

Threat-Informed Vulnerability Prioritisation: Temporal Validation and a Formal Design for Organisational Context

Emad Sherif^{1*}

¹Faculty of Technology, Arts and Culture, De Montfort University, Leicester, UK

*Correspondence: emad.o.sherif@dmu.ac.uk

PAPER INFO

Paper history:

Received 24 April 2026

Accepted 22 September 2026

Citation:

Sherif, E. (2026). Threat-Informed Vulnerability Prioritisation: Temporal Validation and a Formal Design for Organisational Context. In *Journal of Information and Organizational Sciences*, vol. 50, no. 2, pp. 293-308

Copyright:

© 2026 The Authors. This work is licensed under a Creative Commons Attribution BY-NC-ND 4.0. For more information, see <https://creativecommons.org/licenses/by-nc-nd/4.0/>

ABSTRACT

Vulnerability prioritisation remains a critical cyber risk-governance challenge for organisations with limited remediation resources. This study makes two contributions. First, it evaluates a threat-informed Key Risk Indicator combining exploitation probability, technical severity, and weakness-level exploitation prevalence. To prevent temporal leakage, weakness weights were estimated from 87,080 vulnerabilities published between 2020 and 2023 and fixed before evaluation on 59,638 vulnerabilities published between January 2024 and May 2025. For each evaluation CVE, KRI used the first archived EPSS score identified on or within two calendar days after publication. KRI achieved an area under the precision-recall curve of 0.053, compared with 0.014 for severity-only prioritisation, and captured 14.6% of CVSS-weighted KEV value within the top 500 vulnerabilities, compared with 6.6%. Second, the study introduces the Context-Adjusted Key Risk Indicator as a formal organisation-aware extension incorporating control implementation and asset context. Synthetic scenarios illustrate its computation but do not validate organisation-specific effectiveness, which requires real organisational measurements and an organisation-sensitive reference outcome. The findings support threat-informed prioritisation while establishing a design for future organisational validation.

Keywords: Cyber Risk Management, Key Control Indicators, Key Risk Indicators, MITRE ATT&CK, Vulnerability Prioritisation

1. Introduction

Cyber vulnerability management is one of the most resource-intensive functions in organisational security operations. Faced with thousands of newly disclosed vulnerabilities each year, security teams should decide, under constrained time and personnel budgets, which vulnerabilities to remediate first. The de facto standard for these decisions, the Common Vulnerability Scoring System (CVSS) (FIRST, 2005), was designed to measure technical severity in isolation from real-world exploitation context (Allodi & Massacci, 2014). A growing body of empirical evidence has demonstrated that CVSS severity is a poor predictor of whether a vulnerability will be exploited in practice (Jacobs et al., 2023).

To address these limitations, Sherif et al. (2026a) proposed a composite Key Risk Indicator (KRI) that combines exploitation likelihood from the Exploit Prediction Scoring System (EPSS) (FIRST, 2025), technical severity from CVSS, and weakness-level exploitation prevalence derived from the Common Weakness Enumeration (CWE) (MITRE, 2006). This work independently implements the KRI formulation and evaluates it using a leakage-controlled temporal holdout design, rather than treating the earlier reported performance as evidence for the current cohort.

This earlier study reported stronger discrimination for KRI than for CVSS-only prioritisation on a broad historical dataset (Sherif et al., 2026a). Though, this work re-evaluates the formulation on a later temporal holdout cohort. However, the KRI framework is organisation-agnostic, assigning identical risk scores to all organisations facing the same vulnerability regardless of whether the security controls are partially or fully implemented. This renders the framework insensitive to the organisational differences, security control maturity, asset exposure, and security investment that most directly determine residual risk in practice.

From an information systems perspective, vulnerability prioritisation is a decision support problem embedded within organisational risk governance. Existing metrics, including CVSS and threat-informed composite scores, treat the vulnerability as the unit of analysis while leaving the organisation invisible, producing identical rankings for a mature enterprise with fully implemented controls and a resource-constrained organisation facing the same disclosed vulnerability.

CA-KRI reframes prioritisation as an organisational information quality problem wherein the goal is not merely to identify which vulnerabilities are technically dangerous in the abstract, but to deliver context-specific risk intelligence designed to reflect each organisation's actual residual exposure, with that design objective subject to field validation as organisational KCI measurement infrastructure matures. This positioning connects the framework to established information systems constructs, decision support, information quality, and socio-technical risk governance, and distinguishes it from prior purely technical approaches to vulnerability scoring.

An equally important concern, largely unaddressed in prior work, is temporal relevance. Several empirical studies in vulnerability prioritisation draw on CVE datasets spanning multiple decades, thereby mixing current active threats with legacy vulnerabilities that are rarely encountered in today's environments (Alexopoulos et al., 2022; Jacobs et al., 2023). Cyber risk managers and security leaders are concerned with the vulnerabilities disclosed in the past five years, for which temporally aligned threat-intelligence inputs can be reconstructed (Jacobs et al., 2023).

A framework evaluated exclusively on this recent landscape provides a more operationally grounded picture of its practical value. This study therefore addresses two questions: whether the threat-informed KRI baseline improves KEV-oriented vulnerability prioritisation over CVSS on a contemporary temporal holdout cohort, and how organisational control coverage and asset-level exposure context can be incorporated into KRI as a proposed organisation-aware extension. The first question is examined empirically, whereas the second is addressed through formal specification and computational illustration.

This paper makes two distinct contributions. First, it provides an empirical evaluation of the threat-informed KRI baseline against CVSS-only prioritisation on a contemporary CVE cohort. This evaluation examines whether integrating exploitation likelihood, technical severity, and weakness prevalence provides a stronger basis for prioritising vulnerabilities associated with known exploitation evidence. Second, the paper introduces CA-KRI as a formal organisation-aware extension of KRI.

CA-KRI specifies how KCI-based control coverage and asset-level exposure context can be incorporated into the threat-informed score as a modifier intended to represent residual organisational exposure. The synthetic contextual scenarios presented in this study are used to illustrate the computational operation and ranking behaviour of the proposed formulation. They do not validate CA-KRI's organisation-specific effectiveness because the KEV outcome is organisation-independent and the contextual inputs are not derived from real organisational measurements. Validation of CA-KRI therefore requires future research using real KCI and asset data with an organisation-sensitive reference outcome.

The remainder of the paper is structured as follows. Section two surveys the related work. Section three proposes the CA-KRI framework. Section four presents the findings. Section five interprets the findings. Section six concludes the paper.

2. Related Work

2.1. Limitations of CVSS-Based Prioritisation

The CVSS was designed to provide a standardised measure of technical vulnerability severity, enabling consistent communication of risk across vendors, researchers, and practitioners (FIRST, 2005). However, a substantial and growing body of empirical evidence has demonstrated that CVSS severity scores correlate poorly with actual exploitation events in practice. Allodi & Massacci (2014) were among the first to systematically document this misalignment, finding that many high-severity vulnerabilities are never exploited in the wild, while a disproportionate share of actively exploited vulnerabilities carry medium or lower CVSS scores.

Subsequent studies have consistently reinforced this finding; Sabetta & Bezzi (2018) identified systematic gaps between CVSS rankings and real-world attacker behaviour in the context of open-source software vulnerabilities, Jacobs et al. (2021) demonstrated that exploitation probability is better predicted by threat intelligence signals than by severity ratings, and Kosciński et al. (2025) extended these findings to the contemporary vulnerability landscape, confirming that CVSS remains a weak discriminator between exploited and non-exploited vulnerabilities even as scoring methodologies have evolved. The fundamental limitation is structural where CVSS measures the technical characteristics of a vulnerability in isolation, without reference to whether exploitation tooling exists, whether the vulnerability is being actively targeted, or whether the affected system is reachable by an adversary.

