Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), noviembre-diciembre 2026,
Volumen 10, Número 6.

https://doi.org/10.37811/cl_rcm.v10i6

QUANTIFYING A QUALITATIVE RISK REGISTER: PRE- AND POST-MITIGATION MONTE CARLO ANALYSIS OF AN ERP SAP S/4HANA TRANSITION PROJECT FOR A LUBRICANTS MANUFACTURING COMPANY

CUANTIFICACIÓN DE UN REGISTRO CUALITATIVO DE RIESGOS: ANÁLISIS MONTE CARLO DE PRE-MITIGACIÓN Y POST-MITIGACIÓN EN EL PROYECTO DE TRANSICIÓN AL ERP SAP S/4HANA DE UNA EMPRESA FABRICANTE DE LUBRICANTES

Benigno Alfredo Armijos De La Cruz
Universitat Politècnica de València

Jon Arambarri-Basañez
Universidad Europea del Atlántico

DOI: https://doi.org/10.37811/cl_rcm.v10i6.26980

Quantifying a qualitative risk register: Pre- and post-mitigation Monte Carlo analysis of an ERP SAP S/4HANA transition project for a lubricants manufacturing company

Benigno Alfredo Armijos De La Cruz[1]
baarmdel@upv.edu.es
https://orcid.org/0000-0003-4442-2697
Universitat Politècnica de València, Department of Engineering Projects, Valencia
Spain

Jon Arambarri-Basañez
jon.arambarri@uneatlantico.es
https://orcid.org/0000-0002-6450-8562
Universidad Europea del Atlántico, Faculty of Engineering, Santander
Spain

ABSTRACT

Enterprise resource planning projects record their risks on qualitative scales, while schedule and cost commitments require quantitative confidence levels. This study developed and tested a method that converts a qualitative risk register into a Monte Carlo model of schedule and cost and compares the exposure before and after the planned responses. The method was applied to the SAP S/4HANA transition of a lubricants manufacturer, using the 323-task SAP Activate plan, a 20-risk register and 5,000 iterations with common random numbers across three scenarios. The 13 prioritized risks raised the 80th percentile to 592.4 working days and USD 2.79 million, against a plan of 423 days and a budget of USD 1.80 million. Thirteen responses costing 30 days and USD 78,000 lowered it to 519.8 days and USD 2.28 million: contingency reserves fell by 42.9% in time and 51.5% in cost, and the outcome improved in 98.6% of the paired iterations (Wilcoxon test, $p < 0.001$). Committing the 80th percentiles of time and cost separately gave only 74% joint confidence. The results show that an existing register can be quantified without collecting new data.

Keywords: project risk management; Monte Carlo simulation; pre- and post-mitigation analysis; contingency reserve; ERP implementation

[1] Main author.
Correspondence: baarmdel@upv.edu.es

Cuantificación de un registro cualitativo de riesgos: análisis Monte Carlo de pre-mitigación y post-mitigación en el proyecto de transición al ERP SAP S/4HANA de una empresa fabricante de lubricantes

RESUMEN

Los proyectos de implementación de sistemas de planificación de recursos empresariales registran sus riesgos en escalas cualitativas, pero sus compromisos de plazo y costo exigen niveles de confianza cuantitativos. Este estudio desarrolló y probó un método que convierte un registro cualitativo de riesgos en un modelo Monte Carlo de plazo y costo y compara la exposición antes y después de las respuestas planificadas. Se aplicó a la transición a SAP S/4HANA de una fabricante de lubricantes, con el plan SAP Activate de 323 tareas, 20 riesgos y 5.000 iteraciones en tres escenarios. Los 13 riesgos priorizados elevaron el percentil 80 a 592,4 días hábiles y USD 2,79 millones, frente a un plan de 423 días y un presupuesto de USD 1,80 millones. Trece respuestas que costaron 30 días y USD 78.000 lo redujeron a 519,8 días y USD 2,28 millones: las reservas bajaron un 42,9 % en plazo y un 51,5 % en costo, y el resultado mejoró en el 98,6 % de las iteraciones pareadas (Wilcoxon, $p < 0,001$). Comprometer por separado el percentil 80 de plazo y de costo dio solo un 74 % de confianza conjunta. Los resultados muestran que un registro existente puede cuantificarse sin recolectar datos nuevos.

Palabras clave: gestión de riesgos de proyectos; simulación Monte Carlo; pre-mitigación y post-mitigación; reserva de contingencia; implementación de ERP

INTRODUCTION

ERP implementations remain among the riskiest information technology (IT) projects. Large samples of IT projects show fat-tailed cost overruns, where a minority of projects exceeds its budget by several times (Flyvbjerg & Budzier, 2011), and ERP programs add the burden of data migration, custom code, integration and organizational change (Aloini et al., 2007; Ram et al., 2013). Transitions from SAP ERP Central Component (ECC) to SAP S/4HANA concentrate these sources of risk: data structures change, custom developments must be adapted, and legal localization must be certified before the system can operate.

Current standards ask project managers to treat risk both qualitatively and quantitatively. The Project Management Institute standard for risk management in portfolios, programs and projects and ISO 31000:2018 describe a cycle of identification, analysis, response and monitoring in which quantitative analysis supports decisions on reserves and commitments (International Organization for Standardization [ISO], 2018; Project Management Institute [PMI], 2019). Professional practice for schedule and cost risk analysis recommends Monte Carlo simulation over a critical path method (CPM) network, with risks modeled as discrete events that add delay to specific activities (AACE International, 2011; Hulett, 2009). Yet in most IT projects the only risk artifact is a register with probability and impact on five-level scales. The quantitative step is skipped, or it is run as a separate exercise disconnected from the register the team actually maintains (Kwak & Ingall, 2007).

Two gaps follow. First, there is little guidance on how to translate an existing practitioner register into a simulation model without collecting new three-point estimates for every risk, which teams rarely have time to provide. Second, the effect of the planned responses is usually asserted, not measured: registers record a residual level after mitigation, but few studies simulate the inherent and residual registers against the same random draws, and fewer still contrast the estimates with what actually happened. A previous study by the first author and colleagues proposed a pre- and post-mitigation simulation of risk registers (Armijos De La Cruz et al., 2025). The present article extends that approach to a full ERP transition schedule, adds a joint schedule–cost model, and verifies the forecasts against the outcome of the project.

The study addresses three questions. RQ1: how can a qualitative register based on the SAP Activate methodology be converted into a reproducible Monte Carlo model of schedule and cost? RQ2: how much do the planned responses reduce the contingency reserves required at the 80% confidence level? RQ3: how well did the register and the model anticipate the actual outcome of the project? The contribution is a five-step method, tested on a real SAP S/4HANA transition, that keeps the register as the single source of risk data and makes the value of each response visible to the steering committee.

