestamp flag
Exact record duplication	Identical demographic and 30-item rating fields	0	0.0	No corroborated duplicate
Repeated rating vectors	24 records in six recurring 30-item patterns	24	11.7	Pattern flag only; records retained
Straight liners	Zero response range across all 30 closed items	21	10.2	Retained in primary analysis
No-straight liner sample	Excludes the 21 straight liners	184	89.8	Response-quality sensitivity
Direct PPP experience	Reported direct PPP participation	78	38.0	Practitioner benchmark
PPP + infrastructure	Direct PPP and PPP-infrastructure experience	64	31.2	Domain-specific sensitivity
Highest-confidence cohort	PPP + infrastructure and good/excellent PPP knowledge	39	19.0	Conservative sensitivity
Panel B. Respondent characteristics				
Organization type	Government entity	99	48.3	
Organization type	Private-sector company	57	27.8	
Organization type	Consultancy office	20	9.8	
Organization type	Contracting company	13	6.3	
Organization type	Academic/research entity	7	3.4	
Organization type	Other/free-text	9	4.4	
Main field of expertise	Project management	69	33.7	
Main field of expertise	Operations and maintenance	28	13.7	
Main field of expertise	Civil engineering/roads	28	13.7	
Main field of expertise	Risk management	20	9.8	
Main field of expertise	Finance and investment	10	4.9	
Main field of expertise	Contracts and procurement	9	4.4	
Main field of expertise	Other/free-text	41	20.0	
Years of experience	<5 years	41	20.0	
Years of experience	5–<10 years	60	29.3	

Characteristic / screening item	Category / criterion	n	% of N	Result / analytical use
Years of experience	10–<15 years	44	21.5	
Years of experience	15–<20 years	30	14.6	
Years of experience	≥20 years	30	14.6	
Direct PPP exposure	Yes	78	38.0	
Direct PPP exposure	General knowledge only	53	25.9	
Direct PPP exposure	No	74	36.1	
PPP-infrastructure experience	Yes	87	42.4	
PPP-infrastructure experience	No	118	57.6	
Self-reported PPP knowledge	Very low	35	17.1	
Self-reported PPP knowledge	Low	24	11.7	
Self-reported PPP knowledge	Moderate	73	35.6	
Self-reported PPP knowledge	Good	44	21.5	
Self-reported PPP knowledge	Excellent	29	14.1	

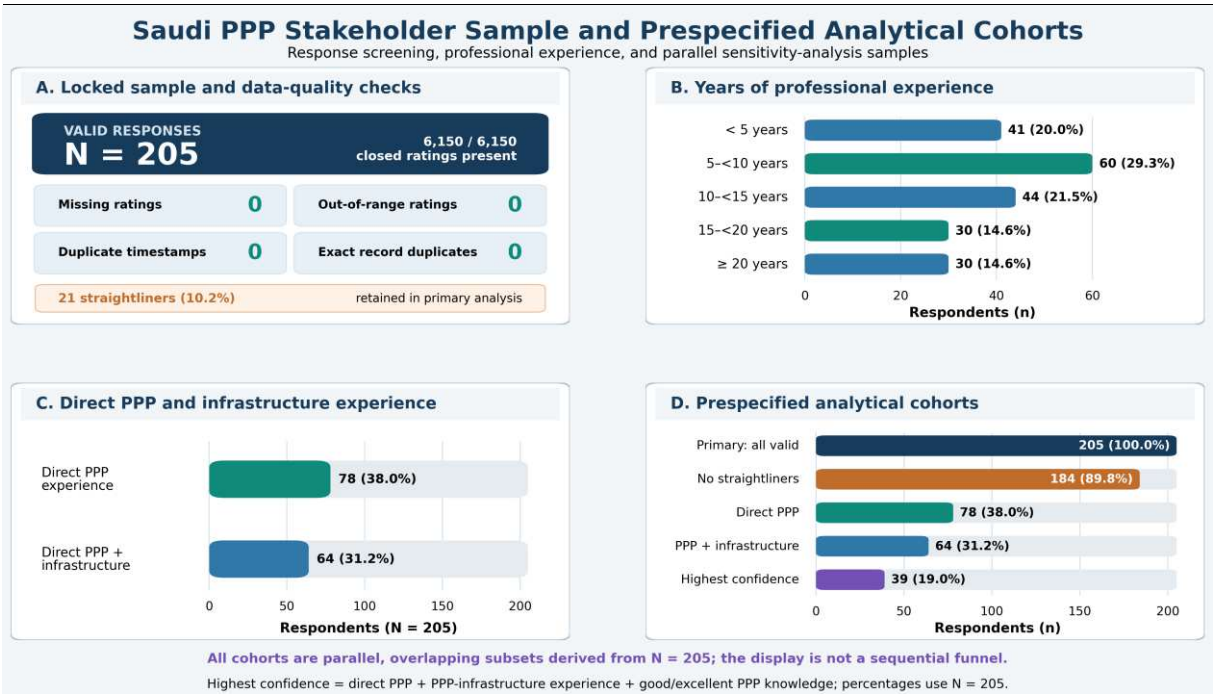


Fig. 10 Stakeholder sample composition and analytical cohorts. Composition of the Saudi PPP stakeholder sample and prespecified analytical cohorts. Panel A summarizes the locked sample and response screening; Panel B reports professional-experience duration; Panel C shows direct PPP and combined PPP-infrastructure experience; and Panel D displays the parallel, overlapping primary and sensitivity-analysis cohorts. Percentages use N = 205; the cohorts do not constitute a sequential funnel.

3.3. Descriptive Probability, Impact, and Criticality Results

Mean probability was between 3.00 and 3.42 for the 15 risks. Mean impact was between 3.39 and 3.72 Cost escalation and overrun (R01) had the highest mean probability (3.42; RII(P,i)=0.685) whilst political and security instability (R08) had the highest proportion of high probability ratings (47.8%). The mean impact was highest for coordination and communication failure (R03) with 3.72; RII(I,i) = 0.743) and the highest percentage of impact ratings of 4–5 (61.5%). Because the underlying ratings are ordinal, we kept medians, interquartile ranges, and full five-category distributions.

The paired respondent level products gave CI values ranging from 0.446 to 0.524. The most significant factor was political and security instability (R08: CI = 0.524; RI = 13.11), followed by cost escalation and overrun (R01: CI = 0.517; RI = 12.93) and coordination and communication failure (R03: CI = 0.515; RI = 12.88). Regulatory and legal uncertainty (R13) had the lowest CI (0.446) followed by R02 and R06 (tie) with CI = 0.482 (average rank = 9.5). The complete factor level results are presented in CI order in Table 14 and the ordinal distributions are linked to the probability-impact map in Figure 11. The bubble size shows the primary CI at the respondent level; the color

shows the predetermined dominance-based tiers, not arbitrary map quadrants. Supplementary Data S4 provides participant-level ratings and full descriptive and diagnostic output.

Table 14 Probability, impact, and respondent-level criticality results.

Panel A. Probability ratings in primary criticality-rank order

Rank	Risk code	Normalized risk factor	Mean	SD	Median [Q1–Q3]	Ratings 4–5 (%)	RII (P)
1	R08	Political and security instability	3.33	1.24	3 [3–4]	47.8	0.665
2	R01	Cost escalation and overrun	3.42	1.10	3 [3–4]	44.9	0.685
3	R03	Coordination and communication failure	3.27	1.09	3 [3–4]	37.1	0.655
4	R07	Planning, scope, and schedule deficiencies	3.25	1.19	3 [3–4]	41.5	0.650
5	R04	Capability and management-system weakness	3.28	1.06	3 [3–4]	42.0	0.657
6	R05	Resource, equipment, and maintenance constraints	3.23	1.16	3 [2–4]	41.5	0.647
7	R11	Design and technical deficiencies	3.22	1.14	3 [3–4]	37.1	0.644
8	R10	Approval and permitting delays	3.22	1.05	3 [3–4]	37.6	0.645
9.5	R02	Contractual and risk-allocation issues	3.20	1.07	3 [3–4]	35.1	0.639
9.5	R06	Land, site, and environmental constraints	3.18	1.17	3 [2–4]	41.0	0.636
11	R09	Demand and market risk	3.23	1.07	3 [3–4]	33.7	0.646
12	R15	Corruption and unethical practices	3.00	1.29	3 [2–4]	33.2	0.601
13	R14	Tax and tariff changes	3.12	1.17	3 [2–4]	37.1	0.623
14	R12	Government intervention and commitment risk	3.08	1.18	3 [2–4]	32.7	0.617
15	R13	Regulatory and legal uncertainty	3.07	1.15	3 [2–4]	31.7	0.615

Panel B. Impact ratings and respondent-level criticality

Risk code	Mean	SD	Median [Q1–Q3]	Ratings 4–5 (%)	RII (I)	CI_i	RI_i
R08	3.61	1.27	4 [3–5]	58.5	0.722	0.524	13.11
R01	3.56	1.08	4 [3–4]	52.7	0.711	0.517	12.93
R03	3.72	1.08	4 [3–5]	61.5	0.743	0.515	12.88
R07	3.60	1.14	4 [3–5]	54.1	0.720	0.502	12.54
R04	3.59	1.10	4 [3–4]	55.1	0.717	0.499	12.47
R05	3.56	1.13	4 [3–4]	55.6	0.712	0.494	12.35
R11	3.58	1.11	4 [3–4]	55.1	0.716	0.492	12.30
R10	3.49	1.11	4 [3–4]	51.2	0.699	0.483	12.07
R02	3.53	1.12	3 [3–4]	49.8	0.706	0.482	12.04
R06	3.53	1.19	4 [3–4]	54.1	0.706	0.482	12.04
R09	3.40	1.15	3 [3–4]	42.4	0.680	0.475	11.87
R15	3.71	1.26	4 [3–5]	58.5	0.741	0.474	11.85
R14	3.52	1.12	4 [3–4]	53.2	0.703	0.469	11.73
R12	3.49	1.20	4 [3–4]	53.2	0.699	0.467	11.68
R13	3.39	1.09	3 [3–4]	42.0	0.677	0.446	11.15

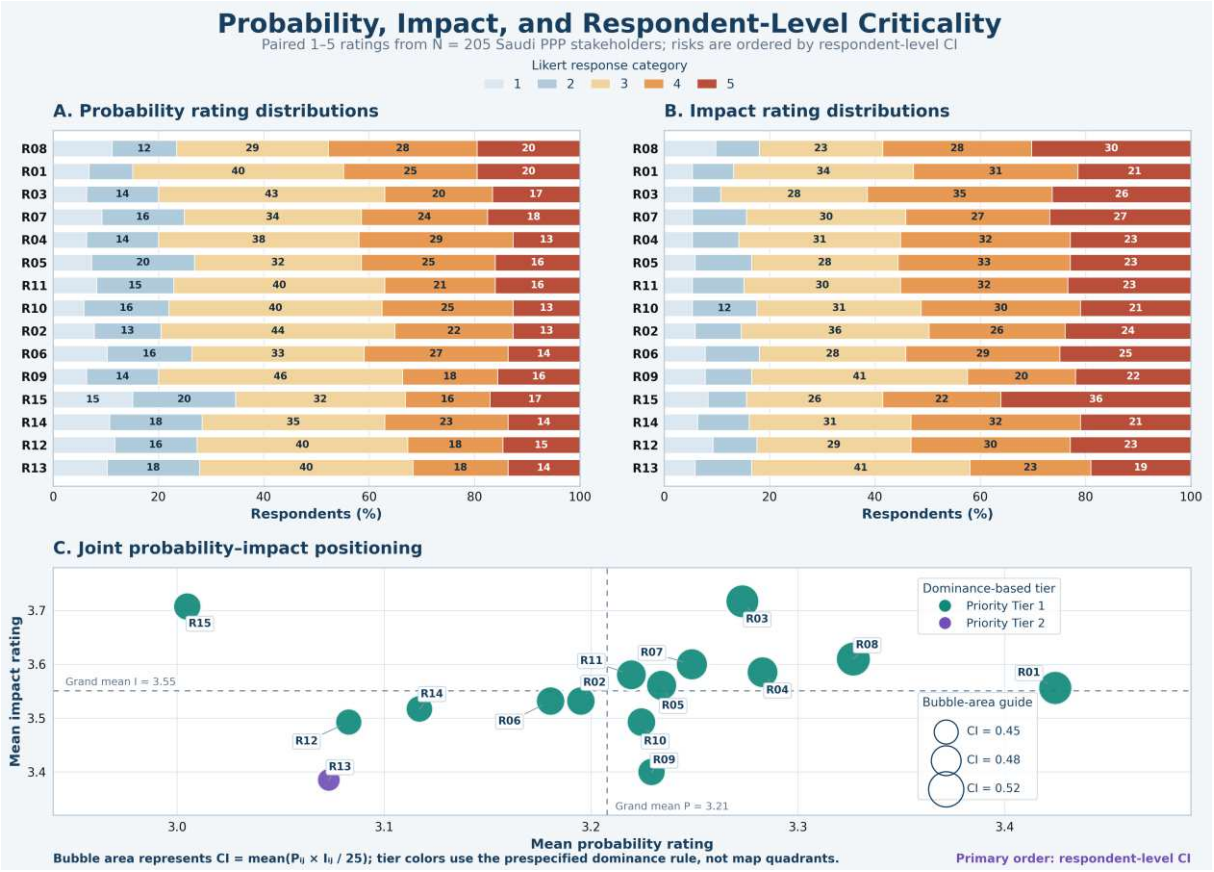


Fig. 11 Probability and impact distributions and joint positioning. Distribution and joint positioning of probability and impact assessments across the 15 PPP risk factors. Panels A and B show the five-category Likert distributions in descending respondent-level criticality order. Panel C plots mean probability against mean impact; bubble area is proportional to CI_i and color denotes the prespecified dominance-based Priority Tier.