From an organisational information systems perspective, the practical consequences of CVSS-based prioritisation are acute for resource-constrained organisations such as SMEs. When prioritisation is driven by CVSS severity, these organisations are systematically directed toward vulnerabilities that are technically rated as dangerous but may not be exploited, while genuinely active threats may be deprioritised. This represents a failure of risk information quality in the IS governance sense wherein the information provided to decision-makers is technically accurate in a narrow sense but lacks the contextual relevance and actionability that effective organisational risk management requires (Savola, 2013). The need for prioritisation frameworks that incorporate exploitation evidence, organisational control coverage, and asset context is therefore not merely a technical optimisation problem but a substantive information systems governance challenge with direct implications for organisational resilience.

2.2. Exploit Prediction and Composite Risk Metrics

The EPSS, developed and maintained by the Forum of Incident Response and Security Teams, addresses the core limitation of CVSS by directly estimating the probability that a given vulnerability will be exploited in the wild within 30 days of disclosure (FIRST, 2025). Rather than measuring technical severity, EPSS is trained on a continuously updated combination of threat intelligence feeds, proof-of-concept availability signals, and historical exploitation data, producing a dynamic, data-driven score that reflects the current attacker landscape rather than a static assessment of vulnerability characteristics (Jacobs et al., 2021). Empirical evaluations have consistently demonstrated that EPSS substantially outperforms CVSS as a predictor of real-world exploitation, offering security teams a materially more accurate basis for triage decisions (Jacobs et al., 2023).

Building on this foundation, Sherif et al. (2026a) proposed a composite KRI that combines EPSS, CVSS base scores, and CWE-based exploitation-prevalence weights. The preprint reported that the composite score outperformed CVSS-only prioritisation on its evaluation dataset and examined the contribution of its individual components through ablation analysis. These earlier findings provide the design basis for the KRI formulation, but they are not treated as empirical evidence for the present study. The current analysis independently estimates the CWE weights on an earlier temporal partition and evaluates the resulting KRI on subsequently published CVEs.

2.3. Security Control Frameworks

Organisations seeking to reduce cyber risk in a structured and measurable way increasingly rely on prescriptive security control frameworks that translate broad security objectives into concrete, implementable safeguards (El Moutaouakil et al., 2026). Among the most widely adopted is the CIS Critical Security Controls (CIS Controls), a prioritised, consensus-based framework comprising 18 control categories and 153 individual safeguards that collectively establish a defence-in-depth security posture (Centre for Internet Security, 2021). The CIS Controls are structured around three Implementation Groups (IGs) that reflect organisational maturity and resource availability where IG1 covers essential cyber hygiene applicable to all organisations, IG2 addresses foundational controls for organisations with moderate security resources, and IG3 encompasses advanced safeguards for organisations with mature security programmes and higher risk profiles (Centre for Internet Security, 2021). This tiered structure makes the framework particularly relevant to resource-constrained environments such as SMEs, where full control adoption is neither feasible nor necessary as a starting point.

While the CIS Controls define what should be implemented, a distinct challenge in cyber risk governance is measuring how well each safeguard has been implemented. To address this, the CIS Controls Assessment Specification (CAS) extends each of the 153 safeguards with a structured measurement specification (CIS, 2025a). This structure enables objective, repeatable, and auditable assessment of control implementation quality, transforming qualitative security posture descriptions into quantitative governance signals suitable

for integration into risk scoring frameworks. In this work, KCIs operationalise the CAS measurement specifications at the safeguard level, producing a score between 0 and 1.

2.4. Connecting Threat Intelligence to Security Controls

A bridge between threat intelligence and organisational defence lies in the structured mapping between adversary behaviour and security controls (Rahman & Williams, 2022). The MITRE ATT&CK framework provides a comprehensive taxonomy of adversary tactics, techniques, and procedures (TTPs) observed in real-world attacks, organised into a matrix that enables organisations to reason systematically about which attacker behaviours their security controls are designed to counter (MITRE, 2025).

Recognising the practical value of connecting this taxonomy to prescriptive defensive guidance, CIS formalised explicit bidirectional mappings between ATT&CK techniques and CIS Controls (Centre for Internet Security, 2024), enabling organisations to identify which safeguards defend against which adversary techniques and, conversely, which techniques remain unmitigated given their current control implementation. This linkage transforms ATT&CK from a descriptive threat catalog into an actionable governance instrument, allowing security teams to reason about defensive coverage.

Building on these mappings, prior research operationalised the ATT&CK-to-CIS linkages into the Cyber Catalog, a structured knowledge base that connects CIS safeguards to ATT&CK technique families and provides KCI definitions and demonstrated automated incident-to-technique classification via a fine-tuned sentence transformer achieving strong rank correlation with expert judgement (Sherif et al., 2026b). The authors established the technical infrastructure through which organisational control coverage can be systematically connected to specific vulnerability exploitation pathways. This study builds on this foundation, using the Cyber Catalog's ATT&CK-to-safeguard mappings to derive per-CVE KCI coverage scores that contextualise the threat-based KRI within each organisation's control environment.

3. Methodology

The methodology comprises two components. The first component is an evaluation of the threat-informed KRI baseline against CVSS prioritisation using CISA KEV as an organisation-independent indicator of known exploitation (CISA, 2021). The second component specifies the CA-KRI formulation and illustrates its computational behaviour using synthetic KCI and asset-context inputs. The first component supports empirical claims concerning exploitation-oriented vulnerability prioritisation. The second demonstrates how organisational control coverage and asset context can be incorporated into the KRI baseline.

3.1. Data Sources and Contemporary Cohort

The dataset integrates the following publicly available resources:

- **CVE Data:** CVE identifiers, CVSS severity ratings (Critical/High/Medium/Low), CVSS base scores, and CWE classifications, sourced from the National Vulnerability Database (NVD). CVEs published between 1 January 2020 and 30 May 2025 are included, yielding a contemporary cohort of 146,718 vulnerabilities.
- **EPSS Scores:** Point-in-time exploitation probabilities were obtained from the historical daily EPSS archives maintained by FIRST and retrieved on 10 September 2026. For each evaluation CVE, the analysis selected the first archived EPSS score identified on or after its NVD publication date, subject to a maximum lag of two calendar days. The evaluation period crossed an EPSS model transition: scores through 16 March 2025 were generated under version v2023.03.01, while scores from 17 March 2025 were generated under version v2025.03.14.
- **KEV Catalog:** CISA KEV inclusion is used as the binary exploitation label. CVEs included in KEV are treated as confirmed positives, while CVEs not included are treated as KEV-unconfirmed rather than confirmed negatives. For CWE-weight estimation, only the training-partition CVEs and their corresponding KEV labels are used. KEV labels belonging to CVEs in the 2024–2025 evaluation partition do not contribute to weight estimation. This preserves temporal separation between model fitting and evaluation while retaining the closed-world limitation associated with treating KEV non-inclusion as the negative evaluation class.
- **Cyber Catalog:** CIS safeguard-to-ATT&CK technique mappings and KCIs, providing the bridge between vulnerability techniques and control coverage scores.