METHODOLOGY

Case and Data Sources

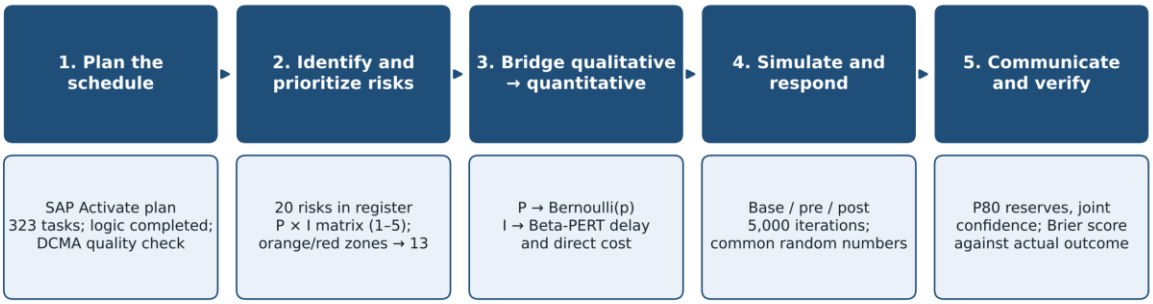
The case is the transition of a lubricants manufacturer in Guayaquil, Ecuador, from SAP ECC 6.0 to SAP S/4HANA. The company produces and sells engine oils, industrial oils and greases through distributors, service centers and fleet customers, and the scope covered order-to-cash, plan-to-produce in its blending and packaging plant, batch and quality management, finance and electronic invoicing. An IT consulting firm based in Guayaquil led the implementation; the first author acted as an external advisor to the firm and was not a member of the project team. The project started on 30 July 2024 with an approved budget of USD 1.80 million.

The study used only existing evidence: project documents, working sessions with the consulting team, and scales agreed with the project director. No interviews or questionnaires were applied. The schedule was the SAP Activate roadmap for the transition to SAP S/4HANA, version 24Q3, which contains 323 leaf tasks across the Discover, Prepare, Explore, Realize, Deploy and Run phases (SAP SE, 2024). The 20 risks of the register were identified in working sessions with the firm, following the register structure of the SAP Activate methodology: cause–event–effect description, category, owner, probability and impact on 1–5 scales, response strategy, status and contingency plan. Closure information on dates, costs and materialized risks was obtained from the firm after the go-live.

The method was implemented by the first author in a web-based risk management system that runs the simulation in the browser; all figures in this article were produced with it. The method has five steps (Figure 1).

Figure 1

Five-step method to quantify a qualitative risk register



Note. Steps 1–3 prepare the model; step 4 runs three scenarios on the same random numbers; step 5 reports reserves and checks the forecasts against the outcome.

Schedule Preparation and Quality Check

The roadmap was imported with its durations and finish-to-start links. The CPM gave a deterministic duration of 423 working days (finish on 12 March 2026) on a five-day calendar without holidays. Because 101 tasks had no successor and 12 more had a total float above 44 working days, their float was unrealistic and could hide near-critical paths in the simulation. The logic was completed with a fixed rule: each open task was linked to the next task of the same deliverable, otherwise to the next task of the same workstream, otherwise to the quality gate of its phase; sequences contrary to the method (for example, configuration before team enablement) were discarded. The rule added 113 links (84 by workstream, 22 by deliverable and 7 by quality gate), for a total of 659 links and 101 critical tasks. The deterministic finish did not change.

The completed schedule was assessed with the 14-point check of the Defense Contract Management Agency (DCMA, 2012), consistent with the schedule assessment guide of the U.S. Government Accountability Office (GAO, 2015). Three tests require a status date and progress data and do not apply to a baseline. The schedule passed 10 of the 11 applicable tests. It failed test 6 (high float): 113 tasks (35.0%) kept more than 44 days of float, mainly in the solution adoption and project management workstreams, which run in parallel to the conversion work. This limitation is discussed in the Discussion.

Qualitative Register and Prioritization

Probability levels 1 to 5 correspond to the bands 1–20%, 21–40%, 41–60%, 61–80% and 81–100%; impact levels range from minimal to very large. The risk level is the product P × I (1–25), and the cells of the matrix are grouped in green, yellow, orange and red zones, where the red zone requires a contingency plan and the orange zone recommends one. Risks in the orange or red zones passed to the quantitative analysis. Of the 20 risks, 13 met this criterion: three in the red zone (R1, R5 and R9) and ten in the orange zone (Figure 2a). The other seven remained under qualitative monitoring.

Bridge from Qualitative Levels to Simulation Inputs

Each prioritized risk i was converted into a discrete event with probability p_i and a delay S_i applied to the tasks it affects. The probability is the midpoint of the band of its level (Table 1):

$$p_i = c(P_i),\ \ c = (10, 30, 50, 70, 90)\%, \quad (1)$$

and the occurrence in each iteration is a Bernoulli trial. The delay follows a modified Beta-PERT distribution (Davis, 2008; Malcolm et al., 1959) whose most likely value m depends on the impact level, $s = (2, 5, 10, 15, 25)$ working days, with minimum $a = 0.5m$ and maximum $b = 2m$:

$$S_i \sim \text{Beta}(\alpha, \beta) \text{ on } [a, b],\ \ \alpha = 1 + \lambda(m - a)/(b - a),\ \ \beta = 1 + \lambda(b - m)/(b - a), \quad (2)$$

$$\text{E}[S_i] = (a + \lambda m + b)/(\lambda + 2),\ \ \lambda = 4. \quad (3)$$

The expected value $\text{VE}_i = p_i \cdot \text{E}[S_i]$ ranks the risks before simulation. Each risk was linked to one or more tasks of the schedule identified in the working sessions; if the risk occurs, a single delay is drawn and added to every linked task, so a risk on a critical task moves the finish of the project, while a risk on a task with float may be partly absorbed (AACE International, 2011; Hulett, 2009). The team could override any converted value; the residual parameters after each response (probability, delay and level) were set by the team for each response.

Cost was modeled jointly with time (AACE International, 2011; GAO, 2020). The total cost C of an iteration is

$$C = F + h \cdot T + e \cdot \Sigma_j D_j + \Sigma_i O_i K_i + \Sigma_i M_i, \quad (4)$$

where F is the fixed cost (30% of the budget), h the daily overhead (40% of the variable budget divided by the deterministic duration), T the simulated project duration, e the effort rate per task-day, D_j the simulated duration of task j, O_i the occurrence indicator, K_i a direct cost drawn from a Beta-PERT whose

most likely value is 0.5%, 1%, 2%, 4% or 7% of the budget according to the impact level, and M_i the cost of the response, counted only in the post-mitigation scenario. With all variables at their deterministic values, C equals the budget.

Table 1 summarizes the default equivalences between qualitative levels and simulation parameters. The team could override any value for an individual risk.