3.4. Response Consistency and Paired Risk Differences

Both rating sets had high consistency in response pattern. The Cronbach's alpha for the 15 probability items was 0.942, and for the 15 impact items was 0.950. Corrected item–total correlations were 0.584 to 0.786 for probability and 0.645 to 0.793 for impact. Alpha-if-deleted values were consistently lower than the corresponding full-scale alpha (probability: 0.936–0.941; impact: 0.945–0.948). The coefficients are measures of the consistency of rating behavior and are not taken as evidence that the 15 substantive risks represent a single latent construct.

The tie-corrected Friedman test showed differences between the paired respondent-level criticality scores, $\chi^2(14) = 44.996$, $p < 0.001$, but Kendall's W = 0.0157 indicated a very small concordance effect. Of the 105 prespecified Wilcoxon contrasts, three remained significant after Holm correction: R08 > R13 (ΔCI = 0.0784; r_rb = 0.454), R03 > R13 (ΔCI = 0.0693; r_rb = 0.435) and R01 > R13 (ΔCI = 0.0714; r_rb = 0.386). The consistency and repeated-measures results are summarized in Table 15; the complete family of comparisons is shown in Supplementary Table S5.

Table 15 Response-pattern consistency and repeated-measures inference.

Panel A. Internal-consistency diagnostics

Item set	k	Cronbach's α	Corrected item–total r, range	α if item deleted, range
Probability ratings	15	0.942	0.584 (R01)–0.786 (R09)	0.936 (R09)–0.941 (R01)
Impact ratings	15	0.950	0.645 (R01)–0.793 (R10)	0.945 (R10)–0.948 (R01)

Panel B. Tie-corrected Friedman test and Kendall's coefficient of concordance

Metric	Value	df / family	p value / interpretation
Friedman χ^2 (tie-corrected)	44.996	14	$p < 0.001$
Tie-correction factor	0.743	—	Applied to χ^2
Kendall's W	0.0157	—	Very small concordance effect
Planned Wilcoxon contrasts	105	15 choose 2	Holm correction across the full family
Holm-significant contrasts	3	105	2.9% of contrasts

Panel C. Wilcoxon contrasts remaining significant after Holm correction

Higher-criticality risk	Lower-criticality risk	ΔCI	Wilcoxon W	Holm p	r_rb
R08 — Political and security instability	R13 — Regulatory and legal uncertainty	0.0784	2116.5	0.0012	0.454
R03 — Coordination and communication failure	R13 — Regulatory and legal uncertainty	0.0693	2478.0	0.0015	0.435
R01 — Cost escalation and overrun	R13 — Regulatory and legal uncertainty	0.0714	3249.0	0.0056	0.386

3.5. Bootstrap Rank Uncertainty and Priority Tiers

The entire set of paired scores for each respondent was retained using participant-vector bootstrap resampling (B = 5,000; fixed seed = 20260809). Table 16 shows the main respondent-level criticality estimate and its percentile confidence interval, median and 95% rank interval, first-place and Top-3 selection probabilities, and prespecified dominance-based tier for each risk. All rows and figure panels maintain the primary CI order. Complete bootstrap scores, ranks, dominance probabilities and tier-construction records are presented in Supplementary Data S6.

Political and security instability (R08) was the top ranked risk with CI = 0.5243 (95% bootstrap CI [0.4835, 0.5647]), median bootstrap rank of 1 (95% interval 1–5), P(Rank = 1) = 51.7%, and P(Top 3) = 89.2%. R01 ranked second (CI = 0.5173; P(Top 3) = 75.6%) and R03 ranked third (CI = 0.5151; P(Top 3) = 76.6%). Beyond these leading factors, Figure 12 shows that the intervals overlapped substantially, supporting uncertainty-aware reporting rather than a deterministic rank cut-off.

Only three directed relationships satisfied both pre-specified dominance conditions: R08 > R13, R01 > R13, and R03 > R13 ($D_{a\beta}$ = 1.000 for each, Holm-adjusted paired contrasts significant). The iterative non-dominated-front algorithm assigned 14 risks to Tier 1 and R13 alone to Tier 2 (CI = 0.4459; median rank = 15; 95% rank interval 13–15; P(Top 3) = 0%). The direct visual evidence for this tier structure is shown in Figure 13. R08, R01 and R03 can be reported as the executive Top 3 with their selection probabilities, Tier 1 as the primary scientific classification.

Table 16 Bootstrap uncertainty, rank-selection probabilities, and priority tiers.

Panel A. Criticality estimates, confidence intervals, and priority tiers

Rank	Risk code	Normalized risk factor	CI_i	95% bootstrap CI	Priority tier
1	R08	Political and security instability	0.5243	[0.4835, 0.5647]	1
2	R01	Cost escalation and overrun	0.5173	[0.4812, 0.5528]	1
3	R03	Coordination and communication failure	0.5151	[0.4796, 0.5499]	1
4	R07	Planning, scope, and schedule deficiencies	0.5017	[0.4642, 0.5401]	1
5	R04	Capability and management-system weakness	0.4989	[0.4646, 0.5337]	1
6	R05	Resource, equipment, and maintenance constraints	0.4939	[0.4570, 0.5309]	1
7	R11	Design and technical deficiencies	0.4919	[0.4542, 0.5288]	1
8	R10	Approval and permitting delays	0.4827	[0.4474, 0.5192]	1
9.5	R02	Contractual and risk-allocation issues	0.4818	[0.4468, 0.5165]	1
9.5	R06	Land, site, and environmental constraints	0.4818	[0.4433, 0.5179]	1
11	R09	Demand and market risk	0.4747	[0.4367, 0.5114]	1
12	R15	Corruption and unethical practices	0.4741	[0.4330, 0.5128]	1
13	R14	Tax and tariff changes	0.4693	[0.4318, 0.5067]	1
14	R12	Government intervention and commitment risk	0.4671	[0.4287, 0.5052]	1
15	R13	Regulatory and legal uncertainty	0.4459	[0.4101, 0.4814]	2

Panel B. Bootstrap rank selection and dominance relationships

Risk code	Median rank	95% rank interval	P(Rank = 1) (%)	P(Top 3) (%)	Dominance relationship
R08	1	[1.0, 5.0]	51.7	89.2	Dominates R13

Risk code	Median rank	95% rank interval	P(Rank = 1) (%)	P(Top 3) (%)	Dominance relationship
R01	2	[1.0, 7.0]	25.8	75.6	Dominates R13
R03	3	[1.0, 6.0]	17.1	76.6	Dominates R13
R07	5	[1.0, 10.0]	2.6	25.2	No joint relation
R04	5	[2.0, 11.0]	1.6	16.0	No joint relation
R05	6	[2.0, 12.0]	0.5	7.5	No joint relation
R11	7	[3.0, 12.0]	0.6	7.0	No joint relation
R10	9	[4.5, 14.0]	0.0	0.6	No joint relation
R02	9	[5.0, 14.0]	0.0	0.4	No joint relation
R06	9	[4.0, 14.0]	0.0	0.7	No joint relation
R09	11	[6.0, 14.0]	0.0	0.0	No joint relation
R15	11	[5.0, 15.0]	0.0	0.8	No joint relation
R14	12	[6.0, 15.0]	0.0	0.2	No joint relation
R12	13	[7.0, 15.0]	0.0	0.0	No joint relation
R13	15	[13.0, 15.0]	0.0	0.0	Dominated by R01, R03, R08

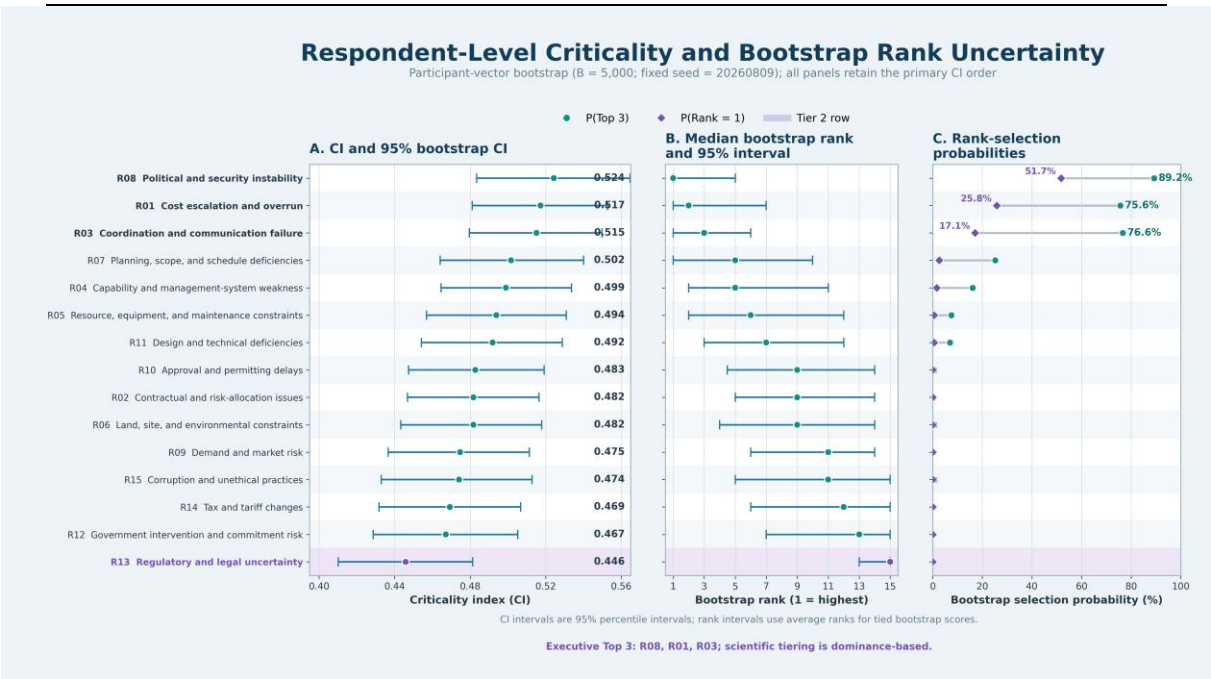


Fig. 12 Criticality estimates and bootstrap rank uncertainty. Respondent-level criticality estimates and bootstrap rank uncertainty of the Saudi PPP risk factors. Panel A reports CI_i with 95% percentile bootstrap intervals; Panel B reports median bootstrap ranks with 95% rank intervals; and Panel C reports P(Top 3) and P(Rank = 1).