CVEs were merged by CVE identifier across all data sources and divided chronologically into two non-overlapping partitions. The training partition comprised 87,080 CVEs published between 1 January 2020 and

31 December 2023 and was used exclusively to estimate the CWE prevalence weights. The held-out evaluation partition comprised 59,638 CVEs published between 1 January 2024 and 30 May 2025 and was used to compare CVSS-only and KRI prioritisation. CWE weights were fixed after estimation on the training partition, and no KEV labels from the held-out evaluation partition contributed to their calculation.

The training partition contained 625 KEV-confirmed CVEs, corresponding to a prevalence of 0.72%, while the held-out evaluation partition contained 237 KEV-confirmed CVEs, corresponding to a prevalence of 0.4%. Because KEV-confirmed vulnerabilities were rare, AUPRC was used as the primary discrimination metric. Uncertainty was estimated using 2,000 paired bootstrap samples drawn with replacement from the held-out evaluation partition and implemented with a fixed random seed of 42. Within each bootstrap sample, CVSS-only and KRI were evaluated on the same resampled CVEs. The bootstrap distributions were used to calculate percentile-based 95% confidence intervals for the paired KRI-minus-CVSS differences. Two-sided bootstrap p-values were calculated as twice the smaller proportion of paired differences at or below zero or at or above zero.

3.2. KRI Computation

The threat-informed KRI used in this study is independently implemented according to the multiplicative formulation introduced by (Sherif et al., 2026a):

$$KRI_{test}(v) = EPSS(v) \times \left(\frac{CVSS(v)}{10} \right) \times CWE_weight_{train}(CWE(v)), v \in D_{2024-2025} \quad (1)$$

For the evaluation cohort, $EPSS(v)$ represents the first historical EPSS score available on or after the CVE's NVD publication date, with a maximum permitted lag of two calendar days. It therefore represents the exploitation probability available at approximately the point of vulnerability publication rather than the CVE's current or end-of-study EPSS score. Where no historical score was available within this window, $EPSS(v)$ was set to zero.

The subscript $train$ emphasises that the CWE weight is learned only from the earlier temporal partition, while v belongs to the later held-out evaluation partition. $CVSS(v)/10$ normalises the CVSS base score to the interval $[0, 1]$, and $CWE_weight_{train}(c)$ represents the exploitation prevalence associated with weakness class c , estimated exclusively from the 2020–2023 training partition. The weight is calculated as:

$$CWE_weight_{train}(c) = \frac{KEV_count_{train}(c)+1}{total_count_{train}(c)+2}, \quad (2)$$

where $KEV_count_{train}(c)$ is the number of KEV-confirmed CVEs assigned to CWE class c in the training partition, and $total_count_{train}(c)$ is the total number of training CVEs assigned to that class. Laplace smoothing with $\alpha = 1$ is applied to avoid zero-valued weights and reduce instability for sparsely represented CWE classes (Chen & Goodman, 1999).

After estimation, the CWE weights are fixed and applied without re-estimation to CVEs in the 2024–2025 evaluation partition. Consequently, no evaluation-period KEV labels contribute to the KRI scores being evaluated. For a CWE class not observed in the training partition, the weight is set to the smoothed global KEV prevalence of the training partition:

$$CWE_weight_{unseen} = \frac{KEV_count_{train}+1}{total_count_{train}+2} \quad (3)$$

This temporal separation prevents information from the held-out evaluation labels from entering the fitted KRI score and ensures that the reported performance reflects application to subsequently disclosed vulnerabilities.

3.3. Illustrative Synthetic Context Scenarios

Real KCI measurements require access to organisation-specific control-assessment data that are generally unavailable in public vulnerability datasets (Boggs et al., 2014). To demonstrate how CA-KRI operates when supplied with varying levels of control coverage, this study uses three reproducible synthetic parameter scenarios. KCI values are generated on the bounded interval $[0,1]$ using Beta distributions, allowing the illustrative scenarios to represent lower, intermediate, and higher levels of safeguard implementation (Hubbard & Seiersen, 2023). The purpose of these values is to introduce controlled variation through which the mathematical and ranking behaviour of the CA-KRI formulation can be examined.

Table 1 summarises three Beta distributions used to generate reproducible differences in assumed control coverage (CIS, 2025b). These parameterisations were selected to produce lower, intermediate, and higher

synthetic KCI values across the bounded interval [0, 1]. They are illustrative assumptions rather than empirically calibrated representations of organisational maturity, CIS Implementation Groups, or observed safeguard adoption (Livieris et al., 2024).

Scenario	Beta(a,b)	Mean KCI
Low coverage	Beta(2,5)	0.29
Medium coverage	Beta(4,4)	0.50
High coverage	Beta(6,2)	0.75

Table 1. Illustrative synthetic KCI control-coverage parameterisations.

The scenarios provide reproducible contextual inputs for illustrating the proposed CA-KRI formulation under different assumed levels of control implementation. They should not be interpreted as measured organisational profiles. The synthetic KCI values are generated independently of the KEV outcome and are used only to demonstrate the generation of contextual inputs. Real KCI measurements may produce different distributions across organisations, safeguards, assets, and vulnerabilities.

3.4. KCI Coverage per CVE

For each CVE, a technique-level KCI score is derived by traversing the Cyber Catalog, developed by Sherif et al. (2026b), in three steps. First, each CVE carries one or more CWE classifications sourced from the NVD. Second, the Cyber Catalog provides a curated lookup table of 54 entries that map CWE weakness families to 12 MITRE ATT&CK technique families. Given a CVE's CWE classification, its ATT&CK technique-family assignment is therefore determined through this predefined mapping, yielding one to two associated technique families per CVE. Third, each identified ATT&CK technique family is linked to the set of CIS safeguards that defend against it, again using the Cyber Catalog's predefined ATT&CK-to-safeguard mappings. This creates a fully traceable pipeline of CVE to CWE to ATT&CK technique family to CIS safeguards, through which organisational control coverage can be connected to vulnerability exploitation pathways.

The mapping is intentionally defined at the ATT&CK technique-family level rather than at the fine-grained individual-technique level. This design reflects the granularity at which CIS safeguard coverage can be meaningfully attributed and ensures complete population coverage across the 2020–2025 CVE cohort used in this study. As a result, all CVEs in the evaluation dataset can be assigned to at least one ATT&CK technique family, avoiding sparsity in downstream KCI coverage computation while maintaining a consistent, reproducible, and operationally interpretable mapping process. The KCI coverage score for each CVE is then computed as the mean KCI score across all safeguards linked to its associated technique family or families:

$$KCI_coverage(v) = \text{mean}(\{ KCI(s) : s \in \text{Safeguards}(\text{Techniques}(v)) \}), \quad (4)$$

where $\text{Safeguards}(\text{Techniques}(v))$ denotes the set of CIS safeguards linked to all ATT&CK techniques associated with vulnerability v , and $KCI(s)$ is the implementation score of safeguard s within the organisation's control environment, bounded in [0,1]. The mean aggregation treats all relevant safeguards as equally weighted contributors to coverage, reflecting the assumption that each defending safeguard addresses a distinct aspect of the exploitation pathway and that partial coverage across multiple safeguards provides proportionally reduced protection. CVEs with no ATT&CK technique mapping would receive the global mean KCI score as a neutral fallback, ensuring no CVE is inadvertently disadvantaged by a data gap. Full population coverage is guaranteed by the Cyber Catalog's 54 entries spanning 12 ATT&CK technique families, which collectively map to all CVEs in the complete 2020–2025 study cohort.