Table 1

Equivalences from qualitative levels to simulation parameters

Level	Probability band	*p* (%)	Impact	Delay *a* / *m* / *b* (d)	Direct cost, most likely (% of budget)
1	1–20%	10	Minimal	1 / 2 / 4	0.5
2	21–40%	30	Small	2.5 / 5 / 10	1
3	41–60%	50	Medium	5 / 10 / 20	2
4	61–80%	70	Large	7.5 / 15 / 30	4
5	81–100%	90	Very large	12.5 / 25 / 50	7

Note. Direct cost uses the same Beta-PERT shape as the delay, with minimum 0.5 and maximum 2 times the most likely value.

Simulation Design

Three scenarios were simulated over the same network. The base scenario models only duration uncertainty: every task received a Beta-PERT at 75%, 100% and 150% of its deterministic duration. The pre-mitigation scenario adds the 13 risks with their inherent parameters. The post-mitigation scenario replaces them with the residual parameters, adds the days of each response to the task that absorbs it, and adds the cost of the responses. The scenarios share common random numbers: each task and each risk has its own seeded stream, and a single occurrence draw per risk and iteration is compared with the inherent and the residual probability. Differences between scenarios therefore reflect the responses, not sampling noise. Each run used 5,000 iterations with seed 20240730; gamma variates for the Beta draws used the method of Marsaglia and Tsang (2000).

The outputs are the empirical distributions of project duration and total cost, summarized by their moments, percentiles, histograms and cumulative distribution functions (S-curves); the contingency reserves, defined as the 80th percentile (P80) minus the deterministic duration and P80 minus the budget; and the bivariate distribution of time and cost. For the latter, the joint commitment is the smallest pair of equal marginal percentiles that are met together in at least 80% of the iterations, and highest-density regions were estimated with a Gaussian kernel and Scott's bandwidth rule (Scott, 1992).

Because the scenarios share random numbers, the effect of the responses was also measured iteration by iteration as the paired difference between the pre- and post-mitigation outcomes, tested with the Wilcoxon signed-rank test (Wilcoxon, 1945). Sensitivity was measured with Spearman rank correlations ρ_k between each input and each output, normalized as a contribution to variance (Vanhoucke, 2010):

$$CV_k = \rho_k^2 / \Sigma_j \rho_j^2. \quad (5)$$

Verification and Ex Post Evaluation

The simulation engine was verified by comparing, for each scenario, the simulated occurrence frequencies of the 13 risks with their assigned probabilities through a chi-square goodness-of-fit test with 13 degrees of freedom. Scenarios were tested separately because they share random numbers. The quality of the register probabilities was evaluated against the risks that actually occurred with the Brier score (Brier, 1950; Gneiting & Raftery, 2007):

$$BS = (1/n) \Sigma_i (p_i - o_i)^2, \quad (6)$$

where $o_i = 1$ if risk i occurred and 0 otherwise. The score was compared with a base-rate forecast that assigns the observed frequency $\bar{o}$ to every risk, whose score is $\bar{o}(1 - \bar{o})$. Discrimination was assessed by comparing the mean probability assigned to the risks that occurred and to those that did not.

Ethical Considerations and Data Availability

The study used project documents and working-session records shared by the consulting firm, with the agreement of the project director; no personal data were collected and no interviews or questionnaires were applied. The names of the firm and of the client company are withheld for confidentiality. The first author served as an external advisor to the consulting firm and received no compensation linked to the results reported here. The risk management system, the schedule with the completed logic and the risk register are available from the corresponding author upon reasonable request.

RESULTS AND DISCUSSION

This section reports the results by scenario, from the qualitative inputs to the simulated distributions, the paired effect of the responses, joint confidence, sensitivity and the ex post verification, and then discusses them against the three research questions.

Prioritized Risks and Responses

Table 2 lists the 13 prioritized risks. Seven are technical and concern data, custom code, integration, testing, costing and quality management; two come from the client organization, and one each from a supplier, consultant resources, the external environment and project management. By expected value, R5 (defects in integrated testing) and R10 (certification of the electronic invoicing add-on) ranked first. The team adopted a mitigation response for each risk (Table 3). Together the responses cost 30 working days and USD 78,000, about 4.3% of the budget, and lowered the risk level by 50.0% to 73.3% (Figure 2). These qualitative inputs only parameterize the simulation; the exposure they produce is quantified in the following subsections.

Table 2

Prioritized risks and their inherent quantitative parameters (pre-mitigation)

ID	Risk (cause → event → effect)	Category	P × I (zone)	*p* (%)	E[*S*] (d)	VE (d)
R1	Legacy material master with duplicate codes and inconsistent units → data cleansing overruns → migration delay	Technical	4 × 4 = 16 (R)	70	16.3	11.4
R2	Duplicate or incomplete customer and vendor records → business partner conversion errors → data rework	Technical	4 × 3 = 12 (O)	70	10.8	7.6
R3	Low availability of key users → incomplete fit-to-standard workshops → late solution design	Client	3 × 4 = 12 (O)	50	16.3	8.1
R4	Changes in pricing and discount policy by channel → pricing reconfiguration → sprint delays	Client	3 × 3 = 9 (O)	50	10.8	5.4
R5	Defects in integrated order-to-cash and plan-to-produce scenarios → extra test cycles → go-live delay	Technical	4 × 5 = 20 (R)	70	27.1	19.0
R7	Interfaces with distributor portal, warehouse system and banks → failed integration tests → interface redesign	Technical	3 × 3 = 9 (O)	50	10.8	5.4
R9	Custom code of the legacy ERP → longer adaptation than estimated → delay of the Realize phase	Technical	4 × 4 = 16 (R)	70	16.3	11.4
R10	Partner add-on for tax localization not certified on time → electronic invoicing tests blocked	Supplier	3 × 5 = 15 (O)	50	27.1	13.5
R11	Mandatory material ledger and actual costing → inventory valuation changes → extra reconciliation	Technical	3 × 4 = 12 (O)	50	16.3	8.1
R12	Turnover of key consultants → learning curve of replacements → lower sprint productivity	Resources	3 × 3 = 9 (O)	50	10.8	5.4
R13	Batch, quality characteristics and certificates of analysis → gaps in the standard → extra extensions	Technical	3 × 4 = 12 (O)	50	16.3	8.1

ID	Risk (cause → event → effect)	Category	P × I (zone)	*p* (%)	E[*S*] (d)	VE (d)
R15	Scheduled power outages in Ecuador → lost working days → delays in test and infrastructure preparation	External	3 × 3 = 9 (O)	50	10.8	5.4
R18	Underestimated commercial reporting scope → growing gap backlog → additional sprints	Project mgmt.	3 × 3 = 9 (O)	50	10.8	5.4

Note. Zone: R = red, O = orange. *p* from Equation (1); E[*S*] is the mean delay if the risk occurs, Equation (3); VE = *p* · E[*S*], in working days.