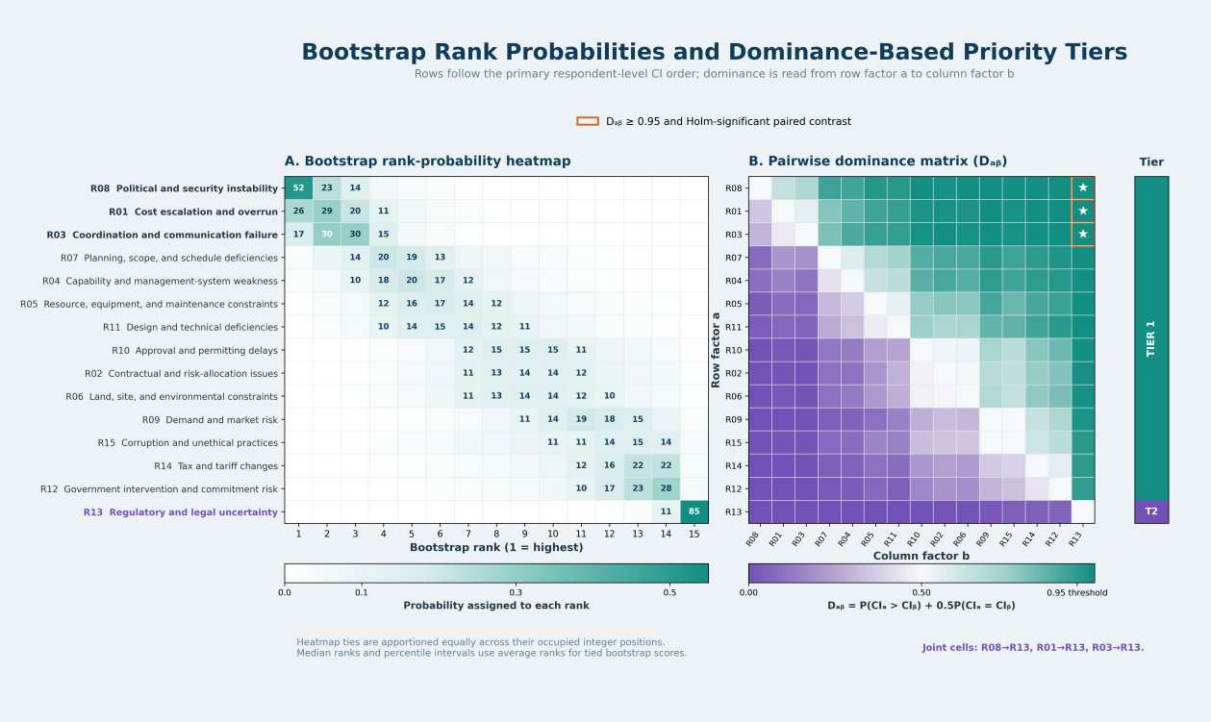


Fig. 13 Bootstrap rank probabilities and dominance-based priority tiers. Bootstrap rank probabilities and dominance-based priority-tier structure. Panel A displays integer-rank probability mass; tied bootstrap positions are apportioned equally for this heatmap, while summary ranks use average ranks. Panel B displays $D_{a\beta} = P(CI_a > CI_\beta) + 0.5P(CI_a = CI_\beta)$, with outlined starred cells marking $D_{a\beta} \geq 0.95$ plus a Holm-significant paired contrast.

3.6. Literature–Market Evidence Alignment

Agreement between corrected study-level recurrence (F_i) and respondent-level market criticality (CI_i) was positive but not statistically significant under the prespecified label-permutation test. Spearman's ρs was 0.400, and the two-sided test based on 100,000 permutations yielded p = 0.1403. Participant-vector bootstrapping (B = 5,000) produced a median ρs of 0.379 and a 95% percentile interval of [0.067, 0.594]. The bootstrap interval conditions on the fixed literature scores and quantifies participant-sampling uncertainty in market ratings, whereas the permutation test evaluates factor-label concordance across only 15 risks. Accordingly, the evidence indicates incomplete positive alignment but does not reject the null of no label-wise concordance at α = 0.05.

To interpret convergence and divergence without reintroducing literature recurrence into a single primary score, F_i and CI_i were converted to tied-rank percentiles L_i and E_i, respectively. The prespecified threshold τ = 0.50 separated high from low evidence on each axis. Table 17 reports the corrected coordinates and alignment class for every factor. Complete factor values, all 100,000 permutation statistics, all 5,000 participant-bootstrap correlations, and IPS sensitivity outputs are provided in Supplementary Data S7.

Table 17 Literature–market concordance and risk-factor alignment.

Code[1]	F_i	L_i	CI_i	E_i	Alignment class
R08	9	0.357	0.5243	1.000	Saudi / emerging risks
R01	11	0.929	0.5173	0.929	Core risks
R03	11	0.929	0.5151	0.857	Core risks
R07	9	0.357	0.5017	0.786	Saudi / emerging risks
R04	11	0.929	0.4989	0.714	Core risks
R05	10	0.679	0.4939	0.643	Core risks
R11	9	0.357	0.4919	0.571	Saudi / emerging risks
R10	8	0.071	0.4827	0.500	Saudi / emerging risks
R02	10	0.679	0.4818	0.393	Established but context-moderated risks
R06	10	0.679	0.4818	0.393	Established but context-moderated risks
R09	8	0.071	0.4747	0.286	Secondary risks

Code[1]	F_i	L_i	CI_i	E_i	Alignment class
R15	9	0.357	0.4741	0.214	Secondary risks
R14	8	0.071	0.4693	0.143	Secondary risks
R12	9	0.357	0.4671	0.071	Secondary risks
R13	10	0.679	0.4459	0.000	Established but context-moderated risks

The corrected classification identified four Core risks (R01, R03, R04, and R05), four Saudi/emerging risks (R08, R07, R11, and R10), three Established but context-moderated risks (R02, R06, and R13), and four Secondary risks (R09, R15, R14, and R12). R08 combined the highest market percentile (E_i = 1.000) with a lower literature percentile (L_i = 0.357), whereas R13 combined a high literature percentile (L_i = 0.679) with the lowest market percentile (E_i = 0.000). These contrasts support separate interpretation of the two evidence dimensions despite the absence of statistically significant overall concordance.

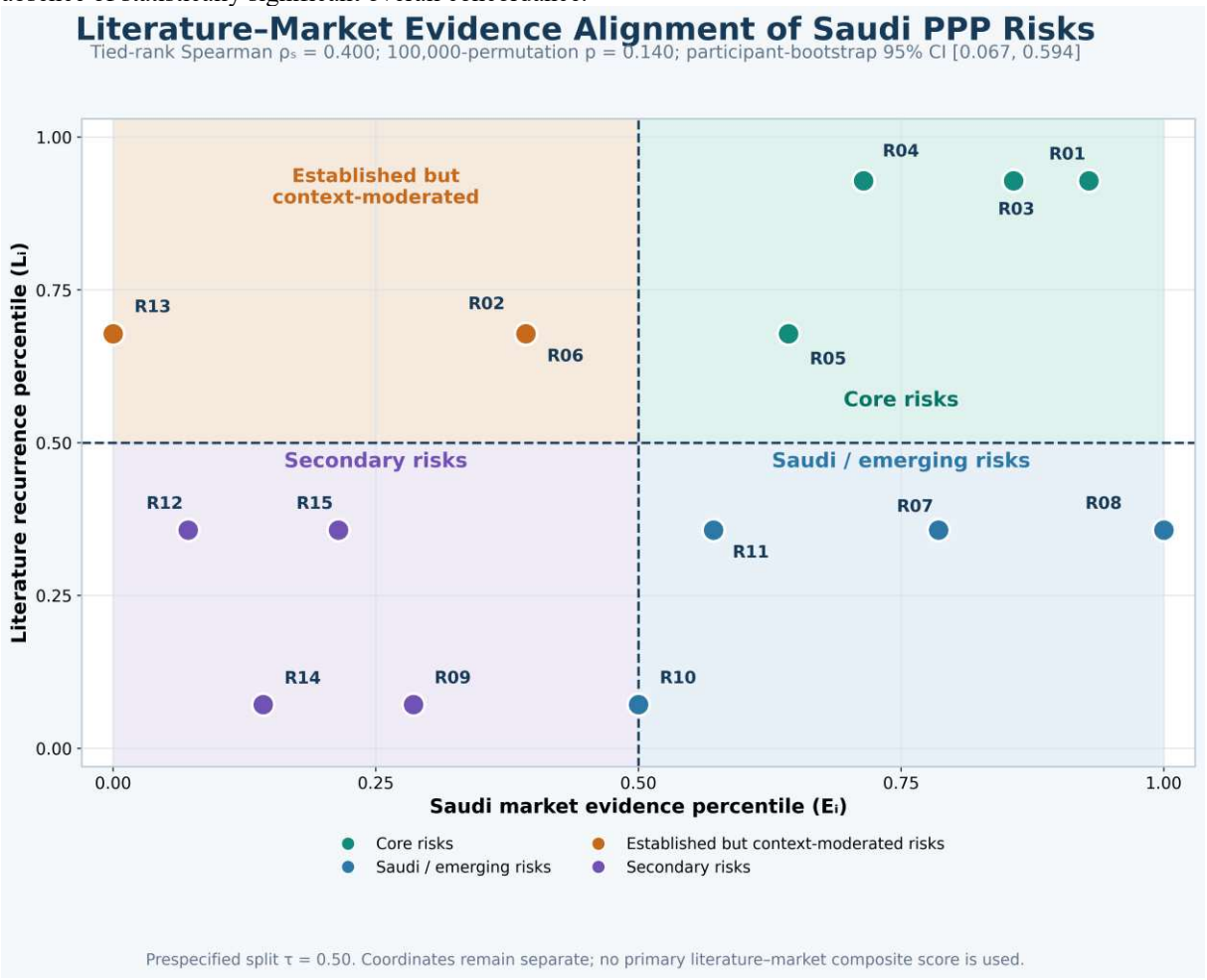


Fig. 14 Literature–market evidence alignment. Corrected literature–market evidence alignment of the 15 Saudi PPP risk factors. The horizontal axis reports the Saudi market percentile E_i and the vertical axis reports the literature recurrence percentile L_i; dashed lines show the prespecified $\tau = 0.50$ split. The corrected tied-rank concordance was $\rho s = 0.400$ (100,000-permutation p = 0.1403; participant-bootstrap 95% CI [0.067, 0.594]). No primary composite score was used.

As visualized in Figure 14, the quadrant structure preserves literature and market evidence as distinct analytical dimensions. No primary composite score was calculated, thereby avoiding double counting because the 15 questionnaire factors had already been shortlisted from the literature base.

3.7. Subgroup Differences and Robustness of Risk Rankings

Subgroup inference used the respondent-level criticality index and controlled the 15 risk-wise tests within each grouping variable by the Benjamini-Hochberg procedure (Table 18, Panel A; Figure 15A). Direct PPP experience was associated with six FDR-significant differences (R01-R05 and R12; q = 0.00484-0.0331), all indicating higher criticality among directly experienced participants (r_rb = 0.204-0.296). The strongest contrast was R01 (U = 6419.5, q = 0.00484, r_rb = 0.296; bootstrap 95% CI [0.139, 0.443]). PPP-infrastructure experience alone yielded

no FDR-significant risk (minimum q = 0.213). Years of experience differentiated 10 risks (epsilon_H^2 = 0.035-0.057), whereas PPP knowledge differentiated all 15 (epsilon_H^2 = 0.057-0.187). The largest knowledge effect occurred for R02, H(4) = 41.409, q = 3.32e-7, epsilon_H^2 = 0.187. Dunn-Holm comparisons localized 12 contrasts across nine risks for experience duration and 65 contrasts for knowledge. All subgroup tests, BH-FDR values, effect estimates, and eligible Dunn-Holm comparisons are provided in Supplementary Data S8.

The full risk order was unchanged after excluding the 21 straightliners ($\rho\rho_s$ = 1.000; maximum rank change = 0). Relative to the primary ranking, correlations remained high for the direct-PPP, combined PPP-infrastructure, and highest-confidence cohorts ($\rho\rho_s$ = 0.858, 0.886, and 0.903, respectively), and was 0.972 under product-of-means scoring (Table 18, Panel B; Figure 15B). The primary Top 3 set (R08, R01, and R03) was preserved in all scenarios except the highest-confidence cohort, where R05 replaced R03 (Jaccard = 0.50). The largest movements were R10 in the direct-PPP cohort (five ranks), R12 in the combined PPP-infrastructure cohort (4.5 ranks due to tied ranks), and R05 in the highest-confidence cohort (four ranks). All 14 primary Tier 1 factors were retained in every scenario; the n = 39 cohort expanded Tier 1 to all 15 factors, reflecting greater rank uncertainty rather than stronger evidence of equivalence. Subgroup findings therefore characterize heterogeneity but do not replace the full-sample priority inference. All six scenario scores, ranks, and stability metrics are provided in Supplementary Data S9.