3.5. Asset-Context Exposure-Penalty

Technique-level KCI coverage alone assigns identical scores to all CVEs within the same ATT&CK technique family, since coverage is derived from the same set of safeguards for each technique. This prevents meaningful re-ranking of vulnerabilities that share a technique mapping but differ in their operational context. To address this limitation, an asset-context Exposure-Penalty is introduced that adjusts the effective control coverage downward for CVEs associated with higher-risk asset configurations:

$$\text{Exposure-Penalty}(v) = \alpha \times \text{internet_exposed}(v) + \beta \times (\text{asset_criticality}(v) = \text{Critical}), \quad (5)$$

where $internet_exposed(v)$ and $(asset_criticality(v) = Critical)$ are binary indicators, $\alpha = 0.20$ and $\beta = 0.10$ are their respective weights, and the total penalty is capped at 0.30. The rationale for these design choices is grounded in CVSS Environmental metrics principles where even well-implemented controls offer reduced residual protection for internet-facing assets, which present a larger attack surface, while critical assets represent higher-value targets for adversaries, justifying elevated residual risk assessments even where controls are strong (Shimizu & Hashimoto, 2026).

In the absence of real organisational asset inventories, the two binary indicators are assigned synthetically to provide reproducible variation for illustrating the CA-KRI calculation. Specifically, $internet_exposed(v)$ is generated as a Bernoulli variable with $p=0.30$, while $asset_criticality(v)=Critical$ is generated as a Bernoulli variable with $p=0.20$. The assignments are generated independently for each CVE using a fixed random seed of 42.

These values are illustrative assumptions. They are generated independently of the KEV label and therefore have no designed relationship with known exploitation status. Consequently, any changes observed after applying these contextual variables cannot be interpreted as evidence that CA-KRI improves or degrades organisation-specific prioritisation. In an operational implementation, the variables would instead be obtained from asset inventories, or other organisational systems that record internet accessibility and asset criticality.

The Exposure-Penalty is combined with KCI coverage to produce the proposed Effective-Coverage modifier:

$$Effective-Coverage(v, org) = KCI_coverage(v, org) \times (1 - Exposure-Penalty(v, org)) \quad (6)$$

Under the assumptions of the proposed formulation, the scaling term reduces the attributed control coverage for vulnerabilities affecting internet-facing or critical assets. When the Exposure-Penalty is zero, Effective-Coverage is equal to the underlying KCI coverage. As the penalty increases, Effective-Coverage decreases accordingly. This relationship represents the intended logic of the proposed design and is not an empirically validated estimate of organisational control effectiveness.

3.6. CA-KRI Formula

Having defined the KRI score and the proposed Effective-Coverage modifier, the two components are combined in the Context-Adjusted Key Risk Indicator:

$$CA-KRI(v, org) = KRI(v) \times (1 - Effective-Coverage(v, org)), \quad (7)$$

where $(1 - Effective-Coverage(v, org))$ is the proposed residual-exposure modifier for vulnerability v in organisation org . Under the assumptions of the formulation, high Effective-Coverage discounts the threat-informed KRI score, while low Effective-Coverage retains a larger proportion of that score. CA-KRI is therefore designed to rank otherwise comparable vulnerabilities differently when their associated control coverage or asset context differs. This interpretation represents the intended decision logic of the proposed model, not a validated estimate of residual organisational risk. Establishing whether the adjustment produces more accurate or useful remediation priorities requires independently measured organisational inputs and an organisation-sensitive evaluation outcome.

CA-KRI combines the threat-informed KRI baseline with the proposed Effective-Coverage modifier. KRI integrates EPSS exploitation probability, normalised CVSS severity, and temporally estimated CWE exploitation prevalence. Effective-Coverage combines safeguard-level KCI measurements linked through the CWE-to-ATT&CK-to-CIS mapping in Equation 4 with the asset-context Exposure-Penalty defined in Equation 5 and the scaling relationship specified in Equation 6.

Under the proposed formulation, higher Effective-Coverage produces a larger discount to the KRI baseline, while lower Effective-Coverage retains a larger proportion of the threat-informed score. The resulting CA-KRI value is intended to support context-sensitive differentiation among vulnerabilities whose organisational control coverage or asset conditions differ. This interpretation describes the intended decision logic of the formal model. It does not establish that CA-KRI accurately measures residual organisational risk or improves remediation decisions, both of which require validation using real organisational inputs and an organisation-sensitive reference outcome.

3.7. Illustrative Worked Example

To illustrate the numerical operation of CA-KRI, consider three hypothetical CVEs with the same threat-informed KRI score of 0.6 but different assumed levels of Effective-Coverage. These values are selected solely

to demonstrate the behaviour of the proposed formula and do not represent observed organisational measurements or validated levels of risk reduction.

As shown in Table 2, an Effective-Coverage value of 0.2 produces a CA-KRI score of $0.6 \times (1 - 0.2) = 0.48$. Increasing Effective-Coverage to 0.5 produces a CA-KRI score of 0.3, while an Effective-Coverage value of 0.8 produces a CA-KRI score of 0.12. Under these assumptions, identical vulnerabilities receive lower context-adjusted scores when the associated control coverage is assumed to be higher. The example demonstrates the intended numerical behaviour of the modifier, and it does not establish that the resulting scores measure residual organisational risk or improve remediation decisions.

CVE	KRI	Effective-Coverage	Residual-exposure modifier 1 – Effective-Coverage	CA-KRI
CVE-A	0.6	0.2	0.8	0.48
CVE-B	0.6	0.5	0.5	0.3
CVE-C	0.6	0.8	0.2	0.12

Table 2. Illustrative calculation of CA-KRI for hypothetical CVEs with an identical KRI score and different Effective-Coverage values.

3.8. Scope of Inference and Validation Boundary

The two components of this study have different evidential boundaries. CISA KEV records whether there is authoritative public evidence that a vulnerability has been exploited (CISA, 2021). It contains no information about the controls implemented by a particular organisation, the exposure or criticality of its affected assets, the consequences of exploitation within that organisation, or the effectiveness of its remediation decisions. KEV is therefore an appropriate reference outcome for evaluating the exploitation-oriented KRI baseline.

The contextual variables used in the CA-KRI illustration are generated independently of KEV. Therefore, similarity or difference between KRI and CA-KRI on KEV-based metrics cannot establish whether the contextual modifier is beneficial, ineffective, or harmful for a particular organisation. Such comparisons describe how the proposed modifier interacts numerically with an organisation-independent exploitation label.

Validation of CA-KRI requires independently measured organisational KCI values, asset exposure and criticality information, and an outcome that reflects organisation-specific residual risk or decision quality. Suitable outcomes could include observed organisation-specific exploitation or incidents, independently elicited expert residual-risk judgements, or longitudinal remediation outcomes.

3.9. Evaluation Metrics

CVSS-only and KRI are evaluated exclusively on the held-out 2024–2025 partition. The KRI scores use CWE weights estimated and fixed on the 2020–2023 training partition. Receiver Operating Characteristic Area Under the Curve (ROC-AUC) measures overall discrimination (Davis & Goadrich, 2006), while Area Under the Precision-Recall Curve (AUPRC) is the primary metric because of the low prevalence of KEV-confirmed CVEs. Precision@k and RVC@k assess performance at operational review thresholds, and severity-stratified Recall@500 examines retrieval across CVSS severity bands (Saito & Rehmsmeier, 2015). All confidence intervals for paired differences and comparative p-values were calculated using 2,000 paired bootstrap samples drawn from the held-out evaluation partition.