Table 3

Response plan and residual parameters (post-mitigation)

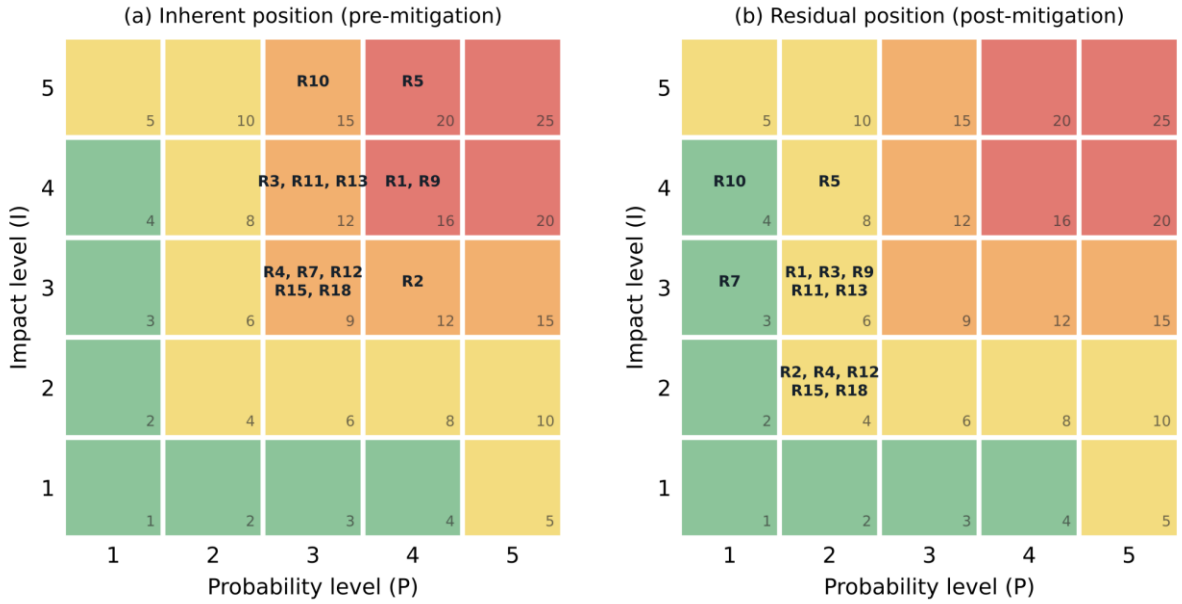
ID	Response (strategy: mitigate)	Cost (d)	Cost (USD k)	Residual P × I	p_{res} (%)	VE_{res} (d)	Level reduction (%)
R1	Early cleansing at source with quality rules per product family and data owners per area	5	12.0	2 × 3 = 6	35	3.0	62.5
R2	Business partner synchronization in the sandbox with duplicate removal and tax ID validation	2	4.0	2 × 2 = 4	30	1.6	66.7
R3	Key-user dedication agreement signed by the sponsor, operational backups, frozen workshop calendar	2	—	2 × 3 = 6	25	2.7	50.0
R4	Change control board and freeze of the commercial policy per sprint	1	—	2 × 2 = 4	25	1.6	55.6
R5	Test automation per sprint, test data with real batches and entry criteria for acceptance tests	5	18.0	2 × 4 = 8	40	6.5	60.0
R7	Early proof of concept of critical interfaces and signed interface contracts	3	7.0	1 × 3 = 3	20	1.3	66.7
R9	Removal of unused custom code before conversion; adaptation of critical objects only	3	9.0	2 × 3 = 6	35	3.0	62.5
R10	Early proof of concept of the add-on and contract clause on certification with penalty	2	5.0	1 × 4 = 4	20	2.2	73.3
R11	Inventory conversion simulated in the sandbox and reconciled with accounting	2	—	2 × 3 = 6	25	2.7	50.0
R12	Succession plan with a shadow consultant per module and retention bonus	1	8.0	2 × 2 = 4	20	1.1	55.6
R13	Gap validation against industry best practices; extensions that keep a clean core	2	—	2 × 3 = 6	25	2.7	50.0
R15	Cloud environments, generator and UPS for the project room, remote work during outages	1	15.0	2 × 2 = 4	20	1.3	55.6

ID	Response (strategy: mitigate)	Cost (d)	Cost (USD k)	Residual P × I	p_{res} (%)	VE_{res} (d)	Level reduction (%)
R18	MoSCoW prioritization; advanced analytics deferred to a second wave	1	—	2 × 2 = 4	25	1.4	55.6
Total		**30**	**78.0**			**31.1**	

Note. Residual delays keep the shape of Equation (2) with the most likely value set by the team. Level reduction = (inherent level − residual level) / inherent level. The sum of expected values fell from 114.3 to 31.1 days; this sum is a ranking aid, not a schedule forecast, because only delays on the driving path reach the finish.

Figure 2

Position of the 13 prioritized risks in the probability–impact matrix before and after the responses



Note. (a) Inherent levels before the responses. (b) Residual levels after the responses. Small numbers show the risk level P × I of each cell.

Distributions of Project Duration and Cost

Figure 3 shows the simulated distributions of total duration and total cost in the three scenarios, and Table 4 summarizes them. Duration uncertainty alone (base scenario) produced a mean of 445.1 working days with a standard deviation of 10.3 days; the probability of finishing within the deterministic plan of 423 days was only 1.2%, and the probability of staying within the budget was 0.02%. The plan sits in the far left tail because the 75/100/150% distributions are right-skewed and because, at each merge point, the longest of several parallel paths drives the finish in every iteration.