Table 18 Subgroup effects and robustness of risk-priority rankings.

Panel A. FDR-controlled subgroup comparisons

Grouping variable	Test and group sizes	BH-FDR result across 15 risks	Minimum q (risk)	Effect-size and post-hoc summary
Direct PPP experience	Mann-Whitney U 78 direct vs 127 other	6/15: R01, R02, R03, R04, R05, R12	0.00484 (R01)	r_rb = 0.204-0.296; all six were higher in the direct-PPP group. Largest: R01 = 0.296.
PPP-infrastructure experience	Mann-Whitney U 87 yes vs 118 no	0/15 after BH-FDR	0.2131 (R01)	Largest |r_rb| = 0.198 (R01); not FDR-significant.
Years of professional experience	Kruskal-Wallis H (df = 4) 41 / 60 / 44 / 30 / 30	10/15: R01, R02, R03, R04, R05, R06, R07, R08, R10, R13	0.0245 (R03)	epsilon_H^2 = 0.035-0.057; 12 Dunn-Holm contrasts across 9 risks.
Self-reported PPP knowledge	Kruskal-Wallis H (df = 4) 35 / 24 / 73 / 44 / 29	15/15: all risk factors	3.32e-07 (R02)	epsilon_H^2 = 0.057-0.187; 65 Dunn-Holm contrasts. Largest: R02 = 0.187.

Panel B. Rank stability against the primary respondent-level criticality ranking

Scenario	n	Spearman $\rho\rho_s$	Maximum rank change	Top-3 Jaccard	Primary Tier 1 retained	Scenario Tier 1 n	Executive Top 3
Primary	205	1.000	0 (reference)	1.00	14/14 (100%)	14	R08 / R01 / R03
No straightliners	184	1.000	0 (all unchanged)	1.00	14/14 (100%)	14	R08 / R01 / R03
Direct PPP	78	0.858	5 (R10)	1.00	14/14 (100%)	14	R01 / R03 / R08
PPP + infrastructure	64	0.886	4.5 (R12)	1.00	14/14 (100%)	14	R01 / R08 / R03
Highest confidence	39	0.903	4 (R05)	0.50	14/14 (100%)	15	R08 / R05 / R01
Product of means	205	0.972	2 (R08)	1.00	14/14 (100%)	14	R01 / R03 / R08

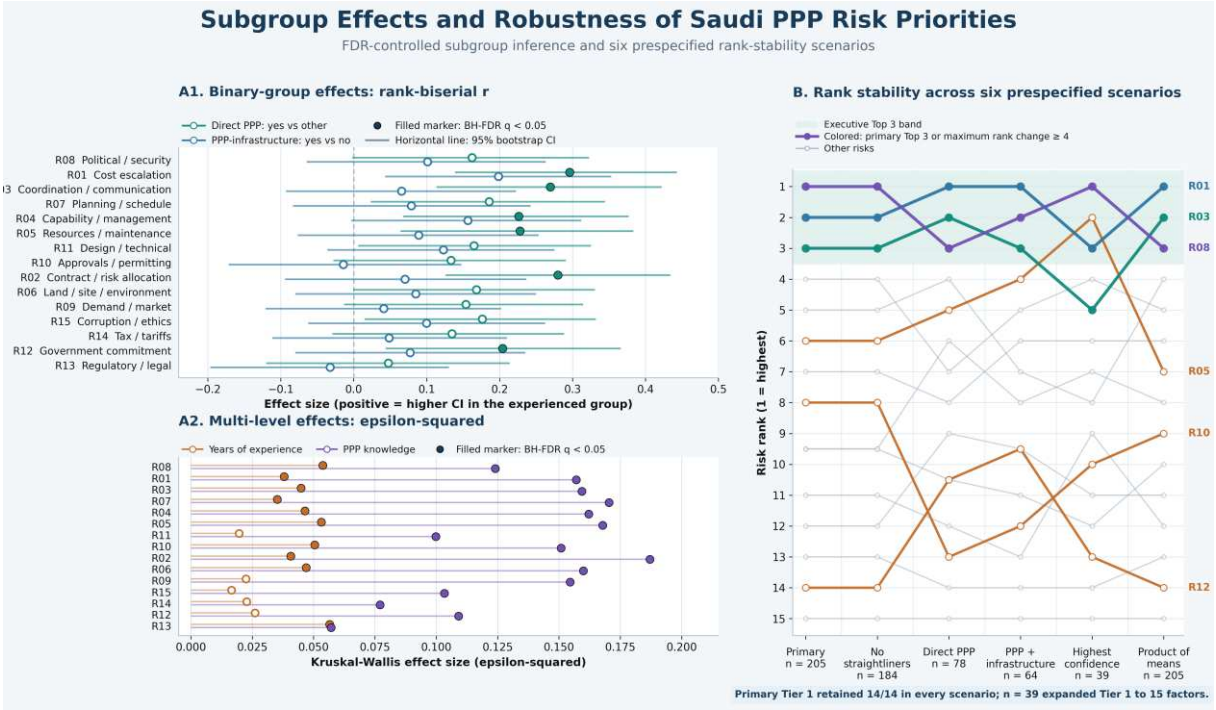


Fig. 15 Subgroup effects and robustness of risk priorities. Subgroup effects and robustness of Saudi PPP risk priorities across prespecified analytical scenarios. Panel A1 reports rank-biserial effects for binary comparisons with 95% group-stratified participant-bootstrap intervals (B = 5,000); positive values indicate higher respondent-level criticality in the experienced group. Panel A2 reports epsilon-squared for years of experience and PPP knowledge. Filled markers survived BH-FDR at q < 0.05; Dunn contrasts were Holm-adjusted. Panel B traces tied criticality ranks across the six scenarios; highlighted lines identify the primary executive Top 3 and factors moving by at least four ranks.

3.8. Qualitative Findings from Open-Ended Responses

Of the 205 valid respondents, 81 (39.5%) supplied nonblank open-ended text. Screening identified 124 blank entries and excluded 54 of the 81 submitted responses as blank-equivalent, yes/no-only, courtesy, comprehensiveness, or off-topic statements. This left 27 substantive responses (13.2% of the full sample; 33.3% of nonblank submissions). Segmentation yielded 49 meaning units, 35 distinct open codes, and 12 final themes. Participant-level prevalence in Table 19 and Figure 16A therefore uses Nq = 27; repeated mentions by one respondent do not increase a theme's count.

Leadership, workforce competence, and decision authority was the most prevalent theme (n = 10, 37.0%), followed by institutional commitment and administrative/contracting continuity (n = 7, 25.9%). Liquidity/payment/macroeconomic exposure and political/security/operational continuity were each mentioned by four participants (14.8%), while technology/cybersecurity/obsolescence and demand realism/social license were each mentioned by three (11.1%). The remaining themes were reported by one or two participants. Nine themes matched or contextually extended at least one of the 15 questionnaire risks; three were classified as emergent candidates: quality and safety assurance, technology/cybersecurity/obsolescence, and institutional memory/interagency knowledge continuity.

The qualitative findings were translated into proposed controls and KRIs (Table 19; Figure 16B) but were not used to recalculate CI values, quantitative ranks, or dominance tiers. No independent second-coder matrix, achieved double-coded proportion, or reconciliation log was available in the locked study files; Cohen's kappa was therefore neither calculated nor claimed for this results section. The full screening, coding, prevalence, mapping, and agreement-status audit is provided in Supplementary Data S10.

Table 19 Qualitative theme prevalence and framework implications.

Theme	n	%	Matched questionnaire factor(s)	Interpretive class	Framework action / KRI implication	Illustrative quotation
T01. Leadership, workforce competence, and decision authority	10	37.0	R04	Contextual extension	Delegated authority and competence matrix; KRIs: qualified-role coverage and decision-turnaround time.	*"The absence of a real leader for such partnerships is a major risk." [Q131]*
T02. Institutional commitment and administrative/contracting continuity	7	25.9	R12; R02; R10; R03	Contextual extension	Continuity and change control; KRIs: award-to-signature days, leadership turnover, and unapproved changes.	*"The project may be awarded while the contract remains unsigned for*

Theme	n	%	Matched questionnaire factor(s)	Interpretive class	Framework action / KRI implication	Illustrative quotation
						more than a year." [Q028]
T03. Liquidity, payment, and macro-financial exposure	4	14.8	R01; R12; R14	Contextual extension	Payment and macro-financial safeguards; KRIs: delay days, liquidity cover, and inflation/FX variance.	*"Financial support and delayed payments." [Q032]*
T12. Political, security, and operational continuity	4	14.8	R08; R05	Contextual extension	Security and supply-continuity plan; KRIs: stoppage days, alerts, and critical-material lead time.	*"Political or security instability directly affects access to project materials." [Q159]*
T07. Technology, cybersecurity, and obsolescence	3	11.1	No direct match	Emergent candidate	Technology, cyber, and upgrade plan; KRIs: unsupported assets, patch compliance, and incidents.	*"There is a risk that technology becomes obsolete over time." [Q114]*
T08. Demand realism and social licence	3	11.1	R09	Contextual extension	Independent demand validation and engagement triggers; KRIs: forecast error, complaints, and fee opposition.	*"Local-community resistance may have a medium-to-high impact." [Q103]*
T04. Regulatory volatility and sustainability compliance	2	7.4	R13; R14; R06	Contextual extension	Regulatory/sustainability register; KRIs: high-severity changes and overdue compliance actions.	*"Sudden legislative and regulatory changes may impose unexpected costs." [Q051]*
T05. Quality and safety assurance	2	7.4	No direct match (closest: R11/R05)	Emergent candidate	Independent quality/HSE assurance; KRIs: NCR closure time, audit findings, and safety incidents.	*"Quality and safety can affect both schedule and project cost." [Q091]*
T06. Environmental and climate resilience	2	7.4	R06	Contextual extension	Climate and environmental controls; KRIs: permit status, hazard exceedances, and adaptation closure.	*"Natural disasters and climate change." [Q103]*
T09. Procurement integrity and contractor-market quality	2	7.4	R15; R04; R02	Contextual extension	Integrity screening and bidder-quality controls; KRIs: bid dispersion, contractor health, and complaints.	*"Execution price wars damage small contractors and increase market concentration." [Q174]*
T10. Planning and schedule discipline	1	3.7	R07	Existing	Time-bound baseline and stage gates; KRIs: schedule variance, missed milestones, and forecast drift.	*"A specific completion time and a future plan are needed." [Q115]*
T11. Institutional memory and inter-agency knowledge continuity	1	3.7	No direct match (closest: R03/R04)	Emergent candidate	Common data environment and lessons repository; KRIs: completeness, retrieval time, and reuse.	*"Failure to preserve infrastructure plans or project history weakens future development." [Q169]*