Precision@k (P@k) measures the fraction of KEV-confirmed exploited CVEs within the top-k ranked vulnerabilities at $k \in \{100, 500, 1000\}$, reflecting realistic analyst daily triage capacity thresholds (Järvelin & Kekäläinen, 2002). Remediation Value Capture@k (RVC@k) measures the fraction of total CVSS-weighted KEV impact recovered by reviewing the top-k ranked CVEs, quantifying risk reduction per unit of analyst effort (Shimizu & Hashimoto, 2026).

Lastly, Severity-Stratified Recall@500 extends this evaluation by reporting recall of KEV-confirmed CVEs within each CVSS severity band, Critical, High, Medium, and Low, when reviewing the top 500 ranked vulnerabilities, revealing whether scoring strategies concentrate remediation attention on the most severe exploited CVEs (Buckley & Voorhees, 2000).

4. Results

4.1. Illustrative KCI Parameter Distributions

Figure 1 presents the synthetic KCI score distributions generated for the three illustrative parameter scenarios. The Low scenario uses a Beta(2,5) distribution and produces a mean KCI score of 0.29, with most values concentrated toward the lower part of the [0, 1] interval. The Medium scenario uses a Beta(4,4) distribution and produces a mean of 0.51, with values distributed approximately symmetrically around the centre of the interval. The High scenario uses a Beta(6,2) distribution and produces a mean of 0.75, with most values concentrated toward the upper part of the interval.

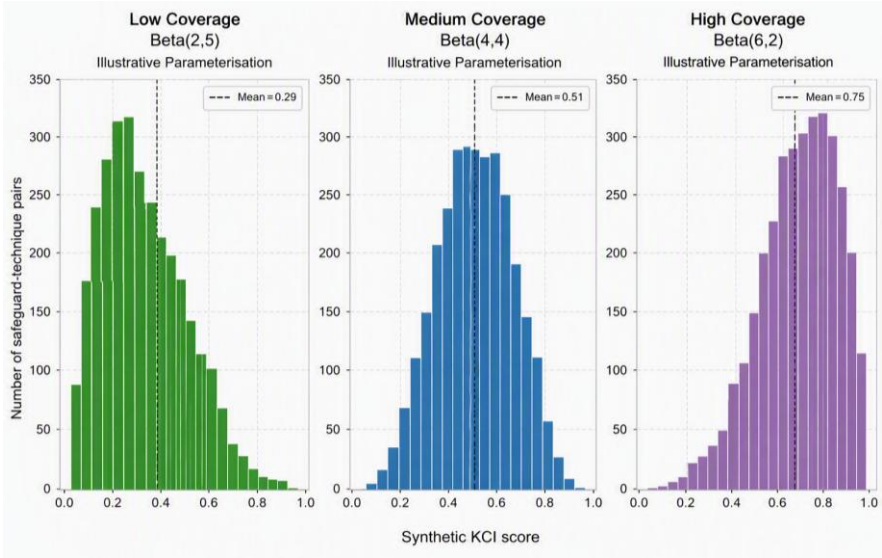


Figure 1. Illustrative synthetic KCI score distributions under control-coverage parameterisations.

The distributions provide reproducible variation for illustrating how different assumed levels of control coverage affect the numerical operation of the CA-KRI formulation. They are not measurements collected from real organisations and should not be interpreted as empirically validated representations of organisational maturity, CIS Implementation Groups, or KCI distributions in practice. Their separation demonstrates only that the selected Beta parameterisations generate distinguishable synthetic input conditions. It does not establish that the resulting CA-KRI rankings are more accurate or operationally beneficial.

4.2. Temporal Holdout Evaluation: KRI versus CVSS-Only

Table 3 reports the evaluation of CVSS-only and KRI on the held-out Dec 2024 to May 2025 cohort. The CWE weights used by KRI were estimated exclusively from the Jan 2020 to Dec 2023 training partition and fixed before evaluation. The reported results therefore contain no category-level leakage from the held-out KEV labels. The comparison tests whether exploitation prevalence learned from earlier CWE patterns, when combined with EPSS and CVSS, generalises to subsequently published vulnerabilities.

Strategy	ROC-AUC	AUPRC	P@100	P@500	P@1000	RVC@100	RVC@500	RVC@1000
CVSS-only	0.765	0.014	0.030	0.026	0.019	0.015	0.066	0.096
KRI	0.807	0.053	0.180	0.070	0.045	0.075	0.146	0.187

Table 3. CVSS-only and KRI performance on the held-out 2024–2025 evaluation cohort.

Using point-in-time EPSS scores, the KRI achieved a ROC-AUC of 0.807, compared with 0.765 for CVSS-only. The KRI therefore provided the strongest overall discrimination between KEV-confirmed and non-KEV vulnerabilities. KRI achieved AUPRC of 0.053 compared with 0.014 for CVSS-only. At the narrowest operational cut-off, KRI achieved the highest P@100, identifying 18 KEV-confirmed vulnerabilities among the top 100, compared with three under CVSS-only prioritisation.

4.3. Discrimination Performance on the Temporal Holdout Cohort

Figure 2 presents the ROC curves for CVSS-only and KRI on the held-out 2024 to 2025 evaluation cohort using point-in-time EPSS scores. KRI achieved a ROC-AUC of 0.807, compared with 0.765 for CVSS-only. The paired difference of 0.042 was statistically supported, with a 95% bootstrap confidence interval of [0.005, 0.075] and $p=0.019$. The result indicates a modest improvement in overall discrimination after preventing later EPSS updates from entering the evaluation.

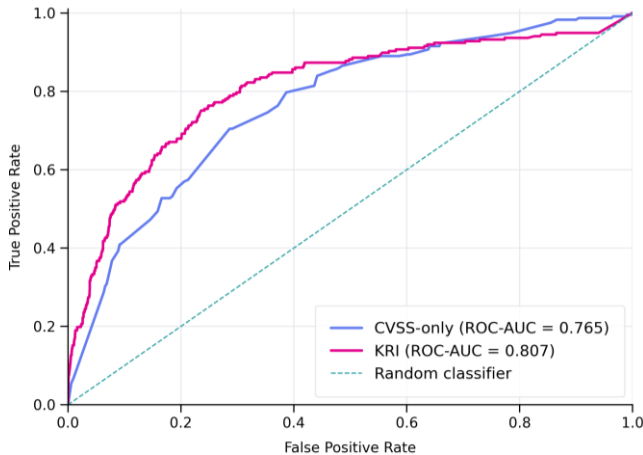


Figure 2. ROC curves for CVSS-only and KRI on the held-out 2024–2025 evaluation cohort.

Figure 3 presents the precision-recall curves for CVSS-only and KRI on the held-out 2024 to 2025 evaluation cohort using point-in-time EPSS scores. KRI achieved an AUPRC of 0.053, compared with 0.014 for CVSS-only. The paired difference of 0.039 was statistically supported, with a 95% bootstrap confidence interval of [0.018, 0.072] and $p<0.001$. Both values exceeded the evaluation-set KEV prevalence of approximately 0.004, while KRI provided a greater concentration of KEV-confirmed vulnerabilities across the precision-recall evaluation.