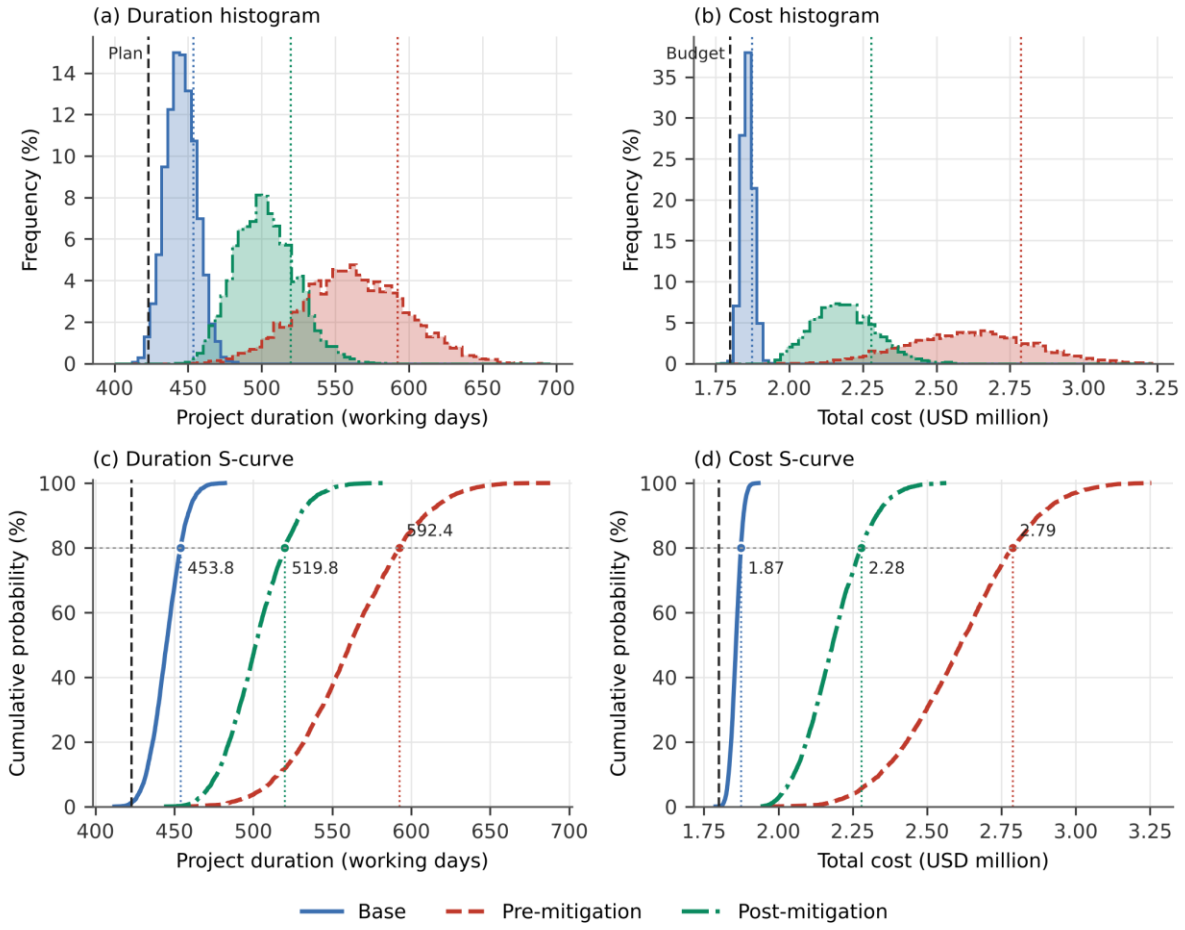
The 13 inherent risks shifted and widened both distributions. The mean duration rose to 562.1 days and its standard deviation to 35.6 days (coefficient of variation 6.3%); the 80% interval between P10 and P90 spanned 92.4 days, from 516.2 to 608.6 days. The mean cost reached USD 2.61 million with a

standard deviation of USD 206,168. No iteration finished within the plan or the budget. After the responses, the distributions moved left and narrowed: the mean duration fell to 502.6 days and its standard deviation to 20.5 days (−42.4%), and the standard deviation of cost fell from USD 206,168 to USD 106,678 (−48.3%). The post-mitigation duration became more right-skewed (skewness 0.29 against 0.08), because the residual risks occur less often but keep long delays when they do.

The S-curves in Figure 3c and 3d give the confidence attached to any commitment. At 80% confidence, the duration fell from 592.4 working days (5 November 2026) before mitigation to 519.8 days (27 July 2026) after it, and the cost from USD 2.79 million to USD 2.28 million. The schedule reserve over the deterministic plan therefore dropped from 169.4 to 96.8 days (−42.9%), and the cost reserve over the budget from USD 988,006 to USD 479,466 (−51.5%).

Figure 3

Simulated distributions of total project duration and total cost



Note. (a) and (b) Histograms. (c) and (d) Cumulative distributions (S-curves) with the 80th percentile marked. Dashed black lines show the deterministic plan (423 working days) and the approved budget (USD 1.80 million). 5,000 iterations, seed 20240730.

Table 4

Descriptive statistics of the simulated project duration and total cost by scenario

Statistic	Base	Pre-mitigation	Post-mitigation	Effect (post − pre)
Duration (working days)				
Mean (standard deviation)	445.1 (10.3)	562.1 (35.6)	502.6 (20.5)	−59.5 (−42.4% SD)
Coefficient of variation; skewness	2.3%; 0.11	6.3%; 0.08	4.1%; 0.29	—
P10 / P50 / P90	432.1 / 445.0 / 458.5	516.2 / 561.1 / 608.6	477.0 / 501.4 / 529.5	—
P80 (finish date)	453.8 (24 Apr 2026)	592.4 (5 Nov 2026)	519.8 (27 Jul 2026)	−72.7
Schedule reserve: P80 − 423 d	30.8	169.4	96.8	−72.7 (−42.9%)
Probability of meeting the plan	1.2%	0.0%	0.0%	—
Total cost (USD)				
Mean (standard deviation)	1,857,951 (19,340)	2,611,709 (206,168)	2,187,877 (106,678)	−423,832 (−48.3% SD)
Coefficient of variation; skewness	1.0%; 0.12	7.9%; 0.03	4.9%; 0.32	—
P10 / P50 / P90 (USD million)	1.83 / 1.86 / 1.88	2.34 / 2.61 / 2.88	2.05 / 2.18 / 2.33	—
P80	1,874,508	2,788,006	2,279,466	−508,539
Cost reserve: P80 − budget	74,508	988,006	479,466	−508,539 (−51.5%)
Probability of meeting the budget	0.02%	0.0%	0.0%	—
Joint behavior				
Pearson correlation, duration–cost	0.92	0.89	0.88	—
P(duration ≤ P80 and cost ≤ P80)	75.6%	74.6%	74.3%	—

Statistic	Base	Pre-mitigation	Post-mitigation	Effect (post − pre)
Joint 80% commitment (date; USD)	28 Apr 2026; 1,878,000	13 Nov 2026; 2,828,000	3 Aug 2026; 2,300,000	—

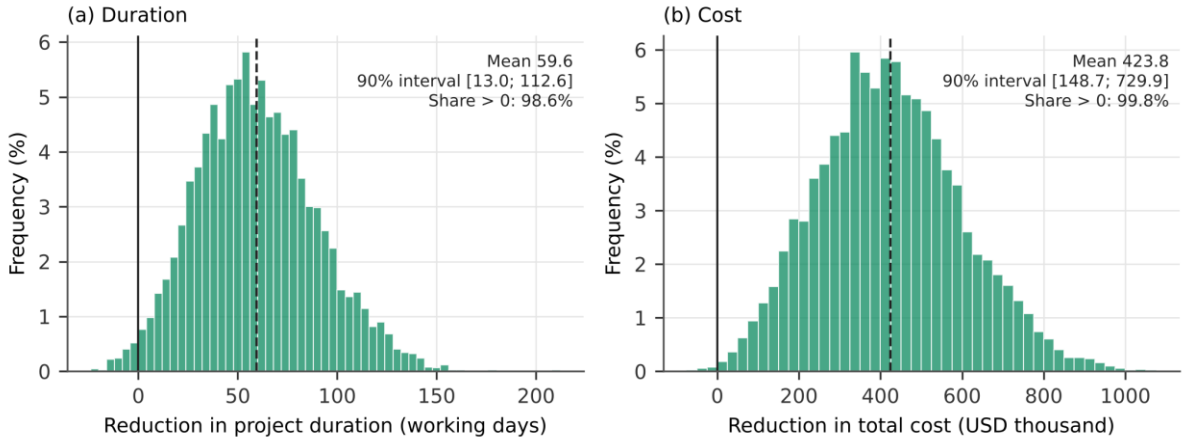
Note. Deterministic plan: 423 working days (12 March 2026); budget: USD 1,800,000. Effects are computed from unrounded values. The joint commitment is the smallest pair of equal marginal percentiles met together in at least 80% of the iterations.