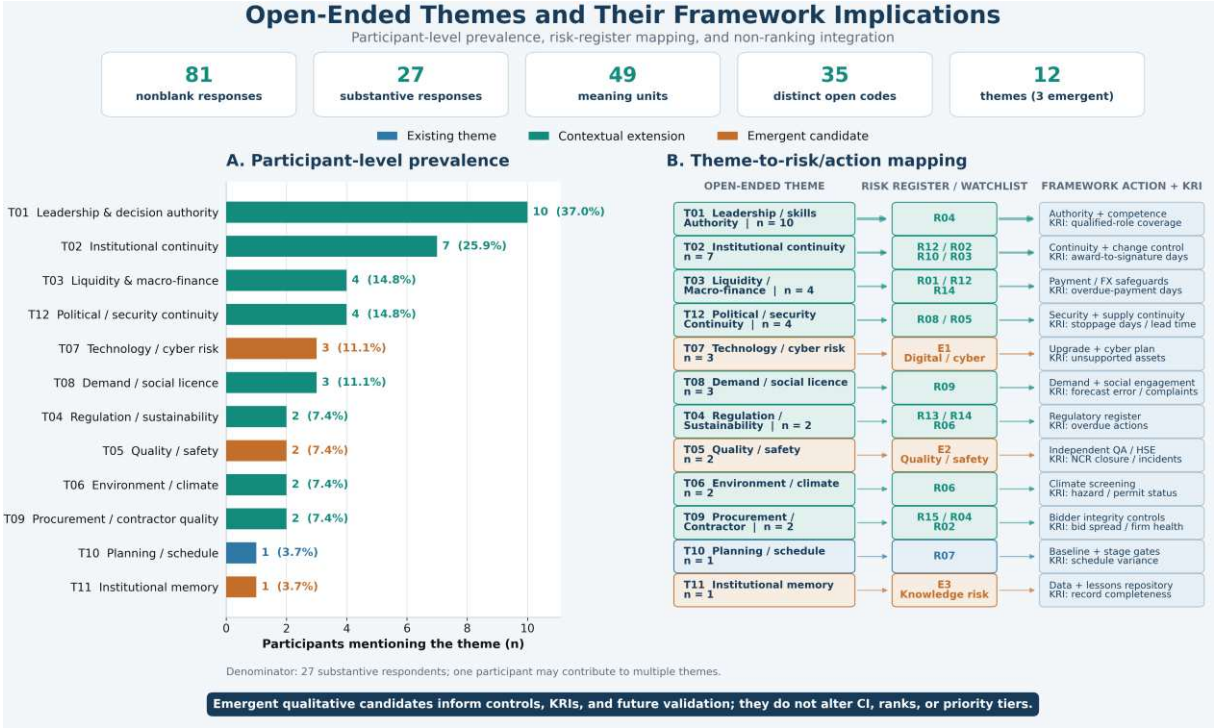


Fig. 16 Qualitative theme prevalence and framework mapping. Prevalence and framework mapping of themes derived from the open-ended responses. Panel A reports participant-level prevalence among the 27 substantive respondents, with each participant counted once per theme. Panel B maps themes to the existing 15-risk register or to an emergent-risk watchlist and then to framework actions and KRIs. Emergent qualitative candidates were retained for framework enrichment and future empirical assessment; they were not inserted into criticality scores, quantitative ranks, or dominance-based priority tiers.

3.9. Proposed Saudi PPP Risk Governance Framework

The integrated analytical outputs were translated into the Proposed Evidence-Based Saudi PPP Risk Governance Framework and its associated operational governance register. The dominance-based procedure identified a 14-factor Tier 1 non-dominated set. R13 (regulatory and legal uncertainty) remained in Tier 2 but was retained within the complete 15-risk register to ensure comprehensive lifecycle monitoring. Within Tier 1, the executive attention set comprised R08, R01, and R03, with respondent-level CI values of 0.524, 0.517, and 0.515 and bootstrap Top-3 probabilities of 89.2%, 75.6%, and 76.6%, respectively.

Table 20 presents the Tier 1 risks in descending order of the primary respondent-level CI and distinguishes the locked analytical evidence from the proposed governance specifications. For each risk, the framework identifies the active lifecycle stages, one accountable functional owner for each active risk-stage, supporting actors, contractual allocation, prevention–mitigation–contingency controls, key risk indicators, trigger thresholds, escalation routes, and post-control residual-risk review. The open-ended findings further informed the framework's emphasis on delegated authority, institutional continuity, liquidity and payment safeguards, political-security continuity, quality assurance, technology resilience, and institutional memory. The complete 15-risk proposed governance register is provided in Supplementary Table S11.

Figure 17 presents the populated framework as an integrated six-stage lifecycle governance system. It links the retained risk factors to stage-specific accountability, risk-specific three-layer response controls, early-warning indicators, trigger thresholds, and a common escalation pathway from the accountable risk owner to executive or external decision-makers. Following any trigger breach or material change in control effectiveness, the framework requires a project-specific post-control reassessment, residual-risk re-rating, and a dated update to the master governance register.

Table 20 Proposed operational governance specifications for Tier 1 risks.

Panel A. Analytical priority evidence and proposed lifecycle accountability and contractual allocation

Risk factor	Priority evidence	Lifecycle stage(s)	Accountable owner	Supporting actors	Contractual allocation
R08 Political and security instability	Tier 1 | rank 1 CI 0.524 (95% CI 0.484-0.565) Fi=9 | Saudi/emerging	L1-L6; event-driven	Contracting-authority security and resilience lead	SPV; EPC/O&M contractors; security agencies; insurers; critical suppliers	Shared: public side bears defined political/change-in-law and force-majeure relief; private side manages site security and continuity controls.

Risk factor	Priority evidence	Lifecycle stage(s)	Accountable owner	Supporting actors	Contractual allocation
R01 Cost escalation and overrun	Tier 1 | rank 2 CI 0.517 (95% CI 0.481-0.553) Fi=11 | Core	L1-L5	Authority commercial lead (L1-L2) - SPV commercial director (L3-L5)	Cost consultant; EPC/O&M contractors; procurement; finance; lenders	Private for controllable delivery and operating costs; shared/indexed for specified inflation, currency, authority change, and compensation events.
R03 Coordination and communication failure	Tier 1 | rank 3 CI 0.515 (95% CI 0.480-0.550) Fi=11 | Core	L1-L6	PPP project director / designated interface manager	Contracting authority; SPV; EPC/O&M contractors; advisers; affected public entities	Shared through a binding RACI, interface obligations, decision service levels, and a common issue-resolution protocol.
R07 Planning, scope, and schedule deficiencies	Tier 1 | rank 4 CI 0.502 (95% CI 0.464-0.540) Fi=9 | Saudi/emerging	L1-L4	Authority PMO (L1-L3) - SPV programme director (L4)	Design team; EPC contractor; cost consultant; independent engineer; permitting leads	Public side retains baseline scope and authority dependencies; private side bears delivery planning and schedule performance after close.
R04 Capability and management-system weakness	Tier 1 | rank 5 CI 0.499 (95% CI 0.465-0.534) Fi=11 | Core	L1-L6	Lead organization's project director, with contracting-authority PPP PMO oversight	HR/workforce leads; technical advisers; SPV; contractors; independent assurance	Each party bears capability risk within its organization; contract sets minimum competence, key-person, substitution, and assurance requirements.
R05 Resource, equipment, and maintenance constraints	Tier 1 | rank 6 CI 0.494 (95% CI 0.457-0.531) Fi=10 | Core	L3-L6	SPV asset and O&M director	EPC/O&M contractors; suppliers; logistics; authority contract manager; insurers	Private for supply, equipment availability, spares, and lifecycle maintenance; relief limited to expressly defined public or force-majeure events.
R11 Design and technical deficiencies	Tier 1 | rank 7 CI 0.492 (95% CI 0.454-0.529) Fi=9 | Saudi/emerging	L2-L4	SPV design manager	EPC designer; independent engineer; authority technical team; operator; specialist reviewers	Private for design integration and fitness for purpose; public side retains output-specification and authority-directed change risk.
R10 Approval and permitting delays	Tier 1 | rank 8 CI 0.483 (95% CI 0.447-0.519) Fi=8 | Saudi/emerging	L1-L4	Contracting-authority permits and interagency-interface lead	Relevant regulators/municipalities; SPV; designers; environmental and utility leads	Public for sovereign approvals and authority-caused delay; private for complete, accurate, and timely applications and compliance evidence.
R02 Contractual and risk-allocation issues	Tier 1 | rank 9.5 CI 0.482 (95% CI 0.447-0.516) Fi=10 | Established but context-moderated	L1-L6	Authority commercial/legal lead (L1-L3) -> contract manager (L4-L6)	PPP unit; SPV; legal and technical advisers; lenders; insurers; dispute board	Allocate each risk to the party best able to control it; record relief, compensation, sharing, insurance, step-in, and dispute mechanisms explicitly.
R06 Land, site, and environmental constraints	Tier 1 | rank 9.5 CI 0.482 (95% CI 0.443-0.518) Fi=10 | Established but context-moderated	L1-L4 and L6	Authority land/environment lead (L1-L3) -> SPV HSE/site-interface lead (L4/L6)	Land agencies; municipalities; environmental regulator; communities;	Public for title, access, and disclosed pre-existing constraints; private for site management and compliance

Risk factor	Priority evidence	Lifecycle stage(s)	Accountable owner	Supporting actors	Contractual allocation
				designers; independent engineer	after handover; shared for defined unknown conditions.
R09 Demand and market risk	Tier 1 | rank 11 CI 0.475 (95% CI 0.437-0.511) Fi=8 | Secondary	L1-L3 and L5	SPV commercial/revenue director, with authority affordability oversight	Demand adviser; operator; lenders; tariff authority; stakeholder-engagement team	Private or shared according to tariff/availability structure and control capacity; avoid blanket transfer where demand is materially policy-dependent.
R15 Corruption and unethical practices	Tier 1 | rank 12 CI 0.474 (95% CI 0.433-0.513) Fi=9 | Secondary	L2-L6	Authority integrity officer (L2) -> SPV compliance officer (L3-L6)	Procurement; internal audit; legal; contractors; lenders; competent oversight bodies	Each party bears responsibility for its own conduct; contract preserves audit, disclosure, suspension, step-in, termination, and recovery rights.
R14 Tax and tariff changes	Tier 1 | rank 13 CI 0.469 (95% CI 0.432-0.507) Fi=8 | Secondary	L2-L5	SPV finance and tax director	Authority commercial lead; tax advisers; lenders; tariff regulator; legal counsel	Private for ordinary tax compliance and forecast risk; shared/public relief only for expressly defined discriminatory or general change-in-law events.
R12 Government intervention and commitment risk	Tier 1 | rank 14 CI 0.467 (95% CI 0.429-0.505) Fi=9 | Secondary	L1-L6	Contracting-authority senior responsible owner (SRO)	Sponsor ministry/entity; PPP unit; treasury/finance; contract manager; SPV; lenders	Public for authorized commitments, approvals, payments, and public-side changes; private for its own compliance and performance obligations.

892 **Panel B. Proposed control, monitoring, escalation, and residual-risk specifications**

Risk factor	Response actions: prevention; mitigation; contingency	KRI	Trigger threshold	Escalation rule	Residual-risk review
R08 Political and security instability	P: Political-security horizon scan; route, workforce, and critical-supply continuity planning. M: Alternative logistics and sourcing; work resequencing; insurance and timely relief notices. C: Incident command, business-continuity activation, safe suspension, and controlled restart.	Security-related lost workdays or critical-supply interruption	Any site closure, >=3 cumulative lost workdays, or critical-supply interruption >7 days.	Immediate notification to the SRO/crisis committee; contractual notice within 1 business day; recovery plan within 48 hours.	Re-rate after every event and monthly while elevated; restart only within the approved project tolerance.
R01 Cost escalation and overrun	P: Reference-class estimate; inflation/currency scenarios; quantified contingency and cost baseline. M: Value engineering, hedging, procurement packaging, and strict change control. C: Release approved contingency, rebaseline, and resolve eligible compensation events.	Estimate-at-completion variance and contingency burn	Forecast cost >5% above approved budget, or >50% of contingency used before 50% physical progress.	Commercial change board within 5 business days; escalate to SRO and lenders when tolerance or covenant headroom is exceeded.	Monthly and at every stage gate; re-rate after each approved material change.
R03					