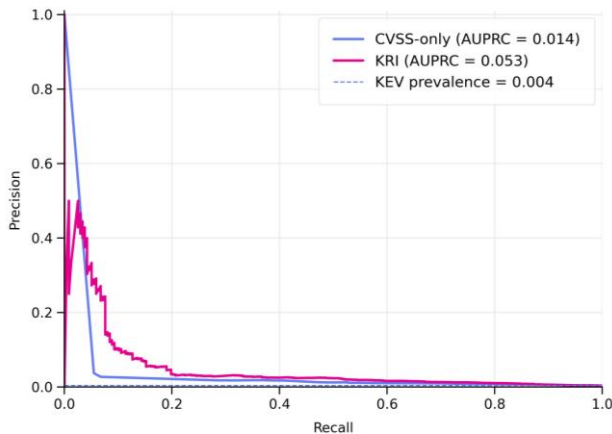


Figure 3. Precision-Recall curves for CVSS-only and KRI on the held-out 2024–2025 evaluation cohort.

4.4. Operational Prioritisation Performance

Figure 4 presents Precision@k for CVSS-only and KRI at review thresholds of $k = 100, 500$, and $1,000$. At $k = 100$, KRI achieved a precision of 0.180 , compared with 0.030 for CVSS-only. At $k = 500$, the corresponding values were 0.070 and 0.026 , while at $k = 1000$, they were 0.045 and 0.019 . The paired KRI-minus-CVSS differences were 0.150 at $k = 100$, 0.044 at $k = 500$, and 0.026 at $k = 1000$. Their respective 95% bootstrap confidence intervals were $[0.050, 0.240]$, $[0.016, 0.066]$, and $[0.008, 0.039]$, with $p=0.002$, $p=0.002$, and $p=0.005$. Although precision declined as the review threshold expanded, KRI retained a statistically supported advantage over CVSS-only at each evaluated threshold.

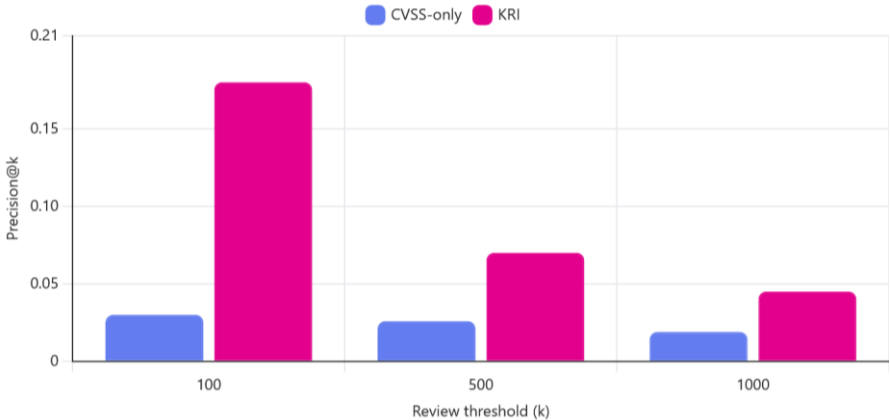


Figure 4. Precision@k for CVSS-only and KRI on the held-out 2024–2025 evaluation cohort.

Figure 5 presents RVC@k for CVSS-only and KRI on the held-out evaluation cohort. At $k = 100$, KRI captured 7.5% of total CVSS-weighted KEV value, compared with 1.5% for CVSS-only. At $k = 500$, KRI captured 14.6% , compared with 6.6% , while at $k = 1000$, the corresponding values were 18.7% and 9.6% . The paired KRI-minus-CVSS differences were 0.06 , 0.08 , and 0.09 , with respective 95% bootstrap confidence intervals of $[0.017, 0.097]$, $[0.014, 0.128]$, and $[0.005, 0.145]$. The corresponding p-values were 0.002 , 0.011 , and 0.037 . KRI therefore captured a greater proportion of CVSS-weighted KEV value than CVSS-only at each evaluated review threshold.



Figure 5. RVC@k for CVSS-only and KRI on the held-out 2024–2025 evaluation cohort.

4.5. Severity-Stratified Recall

Table 4 reports severity-stratified Recall@500 for CVSS-only and KRI on the held-out 2024 to 2025 evaluation cohort. KRI retrieved a greater proportion of KEV-confirmed High- and Medium-severity vulnerabilities, whereas CVSS-only produced slightly higher recall in the Critical category. For Critical-severity vulnerabilities, Recall@500 was 0.117 for CVSS-only and 0.096 for KRI. For High-severity vulnerabilities, recall increased from 0.018 under CVSS-only to 0.209 under KRI, while Medium-severity recall increased from 0 to 0.091. Recall was not calculated for the Low-severity category because the evaluation cohort contained no KEV-confirmed Low-severity CVEs.

CVSS severity	KEV-confirmed	CVSS-only	KRI
Critical	94	0.117	0.096
High	110	0.018	0.209
Medium	33	0	0.091
Low	0	Not applicable	Not applicable

Table 4. Severity-stratified Recall@500 for CVSS-only and KRI on the held-out 2024–2025.

The severity-stratified results show that KRI’s operational advantage was concentrated among High- and Medium-severity KEV vulnerabilities rather than being uniform across all severity bands. Within the overall top 500, KRI captured 20.9% of High-severity and 9.1% of Medium-severity KEV vulnerabilities, compared with 1.8% and 0% under CVSS-only prioritisation. CVSS-only produced slightly higher recall among Critical vulnerabilities, at 11.7% compared with 9.6% for KRI. This pattern is consistent with CVSS-only prioritisation concentrating heavily on the highest technical-severity category, while KRI also surfaces known exploited vulnerabilities outside that category.

5. Discussion

5.1. Interpreting the Contemporary Evaluation

The temporal evaluation controls two sources of leakage. First, the CWE exploitation-prevalence weights were estimated from the Jan 2020 to Dec 2023 training partition and applied without re-estimation to the Jan 2024 to May 2025 evaluation cohort. Second, each evaluation CVE was assigned the first historical EPSS score identified on or within two days after publication. Under this temporally controlled design, KRI achieved a ROC-AUC of 0.807, compared with 0.765 for CVSS-only, and an AUPRC of 0.053, compared with 0.014. The corrected results indicate that KRI retains a statistically supported advantage over CVSS-only.

The operational results further show that the KRI advantage depends on the decision threshold and severity composition of the reviewed set. KRI produced higher Precision@k and RVC@k values than CVSS-only at all three evaluated thresholds. At $k = 500$, for example, KRI doubled the proportion of CVSS-weighted KEV value captured from 6.6% to 14.6%. Severity-stratified analysis showed that this advantage was concentrated among High- and Medium-severity KEV vulnerabilities, while CVSS-only retained slightly higher recall among Critical vulnerabilities. KRI should therefore be interpreted as complementing technical severity with exploitation-oriented information.

5.2. Validation Boundary of CA-KRI

KEV records evidence of exploitation independently of the controls, assets, and business conditions of any organisation. CA-KRI, by contrast, is designed to modify vulnerability priorities according to local control coverage and asset context. KEV therefore does not represent the organisation-specific construct that distinguishes CA-KRI from KRI.

This validation boundary is reinforced by the synthetic contextual inputs, which are generated independently of KEV. Comparisons against KEV could not determine whether a contextual adjustment reflects useful organisation-specific information or merely introduces unrelated ranking variation. The revised empirical analysis therefore evaluates only KRI against CVSS-only prioritisation. CA-KRI is treated as a formal design proposal whose practical value requires independently measured organisational inputs and an organisation-sensitive reference outcome.