Effect of the Responses on Each Iteration

Because the scenarios share random numbers, each iteration yields a paired observation of the same project with and without responses. Figure 4 shows the distribution of these paired reductions. The responses shortened the project in 98.6% of the iterations, by 59.6 working days on average, with a 90% interval from 13.0 to 112.6 days, and lowered the total cost in 99.8% of them, by USD 423,832 on average (90% interval USD 148,703 to 729,852). The Wilcoxon signed-rank test rejected a null median difference for both duration and cost ($p < 0.001$). The reduction of the P80 (72.7 days) exceeds the mean paired reduction (59.6 days) because the responses cut the right tail more than the center: the iterations in which several inherent risks coincide are the ones that change most. Relative to the investment, each day spent on responses removed 2.4 days of schedule reserve at P80, and each dollar removed about USD 6.5 of cost reserve.

Figure 4

Paired reduction produced by the responses in each iteration



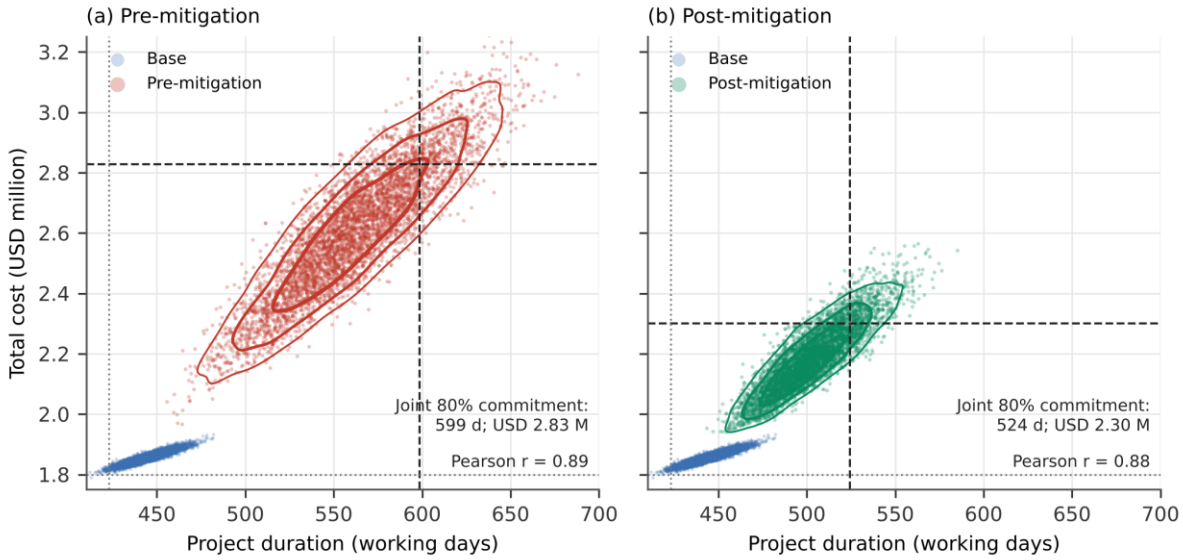
Note. (a) Project duration. (b) Total cost. Each value is the pre-mitigation minus the post-mitigation outcome of one of the 5,000 iterations; the dashed line marks the mean, and values above zero mean that the responses improved the outcome.

Joint Confidence in Time and Cost

Duration and cost were strongly correlated in every scenario (Pearson *r* between 0.88 and 0.92), because daily overhead and risk costs grow with delay (Figure 5). Correlation is not perfect, however, and the consequence matters for commitments: a sponsor who commits the P80 duration and the P80 cost separately obtains a joint confidence of only 74.6% before mitigation and 74.3% after it (Table 4). Reaching a true 80% joint confidence required moving each marginal to its 84.5th percentile before mitigation and to its 84.7th percentile after it. The resulting commitments were 13 November 2026 and USD 2.83 million before mitigation, and 3 August 2026 and USD 2.30 million after it. The responses shrank the 80% joint-probability region, the middle contour in Figure 5, along both axes and moved it towards the plan, although no part of it reaches the plan or the budget.

Figure 5

Bivariate distribution of project duration and total cost



Note. (a) Pre-mitigation and (b) post-mitigation scenarios, each shown against the base scenario. Contours are the highest-density regions containing 50% (thick), 80% (medium) and 95% (thin) of the iterations, estimated with a Gaussian kernel. Dashed lines mark the joint 80% commitment; dotted lines mark the plan and the budget.

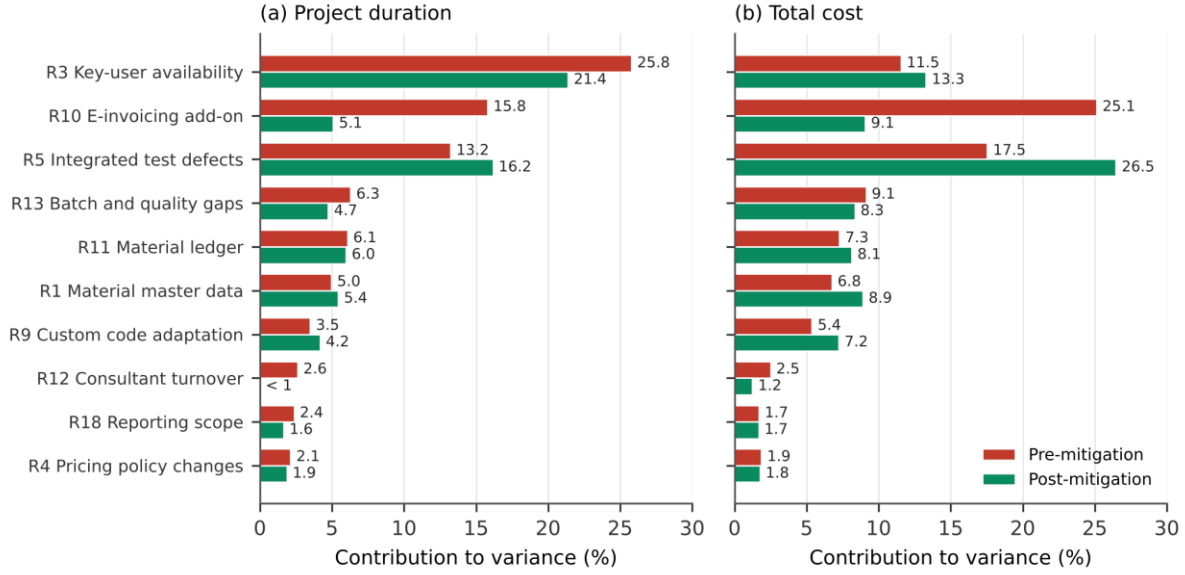
Sensitivity of Duration and Cost

Figure 6 decomposes the variance of duration and cost into the contributions of the risks and the task durations, Equation (5). The ranking differs from the expected-value ranking of Table 2. R3, the low availability of key users, explained 25.8% of the variance of duration before mitigation and still 21.4% after it, although its expected value ranked only fifth. Its delay falls on the fit-to-standard workshops and the design activity, both critical and both 30 days long, so every day it adds reaches the finish. R10