Risk factor	Response actions: prevention; mitigation; contingency	KRI	Trigger threshold	Escalation rule	Residual-risk review
Coordination and communication failure	P: Integrated RACI/interface matrix, common data environment, decision rights, and meeting cadence. M: Cross-party interface workshops, single issue log, and time-bound action closure. C: Executive interface board, dispute-avoidance referral, and coordinated recovery plan.	Critical decisions or interfaces overdue	>=3 critical items >10 business days overdue, or one unresolved dependency blocks a milestone.	Interface board within 2 business days; steering committee if unresolved after 5 business days.	Fortnightly during L2-L4 and monthly during L5-L6; re-rate after material interface failures.
R07 Planning, scope, and schedule deficiencies	P: Stage-gate scope freeze, integrated master schedule, dependency logic, and schedule-risk analysis. M: Critical-path recovery, resequencing, resource leveling, and controlled scope decisions. C: Approved rebaseline, targeted acceleration, and deployment of schedule contingency.	Schedule performance index and critical-path slippage	SPI <0.90 for two reporting periods, or critical-path slippage >30 calendar days.	Recovery board within 5 business days; SRO if forecast completion exceeds contractual tolerance.	Monthly and at stage gates; re-rate after every approved rebaseline.
R04 Capability and management-system weakness	P: Competency matrix, critical-role coverage plan, delegated authority, and minimum key-person criteria. M: Targeted recruitment, training, coaching, and temporary expert augmentation. C: Replace or supplement underperforming teams and appoint independent project-management support.	Qualified critical-role coverage and critical vacancy duration	<90% of critical roles filled, or any critical vacancy remains open >30 calendar days.	PPP PMO capability review; sponsor-approved replacement or augmentation plan within 10 business days.	Monthly role-coverage review and re-rating at each lifecycle handoff.
R05 Resource, equipment, and maintenance constraints	P: Lifecycle asset plan, critical-spares policy, supplier due diligence, and preventive-maintenance programme. M: Alternative suppliers, expedited logistics, condition-based maintenance, and backlog recovery. C: Emergency spares, temporary equipment, service-continuity mode, and recovery mobilization.	Critical-spare cover and overdue preventive-maintenance backlog	Critical-spare cover <30 days, or >10% of scheduled preventive maintenance overdue.	SPV COO within 2 business days; authority contract manager when service KPIs are threatened.	Monthly, after major breakdowns, and before handback condition reviews.
R11 Design and technical deficiencies	P: Independent design reviews, constructability checks, BIM/interface coordination, and design-freeze gates. M: Time-bound design correction, technical-query closure, and controlled design change. C: Independent redesign/remediation, hold-point activation, and affected-work recovery.	Major design non-conformances and late design approvals	>=3 major design NCRs open >30 days, or a critical design package >20 business days late.	Design assurance board; notify independent engineer and authority technical lead within 5 business days.	At every design gate and monthly during construction; re-rate after major NCR closure.
R10 Approval and permitting delays	P: Integrated permit register, early authority engagement, submission-quality gate, and dependency calendar. M: Rapid query closure, dedicated interagency working group, and work resequencing.	Critical permit milestone slippage	A critical permit is >20 business days overdue, or threatens the critical path by >10 days.	Permits board within 5 business days; SRO if unresolved after 10 business days.	Fortnightly until critical approvals close and at each affected stage gate.

Risk factor	Response actions: prevention; mitigation; contingency	KRI	Trigger threshold	Escalation rule	Residual-risk review
	C: Interagency escalation, alternate approval route where lawful, and contractual relief/recovery plan.				
R02 Contractual and risk-allocation issues	P: Bankable allocation matrix, market sounding, lender review, and clause-to-risk crosswalk. M: Joint risk workshops, clarification log, early-warning notices, and controlled contract changes. C: Dispute-avoidance board, mediation/adjudication route, and continuity protections during resolution.	Unresolved high-severity allocation or claim issues	Any high-severity allocation issue remains open at bid/financial-close gate, or >20 business days during delivery.	Joint commercial board; then contractually defined dispute-avoidance mechanism and SRO.	At every gate and monthly while a material claim or ambiguity remains open.
R06 Land, site, and environmental constraints	P: Title/access due diligence, surveys, EIA and climate screening, permit plan, and stakeholder engagement. M: Remediation, resettlement/stakeholder actions, design adaptation, and compliance-action closure. C: Alternate access/design, stage-gate hold, emergency environmental response, and contractual relief where eligible.	Unsecured parcel/permit or overdue high-severity environmental action	Any critical parcel/permit is not secured by its stage gate, or a high-severity action is >30 days overdue.	Land/environment taskforce within 5 business days; gate hold or relief notice when the milestone is endangered.	Monthly and at land/permit/design/hand-back gates; re-rate after material environmental events.
R09 Demand and market risk	P: Independent demand validation, downside scenarios, willingness-to-pay testing, and social-engagement plan. M: Service/pricing adjustments, demand-management measures, targeted engagement, and cost alignment. C: Activate contracted revenue-sharing/support mechanism or restructuring review where trigger conditions are met.	Actual demand or revenue variance from the approved base case	>10% below the approved base for two consecutive reporting periods, or any financing-covenant breach.	Commercial review; notify authority and lenders when a support or covenant trigger is reached.	Quarterly during operation and after tariff, service, or demand shocks.
R15 Corruption and unethical practices	P: Integrity due diligence, conflict declarations, segregation of duties, e-procurement, and whistleblowing channels. M: Independent review, enhanced approvals, control remediation, and supplier re-screening. C: Suspend the affected process/party, preserve evidence, investigate, report as required, and activate step-in/termination rights.	Substantiated integrity breach or unresolved high-risk due-diligence red flag	Any substantiated breach, or a high-risk red flag unresolved for >10 business days.	Immediate audit/ethics committee escalation and notification to competent authorities when legally required.	Continuous event monitoring, quarterly assurance, and re-rating after every substantiated case.
R14 Tax and tariff changes	P: Tax/tariff baseline, scenario testing, contractual change-in-law definition, and compliance calendar. M: Permitted pass-through, hedging, cost efficiency, and financing/covenant management. C: Issue contractual notice, activate relief/renegotiation route, and reassess affordability and bankability.	Unmitigated tax/tariff change effect on annual cost or revenue	Modeled effect >=2% of annual operating cost/revenue, or any financing-covenant breach.	Finance/commercial board within 10 business days and change-in-law notice within the contractual period.	Quarterly and after every enacted or probable material change.

Risk factor	Response actions: prevention; mitigation; contingency	KRI	Trigger threshold	Escalation rule	Residual-risk review
R12 Government intervention and commitment risk	P: Commitment register, governance charter, approval calendar, payment-security arrangements, and succession protocol. M: Continuity/change-control protocol, dependency tracking, and accelerated payment/approval resolution. C: Steering escalation, sponsor/treasury intervention, and contractual compensation-event process where applicable.	Overdue public approval, payment, guarantee, or committed dependency	Payment/approval >30 calendar days late, or a critical commitment is missed by its dependency date.	Steering committee within 5 business days; sponsor/treasury escalation under the governance charter.	Monthly and after leadership, policy, or fiscal events; re-rate at every handoff.

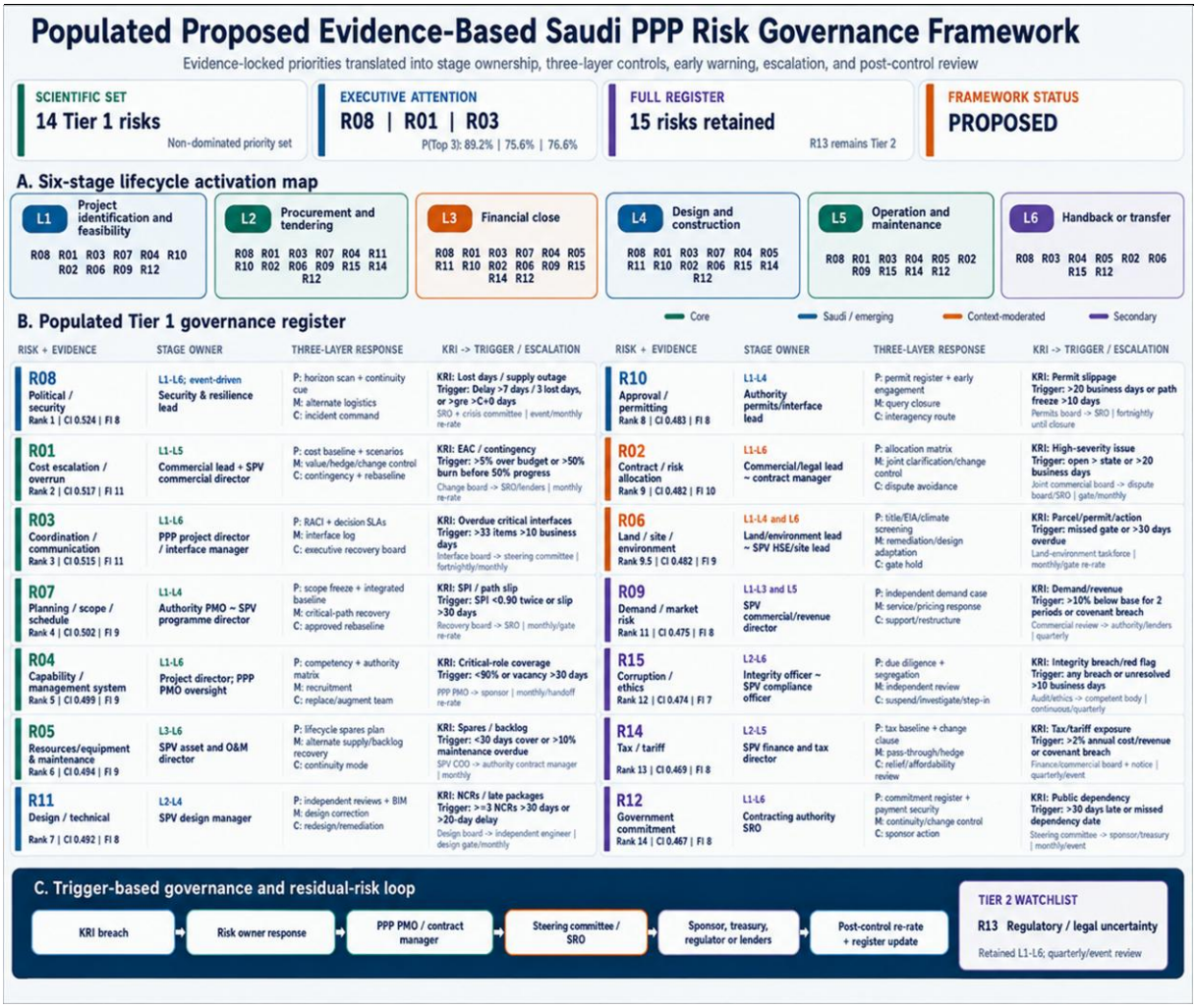


Fig. 17 Proposed evidence-based Saudi PPP risk governance framework. Populated Proposed Evidence-Based Saudi PPP Risk Governance Framework. The six-stage lifecycle ribbon identifies the active stages for the 14 non-dominated Tier 1 factors; R13 is retained in the full register as a Tier 2 watchlist factor. Risk cards report evidence-locked rank and CI values alongside proposed stage owners, three-layer responses, KRIs, trigger thresholds, escalation routes, and residual-risk review. The lower band shows the common trigger-based escalation and post-control reassessment loop. Governance fields and numerical triggers are proposed, project-adaptable specifications; residual risk is to be reassessed using project-specific post-control data.

4. Discussion

4.1. Principal Findings and Contribution

This study connected four analytical tasks that are frequently treated separately in PPP risk research: evidence normalization, context-specific stakeholder assessment, uncertainty-aware prioritization, and operational governance translation. The corrected evidence architecture began with an auditable 43-source corpus, distinguished the 15-source item-level archive from the 14-source recurrence denominator, and retained the final disposition of all

519 extracted statements. This separation is important because repeated expressions within one source should not be interpreted as independent evidence of recurrence. The approach therefore responds to the terminological heterogeneity and inconsistent units of analysis identified in previous PPP risk reviews [1–4].

The field assessment further demonstrated that identifying a frequently reported risk is not equivalent to establishing its importance in a particular market. The 205-participant Saudi stakeholder sample produced locally grounded probability and impact assessments, while respondent-level pairing preserved the dependence between the two ratings. Bootstrap rank distributions, dominance probabilities, subgroup inference, sensitivity scenarios, and qualitative findings were then used to characterize what the ranking could and could not support. This represents a shift from a static risk list toward a traceable evidence-to-governance process, consistent with calls for integrated risk management, continuing ownership, and lifecycle-based allocation in PPP projects [9–12].

4.2. Interpretation of the Leading Saudi PPP Risks

Political and security instability (R08) emerged as the highest-ranked risk, with CI = 0.5243, P(Rank = 1) = 51.7%, and P(Top 3) = 89.2%. This result indicates that Saudi stakeholders assigned substantial combined probability and impact to risks arising from political change, geopolitical exposure, security disruption, and related continuity concerns. The finding is consistent with research showing that political uncertainty can affect contractual confidence, investment conditions, and the long-term performance of PPP arrangements [13]. It also complements earlier Saudi evidence identifying institutional capability and ineffective risk-management implementation as material concerns in independent water and power projects [14,15].