5.3. Theoretical Coherence of CA-KRI

The CA-KRI formula is grounded in the expected-loss framework of risk management. Expected loss can be decomposed as: $E[Loss] \propto Threat \times Vulnerability \times Impact$ (Cox, 2008). KRI captures Threat (EPSS) and Vulnerability (CVSS, CWE). CA-KRI adds a Control dimension: *Effective-Coverage* is intended to represent the extent to which the vulnerability dimension may be reduced through control implementation. The Exposure-Penalty acknowledges that control effectiveness is not asset-agnostic: internet-facing systems and critical assets face higher threat density, reducing the practical benefit of even well-implemented controls. This decomposition is consistent with established risk frameworks including NIST SP 800-30 (NIST, 2012).

5.4. The Role of KCI Data Quality

The organisation-specific value of CA-KRI depends on the validity, granularity, and timeliness of its KCI and asset inputs. The synthetic scenarios used in this study cannot reproduce the dependencies, measurement errors, implementation differences, and asset relationships that may exist in real organisations. They also cannot establish whether variation in control coverage would produce more useful remediation rankings.

Real KCI measurements may generate ranking behaviour that differs from the illustrative scenarios, but the direction and practical value of those differences remain unknown. Future evaluation should therefore examine CA-KRI using independently collected safeguard-level measurements and asset information rather than assuming that real data will necessarily strengthen the contextual signal.

5.5. CA-KRI Within the Broader Risk Management Pipeline

The CA-KRI framework occupies a distinct position within the data-driven cyber risk management pipeline introduced in Sherif et al. (2026b), see Section 2.4 for a full description. The present work extends that pipeline by adding the KCI-based contextual modifier that connects control coverage to individual CVE prioritisation, then incorporates that coverage as a contextual modifier of the KRI score. The result is a vulnerability prioritisation framework that is designed to be threat-aware, control-aware, and asset-context-aware.

It is important to acknowledge that vulnerability prioritisation in practice extends beyond the dimensions captured by CA-KRI. Remediation decisions are also shaped by the business criticality of the processes and functions supported by an affected asset, a technically low-severity vulnerability on a mission-critical system may warrant immediate attention regardless of its KRI score.

Sector-specific regulatory and compliance obligations (e.g., NIS2, PCI-DSS, GDPR) impose additional prioritisation constraints that are external to technical risk metrics. Operational constraints, including maintenance windows, change management capacity, vendor patch availability, and staffing cycles, further bound the set of remediation actions feasible within any given window. CA-KRI is therefore best understood as a decision-support input to a broader governance process rather than a standalone prioritisation authority. Future work integrating CA-KRI with business impact assessments and regulatory requirement mappings would strengthen its applicability as a comprehensive cyber risk governance tool.

5.6. Practical Implications for SMEs

Resource-constrained SMEs face disproportionate vulnerability management challenges due to limited analyst capacity and constrained security budgets (Calvo-Manzano et al., 2025). The evaluated KRI baseline is relevant to resource-constrained SMEs because its principal inputs are publicly available and can support exploitation-oriented triage without requiring organisation-specific control data. Its core data sources, NVD, EPSS, and KEV, are publicly accessible, reducing the information-acquisition barrier for organisations with limited security resources.

On the held-out evaluation cohort, KRI increased Precision@100 from 0.030 for CVSS-only to 0.180 and increased RVC@500 from 6.6% to 14.6%. In practical terms, reviewing the first 100 KRI-ranked vulnerabilities identified 18 KEV-confirmed CVEs, compared with three under CVSS-only prioritisation. These results indicate that constrained review capacity can be directed toward a better concentration of KEV-confirmed vulnerabilities than under severity-only prioritisation.

For resource-constrained organisations, the evaluated KRI baseline may provide a low-cost starting point for exploitation-oriented triage using publicly available data. CA-KRI provides a formal mechanism through which local control coverage and asset context could subsequently be incorporated into that baseline. However, the synthetic analysis does not establish that this adjustment improves SME remediation decisions.

Its operational value and trade-offs require evaluation using real organisational inputs and organisation-sensitive outcomes.

5.7. Limitations and Future Work

The principal limitation of the CA-KRI analysis is not only that its contextual inputs are synthetic, but also that the available evaluation outcome is structurally unable to measure its intended organisation-specific contribution. KEV identifies vulnerabilities with authoritative evidence of exploitation but does not record local control effectiveness, asset exposure, asset criticality, business consequences, or residual organisational risk. The present study therefore does not empirically validate CA-KRI.

Future validation of CA-KRI should combine real organisational inputs with an organisation-sensitive reference outcome. The required inputs include independently measured safeguard-level KCI values, asset exposure, asset criticality, and the relationships between vulnerabilities and affected assets. These measurements should be paired with outcomes such as observed local incidents, security-team remediation decisions, or longitudinal remediation effectiveness.

The held-out Jan 2024 to May 2025 evaluation cohort contains 237 KEV-confirmed vulnerabilities among 59,638 records. Although this provides a temporally relevant evaluation, the low positive prevalence increases uncertainty in precision-based metrics. The reported bootstrap confidence intervals quantify this uncertainty, while future temporal replications may provide more stable estimates as additional recently disclosed vulnerabilities enter KEV.

The KEV catalog is a curated, operationally prioritised signal maintained for actionable threat intelligence across the federal enterprise; it is not an exhaustive register of all exploited CVEs. CVEs may be absent from KEV because they were disclosed recently and have not yet been triaged, because exploitation occurred in targeted or sector-specific campaigns with limited public reporting, or because the threat did not meet CISA's significance threshold despite confirmed exploitation. Treating non-KEV CVEs as the negative evaluation class embeds a closed-world assumption and may introduce false negatives into the reference outcome. Treating KEV-unconfirmed CVEs as the negative evaluation class may introduce outcome misclassification. The direction and magnitude of any resulting bias cannot be assumed to be identical across scoring strategies because unrecorded exploitation may be distributed differently across their rankings.

Precision@k and RVC@k should therefore be interpreted as measuring the concentration and CVSS-weighted value of KEV-confirmed vulnerabilities, not all exploited vulnerabilities. Future evaluations using broader or multi-source exploitation evidence would strengthen the external validity of the KRI findings.

6. Conclusion

This paper presented two contributions to vulnerability prioritisation. First, it evaluated the threat-informed KRI baseline against CVSS-only prioritisation using a contemporary dataset of 146,718 CVEs published between 2020 and 2025. Second, it introduced CA-KRI as a formal organisation-aware extension that incorporates KCI-based control coverage and asset-level exposure context into the KRI framework.

The temporally isolated evaluation indicates that KRI improves KEV-oriented vulnerability prioritisation relative to CVSS-only on CVEs published between 2024 and 2025. Using point-in-time EPSS scores, KRI achieved a ROC-AUC of 0.807, compared with 0.765 for CVSS-only, and an AUPRC of 0.053, compared with 0.014. At $k = 500$, KRI captured 14.6% of CVSS-weighted KEV value, compared with 6.6% for CVSS-only. The paired ROC-AUC difference was 0.042, with a 95% bootstrap confidence interval of [0.005, 0.075] and $p = 0.019$. KRI also produced statistically supported improvements in AUPRC, Precision@k, and RVC@k at the evaluated review thresholds. Because the CWE weights were estimated exclusively from the 2020 to 2023 training partition and the KRI calculation used point-in-time EPSS scores, the evaluation limits both category-level label leakage and look-ahead bias arising from later EPSS updates.