explained 15.8% of the duration variance and 25.1% of the cost variance before mitigation; its response, which lowered the probability to 20%, cut these shares to 5.1% and 9.1%, the largest reduction of any risk. R5 kept a high share after mitigation (16.2% of duration and 26.5% of cost), because its residual impact remained large. After mitigation, three task durations entered the top 15 contributors for the first time (sandbox setup, integrated testing and solution design, about 3–4% each), a sign that the residual exposure is shifting from discrete risks to the normal variability of the schedule.

Figure 6

Contribution of the main risks to the variance of project duration and total cost



Note. (a) Project duration. (b) Total cost. The ten risks with the highest pre-mitigation contribution to duration are shown before and after the responses. Contributions are normalized squared Spearman rank correlations; "< 1" indicates that the risk fell outside the 15 largest contributors.

Verification and Contrast with the Outcome

The simulated occurrence frequencies agreed with the assigned probabilities in both scenarios (chi-square with 13 degrees of freedom: 19.94, $p = 0.097$ before mitigation; 8.74, $p = 0.792$ after it), so the engine reproduced its inputs without detectable bias.

The project went live on 27 April 2026 at a cost of approximately USD 2.5 million, and the planned responses were executed. Three risks materialized: R1, R5 and R9, precisely the three red-zone risks. Table 5 summarizes the evaluation of the probabilities. The register discriminated well: the risks that occurred had received a mean inherent probability of 70.0%, against 44.1% for those that did not. Calibration was weaker. Over the 20 risks, neither the inherent probabilities (Brier score 0.198) nor the probabilities in force after the executed responses (0.136) improved on the base-rate forecast (0.128),

because the register assigned high probabilities to many risks that did not occur. Only within the 13 prioritized risks did the residual probabilities outperform the base rate (0.136 versus 0.178). With three events, these scores are indicative.

Table 5

Discrimination and calibration of the register probabilities against the outcome

Measure	Inherent probabilities	After executed responses	Base-rate forecast
Mean probability: occurred / not occurred (20 risks)	70.0% / 44.1%	36.7% / 27.4%	—
Brier score, 20 risks (3 occurred)	0.198	0.136	0.128 ($\bar{o}$ = 15%)
Brier score, 13 prioritized risks (3 occurred)	0.232	0.136	0.178 ($\bar{o}$ = 23.1%)

Note. "After executed responses" uses the residual probability for the 13 prioritized risks and the inherent probability for the other seven. A Brier score of 0.25 corresponds to an uninformative forecast of 50%.

The outcome can also be located on the simulated distributions. The go-live came 66 working days after the cutover planned for 23 January 2026 (working day 455 against day 389). Adding the 34 working days of hypercare, handover and closure that the plan places after the cutover gives an equivalent project duration of about 489 days. This value falls at the 27th percentile of the post-mitigation distribution and at the 2nd percentile of the pre-mitigation one: the post-mitigation forecast was conservative in time. In cost, the outcome went the other way. The USD 2.5 million spent at go-live, before hypercare and closure, lay at the 30th percentile of the pre-mitigation distribution but above the 99th percentile of the post-mitigation one, beyond its P80 of USD 2.28 million and its joint commitment of USD 2.30 million.

DISCUSSION

Regarding RQ1, the case shows that a register maintained under the SAP Activate methodology can feed a Monte Carlo model directly. Three decisions made this possible: a fixed equivalence between levels and parameters, which removes the need for new three-point estimates per risk; explicit links between each risk and the tasks it affects, which let the network decide how much of each delay reaches the finish; and common random numbers, which isolate the effect of the responses and turn each iteration into a paired comparison. The register remained the single source of risk data, so updating a level or a response updates the forecast. This answers the concern that quantitative analysis in IT projects is run apart from the register the team maintains (Kwak & Ingall, 2007), and it operationalizes

the integrated cost and schedule analysis recommended by practice standards (AACE International, 2011; Hulett, 2009).

Regarding RQ2, the responses roughly halved the reserves required at 80% confidence, at a cost of about 4% of the budget, and the paired analysis shows that this gain is not an artifact of two lucky samples: it held in more than 98% of the iterations. Three results go beyond what a qualitative matrix can show. First, the responses reduced dispersion as much as location, which is what narrows the reserve. Second, the sensitivity ranking disagreed with the expected-value ranking: R3 carried a modest expected value but dominated the variance because it falls on long critical tasks, and its response left most of that contribution in place. This is the kind of finding that justifies schedule-based simulation over register arithmetic (Hulett, 2009; Vanhoucke, 2010). Third, committing marginal P80 values separately delivers only about 74% joint confidence, so a sponsor who needs 80% for both time and cost must use the joint commitment. This point is rarely made explicit in practice, although cost and schedule guidance asks for confidence levels on both (AACE International, 2011; GAO, 2020).

Regarding RQ3, the evidence is mixed and informative. The register identified the right risks: the three that occurred were the three red ones. This pattern is common in elicited probabilities, which tend to rank events more reliably than they calibrate them (Gneiting & Raftery, 2007; Vose, 2008). The probabilities were too high for the risks that did not occur, which is the expected pattern of a cautious register and explains why calibration did not beat the base rate across the 20 risks. Two further results were unexpected. First, all three red risks occurred despite their responses, which suggests that the responses acted more on impact than on probability; the residual probabilities set by the team were optimistic. Second, the model was conservative in time, with the outcome at the 27th percentile, but optimistic in cost, with the outcome above the 99th percentile. The schedule bias has two plausible sources in the model: the days of each response were added to a task as if they consumed its calendar, although they can run in parallel, and risks were placed on tasks that already carried duration uncertainty, which can count part of the variability twice. The cost bias comes from Equation (4), which ties extra cost to extra time through a correlation near 0.9; in practice the firm recovered time by adding resources, an acceleration that raises cost while shortening the schedule and that the model does not represent.