Cost escalation and overrun (R01) ranked second and had the highest mean probability, while coordination and communication failure (R03) ranked third and had the highest mean impact. Their positions are substantively plausible because PPP projects combine long contractual periods, multiple public and private organizations, complex financing arrangements, and interdependent design, construction, and operational decisions. Financial-risk reviews have similarly emphasized the continuing importance of cost, financing, revenue, and payment exposures [2,16]. Integrated risk and operational-management studies also show that weak coordination can transmit local problems across organizational and lifecycle boundaries [3,8,9]. Accordingly, the leading risks should not be interpreted as isolated events: political or security disruption may affect financing and supply conditions, while inadequate coordination may delay responses and intensify cost consequences.

Capability and management-system weakness (R04) was among the most recurrent literature factors and remained a Core risk in the evidence-alignment classification. This supports the view that PPP performance depends not only on contractual allocation but also on the capacity of participating organizations to interpret obligations, exchange information, manage interfaces, and implement controls. Saudi construction research has likewise highlighted management capability and broader institutional barriers as important determinants of sustainable project performance [30,31]. The present findings therefore indicate that capability-building should be treated as a risk-control requirement rather than as a general organizational aspiration.

Regulatory and legal uncertainty (R13), by contrast, had the lowest stakeholder criticality despite relatively high literature recurrence. This result should not be interpreted as evidence that regulatory risk is absent. It may instead indicate that respondents perceived this risk as more familiar, formally allocated, or manageable than political, cost, coordination, and capability exposures. Risk taxonomies and allocation studies consistently recognize legal and regulatory conditions as fundamental PPP concerns [1,4,12]. The divergence observed here illustrates why literature frequency should remain analytically distinct from current market assessment. R13 was therefore retained in the complete governance register and Tier 2 watchlist rather than being removed from monitoring.

4.3. Ranking Uncertainty and the Meaning of the Priority Tiers

Although the Friedman test detected statistically significant differences among the 15 paired criticality scores, Kendall's W = 0.0157 indicated very limited overall separation. Only three of the 105 Wilcoxon comparisons remained significant after Holm correction, and each involved one of the three leading risks exceeding R13. These results show that statistical significance at the global level does not imply a sharply ordered hierarchy across all risks. The high response-consistency coefficients for probability and impact indicate consistent rating behavior, but they do not establish that the 15 risks represent a single latent construct or that their criticality values are widely separated.

The bootstrap results provide a more informative interpretation. R08 had the greatest probability of occupying first place, but its 95% rank interval extended from 1 to 5. R01 and R03 also had substantial Top-3 probabilities, while confidence intervals for several factors overlapped. Consequently, the reported rank order should be understood as an evidence-based summary of central tendency rather than as a deterministic sequence that will reproduce exactly in every stakeholder sample or project.

The assignment of 14 risks to the non-dominated Tier 1 set deserves particular attention. It does not mean that all 14 risks are equally critical or require identical resources. Instead, it means that the prespecified evidence did not establish stable directed superiority for most risk pairs under both the dominance and multiplicity-adjusted comparison rules. The large Tier 1 set is therefore an honest representation of limited statistical separation rather than a weakness to be corrected by imposing an arbitrary cut-off. The executive Top 3 and the scientific priority tiers serve different purposes: R08, R01, and R03 provide a concise management-attention set, whereas Tier 1 preserves

all risks that could not be shown to be dominated robustly. R13 remains monitored in Tier 2 and can still be escalated when its project-specific trigger is breached.

4.4. Literature–Market Alignment and Stakeholder Heterogeneity

The relationship between literature recurrence and Saudi market criticality was positive but not statistically significant under the prespecified label-permutation test ($\rho s = 0.400$; $p = 0.1403$). This result suggests partial convergence rather than either complete agreement or complete disconnection. Cost escalation, coordination failure, capability weakness, and resource-related risks were prominent in both evidence layers, whereas political and security instability was more prominent in the Saudi stakeholder assessment than in the recurrence ranking. Conversely, regulatory and legal uncertainty received stronger support from the literature dimension than from current stakeholder ratings.

The participant-bootstrap interval and permutation result address different questions and should not be treated as contradictory. Participant-vector resampling conditions on the fixed literature scores and evaluates uncertainty associated with the stakeholder sample. The permutation test instead evaluates whether the observed matching of factor labels across only 15 risks is stronger than expected under random label assignment. The primary inference regarding literature–market concordance therefore remains the non-significant permutation result. The positive bootstrap distribution indicates that the observed direction was reasonably stable to participant resampling, but it does not establish general concordance across the broader universe of PPP risk factors.

This distinction supports the decision not to combine literature and market percentiles into a new primary composite score. Because the questionnaire factors were already derived from the literature base, adding literature recurrence again to market criticality could double count the same evidence. The quadrant classification instead retains both dimensions and makes their convergence or divergence visible. This approach is consistent with the broader literature showing that PPP risk salience varies by sector, delivery arrangement, national setting, and stakeholder position [5–8].

Subgroup findings further demonstrated that stakeholder experience influenced the level of assigned criticality. Participants with direct PPP experience reported significantly higher criticality for six risks, while years of experience differentiated ten risks and PPP knowledge differentiated all 15. These effects may reflect greater recognition of interdependencies, contractual exposure, or implementation difficulty among more experienced participants. However, PPP-infrastructure experience alone produced no FDR-significant differences. The results therefore do not support replacing the heterogeneous full sample with a narrowly defined subgroup. Instead, they indicate that experience should be retained as an interpretive dimension when applying the framework.

The sensitivity analyses reinforce this conclusion. Excluding the 21 straightliners left the complete ranking unchanged, and the primary order remained strongly correlated with rankings obtained from the direct-PPP, combined-experience, highest-confidence, and alternative-scoring scenarios. The Top 3 was also preserved in all scenarios except the smallest highest-confidence cohort, in which R05 replaced R03. This isolated change is consistent with increased sampling uncertainty at $n = 39$ rather than evidence that the full-sample ranking is unreliable. The stability of Tier 1 membership across the scenarios provides further support for the distinction between a broad scientific priority set and a more selective executive-attention set.

4.5. Qualitative Findings and Governance Translation

The open-ended responses added contextual detail without changing the quantitative hierarchy. Leadership, workforce competence, and decision authority was the most prevalent theme, followed by institutional commitment and administrative or contracting continuity. These themes help explain why coordination, capability, and management-system risks remained prominent. They also extend the interpretation from identifying risk conditions to considering whether organizations possess the authority, knowledge, continuity, and response capacity required to manage them.

Three qualitative themes—quality and safety assurance, technology, cybersecurity and obsolescence, and institutional memory and interagency knowledge continuity—were classified as emergent candidates. Their identification is consistent with the growing importance of data-driven monitoring, digital procurement, process mining, BIM-supported operation, and dynamic safety systems in construction and infrastructure management [32,34,39–42]. However, these themes were not inserted into the quantitative ranking because they were derived from only 27 substantive open-ended responses and were not assessed through the closed probability–impact instrument. They are more appropriately retained as watchlist items for future measurement and project-specific review.

The Proposed Evidence-Based Saudi PPP Risk Governance Framework translates the combined findings into six lifecycle stages and a common governance logic. Its principal contribution is not another numerical score, but the connection of analytical evidence to ownership, contractual allocation, prevention, mitigation, contingency planning, KRIs, trigger thresholds, escalation, and residual-risk reassessment. This is consistent with lifecycle-oriented PPP risk planning and integrated management principles [9–12,43]. The framework also preserves analytical and governance fields separately: CI values, ranks, uncertainty, tiers, and evidence classes are imported without reweighting, while owners, controls, thresholds, and review periods remain explicit project-adaptable specifications. The framework should therefore be used as a structured decision support and governance-register template rather than as a universal prescriptive model. Its proposed thresholds must be calibrated to project sector, contractual

structure, lifecycle stage, organizational risk appetite, and available control data. A risk outside the executive Top 3 may still require immediate escalation if its KRI breaches the approved project threshold. Conversely, a highly ranked risk should not automatically receive additional resources when existing controls have demonstrably reduced its residual exposure.

4.6. Strengths and Limitations

Several methodological features strengthen the credibility of the findings. First, the study distinguishes the full evidence corpus, the item-level archive, the recurrence-eligible subset, the normalized register, and the questionnaire shortlist. Second, every extracted statement retains a final disposition and traceable source relationship. Third, the primary criticality index preserves respondent-level probability–impact pairing instead of relying only on a product of aggregate means. Fourth, rank uncertainty is reported through participant-vector bootstrap resampling and dominance probabilities. Fifth, literature and market evidence remain separate, avoiding double counting. Finally, subgroup, response-quality, sample, and scoring sensitivity analyses were conducted without replacing the prespecified primary analysis.

The study nevertheless has important limitations. The stakeholder survey used non-probability purposive recruitment, and the number of eligible invitees reached was unavailable; consequently, a conventional response rate could not be calculated. Direct PPP participation was reported by 38.0% of respondents, and the highest-confidence subgroup contained only 39 participants. The results therefore represent the assessed Saudi stakeholder sample and should not be generalized mechanically to every Saudi PPP sector, organization, or project. The cross-sectional self-reported design also captures perceived probability and impact rather than observed project loss data.

4.7. Practical and Research Implications

For Saudi public authorities and private partners, the results support a two-level management strategy. R08, R01, and R03 form an executive-attention set requiring visible senior oversight, continuity planning, cost control, and interface management. The wider Tier 1 set should remain subject to structured ownership and monitoring because the evidence did not justify excluding most factors from enhanced attention. R13 and qualitative watchlist risks should remain in the master register and be escalated whenever project-specific indicators show material change.

The results also indicate that PPP governance should emphasize organizational interfaces rather than treating each risk as the responsibility of one isolated department. Political and security continuity, financing and payment safeguards, coordination authority, capability development, information exchange, and technology resilience require linked actions across public authorities, project companies, financiers, contractors, operators, and regulators. Contractual allocation should consequently be supported by measurable KRIs and escalation rules rather than being treated as a one-time transfer of responsibility.

Future research should update the evidence corpus through a prospectively documented search and independent coding process, administer the questionnaire to larger sector-stratified samples, and compare transport, water, energy, housing, health, and education PPPs. Longitudinal project data should be used to examine whether changes in KRIs predict delays, cost escalation, disputes, service interruptions, or residual-risk movement. Repeated stakeholder assessments could also determine whether the observed literature–market divergences persist as the Saudi PPP market develops. Most importantly, the proposed governance register should be applied prospectively to real projects, with ownership actions, trigger breaches, escalations, control completion, and residual-risk outcomes recorded over time. Such evidence would determine whether the framework improves the timeliness, transparency, and consistency of PPP risk governance.

5. Conclusions

This study developed an uncertainty-aware, evidence-based approach for prioritizing and governing public-private partnership risks in Saudi Arabia. By linking a traceable literature-derived risk archive with stakeholder probability-impact assessment, bootstrap rank uncertainty, dominance-based tiering, subgroup and sensitivity testing, and qualitative contextualization, the study moved beyond a simple ranked list toward an operational governance logic. The results identified political and security instability, cost escalation and overrun, and coordination and communication failure as the executive Top 3 risks, while the dominance analysis showed that most risks should remain under enhanced Tier 1 monitoring because statistical separation between many factors was limited. Regulatory and legal uncertainty was retained as a Tier 2 watchlist risk rather than excluded from the governance register. The proposed framework translates these findings into lifecycle ownership, contractual allocation, response controls, KRIs, trigger thresholds, escalation routes, and residual-risk review. Its value lies in providing Saudi PPP decision-makers with a transparent and adaptable structure for connecting empirical risk evidence to practical governance action. The framework remains proposed and should be refined through future project-level implementation, sector-specific calibration, longitudinal KRI data, and independent external assessment across different Saudi PPP sectors.

Declarations

Availability of data and materials

The analytical datasets and executable scripts supporting the reported results are provided in Supplementary Data S1 and S3–S10 and Supplementary Code S12. The exact questionnaire is provided in Supplementary File S2.

Competing interests

The authors declare no competing interests.

Funding

This research received no external funding.

Authors' contributions

Conceptualization, M.H.A. and F.A.S.; methodology, M.H.A., F.A.S. and M.O.R.; software, M.H.A.; validation, M.H.A., F.A.S., M.O.R. and D.M.M.; formal analysis, M.H.A.; investigation, M.H.A.; resources, M.H.A., F.A.S., M.O.R. and D.M.M.; data curation, M.H.A.; writing – original draft preparation, M.H.A.; writing – review and editing, F.A.S., M.O.R. and D.M.M.; visualization, M.H.A.; supervision, F.A.S., M.O.R. and D.M.M.; project administration, M.H.A. All authors have read and agreed to the final version of the manuscript.