The study specifies CA-KRI as an extension of KRI that incorporates KCI-based control coverage and asset context into vulnerability scoring. The synthetic scenarios illustrate the generation of reproducible contextual inputs for the proposed modifier. They do not establish whether CA-KRI produces accurate or beneficial organisation-specific rankings. The similarity between KRI and CA-KRI on KEV-based metrics is not treated as evidence of effective contextual re-ranking. KEV is organisation-independent, and the synthetic contextual inputs are generated independently of it. The findings therefore apply to the KRI baseline, while validation of CA-KRI requires real organisational measurements.

This work extends research on data-driven cyber risk management in two ways. It examines a threat-informed KRI as an alternative to CVSS-only prioritisation, and it provides a design for incorporating organisational control coverage and asset context through CA-KRI. The latter should be regarded as a research

proposition for subsequent field evaluation. Organisations with established CIS Controls assessment processes, asset-management systems, and sufficiently detailed security-incident records would provide appropriate settings for evaluating whether CA-KRI produces more useful residual-risk and remediation decisions.

Data Availability

To support reproducibility, the datasets used in this study are publicly available at <https://doi.org/10.5281/zenodo.22705694>.

References

- Alexopoulos, N., Brack, M., Wagner, J. P., Grube, T., & Mühlhäuser, M. (2022). How long do vulnerabilities live in the code? A large-scale empirical measurement study on FOSS vulnerability lifetimes. In *Proceedings of the 31st USENIX Security Symposium* (pp. 359–376). USENIX Association.
- Allodi, L., & Massacci, F. (2014). Comparing vulnerability severity and exploits using case-control studies. *ACM Transactions on Information and System Security*, 17(1), Article 1, 1–20. <https://doi.org/10.1145/2630069>.
- Boggs, N., Zhao, H., Du, S., & Stolfo, S. J. (2014). Synthetic data generation and defense in depth measurement of web applications. In A. Stavrou, H. Bos, & G. Portokalidis (Eds.), *Research in attacks, intrusions and defenses: 17th International Symposium, RAID 2014 (Lecture Notes in Computer Science, Vol. 8688, pp. 234–254)*. Springer. https://doi.org/10.1007/978-3-319-11379-1_12.
- Buckley, C., & Voorhees, E. M. (2000). Evaluating evaluation measure stability. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 33–40). Association for Computing Machinery. <https://doi.org/10.1145/345508.345543>.
- Calvo-Manzano, J. A., San Feliu, T., Herranz, Á., Mariño, J., Fredlund, L. Å., & Moreno, A. M. (2025). CyberESP: An integrated cybersecurity framework for SMEs. *Journal of Software: Evolution and Process*, 37(9), e70050. <https://doi.org/10.1002/smr.70050>.
- Centre for Internet Security. (2021). CIS Critical Security Controls Version 8. <https://www.cisecurity.org/controls/v8>.
- Centre for Internet Security. (2024). CIS Controls v8 Master Mapping to MITRE Enterprise ATT&CK v82. <https://www.cisecurity.org/insights/white-papers/cis-controls-v8-master-mapping-to-mitre-enterprise-attck-v82>.
- Centre for Internet Security. (2025a). CIS Controls Assessment Specification. <https://www.cisecurity.org/controls/cis-controls-assessment-specification>.
- Chen, S. F., & Goodman, J. (1999). An empirical study of smoothing techniques for language modeling. *Computer Speech & Language*, 13(4), 359–394.
- Centre for Internet Security. (2025b). CIS Community Defense Model v2.0. <https://www.cisecurity.org/insights/white-papers/cis-community-defense-model-2-0>.
- Cybersecurity and Infrastructure Security Agency. (2021). Known Exploited Vulnerabilities Catalog. <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>.
- Cox, Jr, L. A. (2008). Some limitations of “Risk = Threat × Vulnerability × Consequence” for risk analysis of terrorist attacks. *Risk Analysis: An International Journal*, 28(6), 1749–1761.
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233–240). Association for Computing Machinery. <https://doi.org/10.1145/1143844.1143874>.
- El Moutaouakil, S., Lindström, J., & Andersson, K. (2026). Evaluating the Effectiveness of Information Security Management Systems: An Analysis Framework and Key Metrics. *Journal of Cybersecurity and Privacy*, 6(2), 73.
- Forum of Incident Response and Security Teams. (2005). Introduction to CVSS. <https://www.first.org/cvss/v1/intro>.
- Forum of Incident Response and Security Teams. (2025). The Exploit Prediction Scoring System. <https://www.first.org/epss/>.
- Hubbard, D. W., & Seiersen, R. (2023). *How to measure anything in cybersecurity risk*. John Wiley & Sons.

- Jacobs, J., Romanosky, S., Edwards, B., Adjerid, I., & Roytman, M. (2021). Exploit Prediction Scoring System (EPSS). *Digital Threats: Research and Practice*, 2(3), Article 20, 1–17. <https://doi.org/10.1145/3436242>.
- Jacobs, J., Romanosky, S., Suciu, O., Edwards, B., & Sarabi, A. (2023). Enhancing vulnerability prioritization: Data-driven exploit predictions with community-driven insights. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 194–206).
- Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4), 422-446.
- Koscinski, V., Nelson, M., Okutan, A., Falso, R., & Mirakhorli, M. (2025). Conflicting scores, confusing signals: An empirical study of vulnerability scoring systems. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1904-1918).
- Livieris, I. E., Alimpertis, N., Domalis, G., & Tsakalidis, D. (2024). An evaluation framework for synthetic data generation models. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 320-335).
- MITRE. (2006). Common Weakness Enumeration. The MITRE Corporation. <https://cwe.mitre.org/>.
- MITRE. (2025). MITRE ATT&CK. <https://attack.mitre.org/>.
- National Institute of Standards and Technology. (2012). Guide for conducting risk assessments (NIST Special Publication 800-30 Revision 1). <https://doi.org/10.6028/NIST.SP.800-30r1>.
- Rahman, M. R., & Williams, L. (2022). An investigation of security controls and MITRE ATT&CK techniques. *arXiv preprint arXiv:2211.06500*. <https://doi.org/10.48550/arXiv.2211.06500>.
- Sabetta, A., & Bezzi, M. (2018). A practical approach to the automatic classification of security-relevant commits. In *2018 IEEE International conference on software maintenance and evolution (ICSME)* (pp. 579-582).
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
- Savola, R. M. (2013). Quality of security metrics and measurements. *Computers & Security*, 37, 78-90.
- Sherif, E., Yevseyeva, I., Basto-Fernandes, V., & Cook, A. (2026a). Bridging the gap between security metrics and key risk indicators: An empirical framework for vulnerability prioritization. *arXiv preprint arXiv:2603.12450*. <https://doi.org/10.48550/arXiv.2603.12450>.
- Sherif, E., Yevseyeva, I., Basto-Fernandes, V., & Cook, A. (2026b). Operationalising cyber risk management: Connecting cyber incidents to MITRE ATT&CK techniques, controls, and metrics. *Cybersecurity & Cybercrime*, 1(10), 29–53. <https://doi.org/10.5604/01.3001.0055.8282>.
- Shimizu, N., & Hashimoto, M. (2026). Vulnerability management chaining: An integrated framework for efficient cybersecurity risk prioritization. *IEEE Access*, 14, 31407-31424.