Ethics approval

Not applicable.

Consent to participate

Participants were informed that the data would be used for academic purposes; however, the preserved materials did not contain a complete participant-information and informed-consent record.

Acknowledgements

AI-assisted support during manuscript preparation is described in Section 2.9.

Supplementary information

The following supporting information accompanies this article: Figure S1: archived 43-source evidence corpus, temporal-eligibility rule, and analytical-role flow; Supplementary File S1 and Supplementary Data S1: literature-evidence audit, complete 43-source corpus and final roles, temporal eligibility and exception log, 519-record cross-walk, focused 134-record adjudication log, separate legacy working-record log, 41-factor recurrence register, questionnaire-boundary decision, and source-flow traceability record; Supplementary File S2: exact questionnaire, codebook, and survey-administration and participant-information evidence audit; Supplementary Data S3: screening, cohorts, ratings, and administration metadata; Supplementary Data S4: scoring, consistency, Friedman, average-rank, and explicit Wilcoxon-implementation diagnostics; Supplementary Table S5: all 105 Wilcoxon contrasts and the three Holm-significant contrasts; Supplementary Data S6: bootstrap, dominance, pairwise-decision, and tier outputs; Supplementary Data S7: literature–market concordance, factor coordinates, alignment classes, permutation and participant-bootstrap distributions, and IPS sensitivity; Supplementary Data S8–S10: subgroup inference, robustness, and qualitative audit; Supplementary Table S11: full proposed Saudi PPP risk-governance register; Supplementary Code S12: version-controlled reproducibility package, verification scripts, reference outputs, and SHA-256 manifest.

References

1. Rasheed N, Shahzad W, Khalfan M, Rotimi JOB (2022) Risk identification, assessment, and allocation in PPP projects: A systematic review. Buildings 12:1109. https://doi.org/10.3390/buildings12081109
2. Akomea-Frimpong I, Jin X, Osei-Kyei R (2021) A holistic review of research studies on financial risk management in public–private partnership projects. Eng Constr Archit Manag 28:2549–2569. https://doi.org/10.1108/ECAM-02-2020-0103
3. Jiang W, Yang Q, Jiang J, Martek I, Gao F (2022) Operational risk management of public–private partnership infrastructure projects: A bibliometric literature review. Buildings 12:1905. https://doi.org/10.3390/buildings12111905
4. Le PT, Kirytopoulos K, Chileshe N, Rameezdeen R (2022) Taxonomy of risks in PPP transportation projects: A systematic literature review. Int J Constr Manag 22:166–181. https://doi.org/10.1080/15623599.2019.1615756
5. Ameyaw EE, Chan APC (2013) Identifying public–private partnership risks in managing water supply projects in Ghana. J Facil Manag 11:152–182. https://doi.org/10.1108/14725961311314651
6. Akhtar M, Mufti NA, Mubin S, Saleem MQ, Zahoor S, Ullah S (2023) Identification of various execution modes and their respective risks for public–private partnership infrastructure projects. Buildings 13:1889. https://doi.org/10.3390/buildings13081889
7. Li Z, Wang H (2023) Exploring risk factors affecting sustainable outcomes of global public–private partnership projects: A stakeholder perspective. Buildings 13:2140. https://doi.org/10.3390/buildings13092140
8. Liu B, Li J, Wang D, Liu H, Wu G, Yuan J (2024) Public–private partnerships: A collaborative framework for ensuring project sustainable operations. Eng Constr Archit Manag 31:264–289. https://doi.org/10.1108/ECAM-12-2021-1124
9. Fischer K, Leidel K, Riemann A, Alfen HW (2010) An integrated risk management system for PPP projects. J Financ Manag Prop Constr 15:260–282. https://doi.org/10.1108/13664381011087515
10. Fatima A, Mubin S, Masood R (2024) Risk-based integrated performance assessment framework for public–private partnership infrastructure projects. Organ Technol Manag Constr 16:251–274. https://doi.org/10.2478/otmcj-2024-0018
11. Mazher KM (2019) Risk assessment and allocation model for public–private partnership infrastructure projects in Pakistan. PhD thesis, The Hong Kong Polytechnic University, Hong Kong. https://theses.lib.polyu.edu.hk/handle/200/10035. Accessed 17 Aug 2026
12. Xiang P, Zhang Q, Jiang Q, Liu Z (2022) Operational risk allocation in urban rail transit public–private partnership projects. Front Environ Sci 10:900322. https://doi.org/10.3389/fenvs.2022.900322
13. Najeeb M (2024) Political uncertainty in public–private partnership projects: a flexible risk mitigation framework for Jordan's challenging landscape. Doctoral dissertation, University of Bedfordshire, Luton. http://hdl.handle.net/10547/626392. Accessed 17 Aug 2026

14. Alsulaiman Y, Bowles G, Ogunlana S (2014) Evaluating risk management in independent water and power plant projects in Saudi Arabia. In: Proceedings of the 30th Annual ARCOM Conference, Portsmouth, 1–3 September 2014, pp 1409–1417
15. Alsulaiman Y, Bowles G, Ogunlana S (2016) A study of the elements that lead to an ineffectiveness of risk management implementation in independent water and power plant projects in Saudi Arabia. In: Alford N, Fréchet J (eds) Proceedings of the Eighth Saudi Students Conference in the UK. Imperial College Press, London. https://doi.org/10.1142/9781783269150_0059
16. Akomea-Frimpong I, Jin X, Osei-Kyei R, Pariafsai F (2023) Critical managerial measures on financial risks of sustainable public–private partnership projects: A PRISMA review. J Financ Manag Prop Constr 28:398–422. https://doi.org/10.1108/JFMPC-12-2021-0070
17. Awuah E (2023) Enhanced key success indicators of a risk management framework for public–private partnerships smart city projects. J Contemp Manag 20:89–122. https://doi.org/10.35683/jcman1068.216
18. Bakhtawar B, Thaheem MJ, Arshad H, Tariq S, Mazher KM, Zayed T, Akhtar N (2022) A sustainability-based risk assessment for P3 projects using a simulation approach. Sustainability 14:344. https://doi.org/10.3390/su14010344
19. Kuru K, Artan D (2024) Riesgo: A knowledge-based qualitative risk assessment system for PPP projects. Buildings 14:953. https://doi.org/10.3390/buildings14040953
20. Fathi M (2024) A structural equation model on critical risk and success in public–private partnership: Exploratory study. J Risk Financ Manag 17:354. https://doi.org/10.3390/jrfm17080354
21. Alqahtani FK, Al-Jadhai S, Alromihy H, Alsaud M, Sherif M, Mohamed AG (2024) Appraising critical success factors in sustainable housing projects: A comparative study of PPP modalities in Saudi Arabia. Heliyon 10:e32854. https://doi.org/10.1016/j.heliyon.2024.e32854
22. Biygautane M, Clegg S (2024) Public–private partnerships for school infrastructure development: International review and recommendations for Saudi Arabia. J Infrastruct Policy Dev 8:3672. https://doi.org/10.24294/jipd.v8i6.3672
23. Bin Jubair SM, Singh JSK (2022) Critical success factors of public–private partnership projects in the Kingdom of Saudi Arabia. Webology 19:1521–1540. https://doi.org/10.14704/web/v19i1/web19102
24. Okoro C, Nnaji C, Chileshe N, Tembo J (2023) Perceptions of public–private partnerships transportation project success factors in developing countries: An explanatory sequential investigation. J Constr Dev Ctries 28:295–327. https://doi.org/10.21315/JCDC-07-22-0123
25. Sanda YN, Anigbogu NA, Molwus JJ (2016) Managing risks in public–private partnerships in housing in Nigeria: Methodological perspective. J Sustain Dev 9:152. https://doi.org/10.5539/jsd.v9n5p152
26. Lima SMCA (2021) PPP: a risk approach in the water sector. PhD thesis, Iscte–Instituto Universitário de Lisboa, Lisbon. https://hdl.handle.net/10071/22688. Accessed 17 Aug 2026
27. Fong PSW (ed) (2025) Public–private partnerships (PPPs) for construction project deliveries. MDPI, Basel. ISBN 978-3-7258-4300-8. https://doi.org/10.3390/books978-3-7258-4300-8
28. Onour IA (2012) Implementing privatization strategy of Saudi Arabia: issues and challenges. SSRN Electron J. https://doi.org/10.2139/ssrn.2189804
29. Aladayleh KJ, Aladaileh MJ (2024) Applying analytical hierarchy process to BIM-based risk management for optimal performance in construction projects. Buildings 14:3632. https://doi.org/10.3390/buildings14113632
30. Alghuried A (2026) Assessing the critical success factors for the sustainable construction project management of Saudi Arabia. J Asian Archit Build Eng 25:598–616. https://doi.org/10.1080/13467581.2025.2454616. First published online 30 January 2025
31. Al-Otaibi A, Bowan PA, Alabdullatief A, Albaiz M, Salah M (2025) Barriers to sustainable building project performance in developing countries: A case of Ghana and the Kingdom of Saudi Arabia. Sustainability 17:3539. https://doi.org/10.3390/su17083539
32. Cao Y, Kamaruzzaman SN, Aziz NM (2022) Building information modeling capabilities in the operation and maintenance phase of green buildings: A systematic review. Buildings 12:830. https://doi.org/10.3390/buildings12060830
33. Fredson G, Adebisi B, Ayorinde OB, Onukwulu EC, Adediwin O (2023) Strategic risk management in high-value contracting for the energy sector: Industry best practices and approaches for long-term success. Int J Manag Organ Res 2:16–30. https://doi.org/10.54660/ijmor.2023.2.1.16-30
34. Khodabakhshian A, Re Cecconi F (2022) Data-driven process mining framework for risk management in construction projects. IOP Conf Ser Earth Environ Sci 1101:032023. https://doi.org/10.1088/1755-1315/1101/3/032023
35. Nawaz A, Su X, Din QMU, Khalid MI, Bilal M, Shah SAR (2020) Identification of health and safety factors involved in infrastructure projects in developing countries: A sequential mixed-method approach. Int J Environ Res Public Health 17:635. https://doi.org/10.3390/ijerph17020635
36. Nsefu MK, Thole P (2024) Development of risk management framework for financing of national road projects by the local banks. Eur J Nat Soc Sci-Novus 2:01000155.
37. Oyetunji TS, Erinjogunola FL, Ajirotutu RO, Adeyemi AB, Ohakawa TC, Adio SA (2025) A unified risk management framework for cost and resource optimization in housing development projects. Gulf J Adv Bus Res 3:985–997. https://doi.org/10.51594/gjabr.v3i4.130
38. Salzano A, Cascone S, Zitiello EP, Nicolella M (2024) Construction safety and efficiency: Integrating building information modeling into risk management and project execution. Sustainability 16:4094. https://doi.org/10.3390/su16104094
39. Siu MFF, Leung WYJ, Chan WMD (2018) A data-driven approach to identify, quantify, and analyse construction risk for Hong Kong NEC projects. J Civ Eng Manag 24:592–606. https://doi.org/10.3846/jcem.2018.6483
40. Sultan B, Alhammad I, AlOthman AA, AlSehli G (2025) Bridging the literature gap on eProcurement systems: Insights from Saudi Arabia's sustainable development transition. Sustainability 17:3429. https://doi.org/10.3390/su17083429

41. Wang J, Chen Z, Song Y, Liu Y, He J, Ma S (2024) Data-driven dynamic Bayesian network model for safety resilience evaluation of prefabricated building construction. Buildings 14:570. https://doi.org/10.3390/buildings14030570
42. Waqar A, Nisar S, Qadir S (2025) BIM-driven sustainable safety management solutions for downstream oil and gas construction projects. Environ Chall 18:101058. https://doi.org/10.1016/j.envc.2024.101058
43. Yaworsky RA (1994) Risk planning for large and BOT projects: a holistic framework. PhD thesis, University of British Columbia, Vancouver. https://library-archives.canada.ca/eng/services/services-libraries/theses/Pages/item.aspx?id-Number=32545955. Accessed 18 Aug 2026
44. Khallaf R, Guevara J, Mendez-Gonzalez P, Castelblanco G (2024) A system dynamics model for a national PPP program: the Egyptian project portfolio. In: Construction Research Congress 2024. https://doi.org/10.1061/9780784485286.051

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SubMaterials.rar