Limitations and Further Research

The findings should be read within the boundaries of the design. The study is a single case: one transition from SAP ECC to SAP S/4HANA, in one lubricants manufacturer, led by one consulting firm. The schedule followed the standard SAP Activate roadmap rather than a plan built from scratch, and the register reflected the judgment of one project team. Other industries, implementation scenarios, such as a new implementation or a selective data transition, or organizations with different risk cultures may produce other rankings, other response returns and other gaps between forecast and outcome. The results therefore support the feasibility of the method, not a general estimate of its benefits.

The ex post verification rests on a small number of events. Only three of the 20 risks materialized, so the Brier scores and the discrimination measures are indicative rather than conclusive; a single additional event would have changed them noticeably. In addition, the outcome of the project is one realization of the uncertainty the model describes. A go-live at the 27th percentile of the post-mitigation distribution is consistent with a sound forecast, but it neither confirms nor refutes the model on its own. Assessing calibration with the rigor that proper scoring rules allow requires many forecasts and many outcomes (Gneiting & Raftery, 2007), which a single project cannot provide.

Several inputs depended on expert judgment. The residual probabilities and delays after each response were set by the team, not observed, and the ex post results suggest that they were optimistic: all three red-zone risks occurred despite their responses. The fixed equivalence between qualitative levels and simulation parameters, although agreed with the project director, imposes the same delay range on every risk of a given impact level, whatever the task it affects. These choices make the method fast to apply, but they transfer the known biases of elicited estimates, such as anchoring on the scale midpoints and overconfidence in the effect of responses, into the simulation (Vose, 2008).

The model structure adds further assumptions. Risks were sampled independently, although data quality, custom code adaptation and test defects are likely to share common causes in ERP projects (Aloini et al., 2007); positive correlation would widen the tails of both distributions and increase the reserves reported here. The days of each response were added to a task as if they consumed its calendar, although many responses run in parallel with the work, and risks were placed on tasks that already carried duration uncertainty, which may count part of the variability twice. Both effects push the

duration upward and may explain why the model was conservative in time. Conversely, Equation (4) ties extra cost to extra time and does not represent the acceleration that the firm applied in practice, adding resources to recover the schedule; this omission may explain why the model was optimistic in cost.

Finally, the schedule itself limits the analysis. The calendar ignored national holidays, which shifts the dates but not the working-day durations. The completed network still failed the high-float test of the DCMA assessment, because the solution adoption and project management workstreams run in parallel to the conversion work. Excess float can hide near-critical paths and understate the merge bias that drives part of the gap between the deterministic plan and the simulated distributions (GAO, 2015; Hulett, 2009).

These limitations point to five lines of further research. First, correlation between risks should be modeled explicitly, for example through shared risk drivers that act on several tasks at once, as recommended for integrated cost and schedule analysis (AACE International, 2011; Hulett, 2009). Second, the cost model should include the time–cost trade-off of acceleration, so that the simulation can represent a project that buys time with money. Third, the residual parameters should be treated as hypotheses to be updated: rerunning the model at each phase gate with the risks that have occurred, the responses that have been executed and the progress achieved would turn a planning tool into a monitoring tool. Fourth, applying the method across a portfolio of ERP projects would provide the outcomes needed to calibrate the level-to-parameter equivalences, in line with reference-class forecasting (Flyvbjerg, 2006). Fifth, comparative studies could test whether the paired pre- and post-mitigation design changes the decisions of steering committees, for instance the funding of responses or the commitment dates they accept, which is the practical value the method claims.

CONCLUSIONS

A qualitative risk register built under the SAP Activate methodology was converted into a reproducible Monte Carlo model of schedule and cost without collecting new data. In the SAP S/4HANA transition studied, the 13 prioritized risks raised the mean duration from 445.1 to 562.1 working days and the reserves required at 80% confidence to 169.4 days and USD 988,006. Thirteen responses costing 30 days and USD 78,000 reduced those reserves by 42.9% and 51.5%, cut the standard deviations of

duration and cost by 42.4% and 48.3%, and improved the outcome in more than 98% of the paired iterations. Separate P80 commitments for time and cost gave only about 74% joint confidence. The register identified the risks that occurred, but its absolute probabilities were not better calibrated than a base-rate forecast, and the model proved conservative in time and optimistic in cost.

In practice, the method gives steering committees distributions, reserves, joint commitments and the contribution of each risk, all from the same register the team already maintains, and it can be rerun at each phase gate to recalibrate. The ex post contrast shows that the main gains in accuracy will come from the cost model and the residual parameters, not from the probability scale. The consulting firm plans to apply the method to projects of greater complexity, which will provide the outcome data needed to calibrate it.

REFERENCES

AACE International. (2011). *Integrated cost and schedule risk analysis using Monte Carlo simulation of a CPM model* (Recommended Practice No. 57R-09). AACE International.

Aloini, D., Dulmin, R., & Mininno, V. (2007). Risk management in ERP project introduction: Review of the literature. *Information & Management, 44*(6), 547–567. https://doi.org/10.1016/j.im.2007.05.004

Armijos De La Cruz, B. A., Castro Esparza, J. R., Leoro Benítez, J. F., & Armijos De La Cruz, W. Y. (2025). Leveraging Monte Carlo simulation for project risk management: A study on pre-mitigation and post-mitigation techniques in risk registers. *Ciencia Latina Revista Científica Multidisciplinar, 9*(5), 6975–6993. https://doi.org/10.37811/cl_rcm.v9i5.20089

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review, 78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Davis, R. (2008). Teaching project simulation in Excel using PERT-beta distributions. *INFORMS Transactions on Education, 8*(3), 139–148. https://doi.org/10.1287/ited.1080.0013

Defense Contract Management Agency. (2012). *Earned value management system (EVMS) program analysis pamphlet* (DCMA-EA PAM 200.1). Defense Contract Management Agency.

Flyvbjerg, B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal, 37*(3), 5–15. https://doi.org/10.1177/875697280603700302

Flyvbjerg, B., & Budzier, A. (2011). Why your IT project may be riskier than you think. *Harvard Business Review, 89*(9), 23–25.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association, 102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Hulett, D. T. (2009). *Practical schedule risk analysis*. Gower.

International Organization for Standardization. (2018). *Risk management—Guidelines* (ISO Standard No. 31000:2018). https://www.iso.org/standard/65694.html

Kwak, Y. H., & Ingall, L. (2007). Exploring Monte Carlo simulation applications for project management. *Risk Management, 9*(1), 44–57. https://doi.org/10.1057/palgrave.rm.8250017

Malcolm, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W. (1959). Application of a technique for research and development program evaluation. *Operations Research, 7*(5), 646–669. https://doi.org/10.1287/opre.7.5.646

Marsaglia, G., & Tsang, W. W. (2000). A simple method for generating gamma variables. *ACM Transactions on Mathematical Software, 26*(3), 363–372. https://doi.org/10.1145/358407.358414

Project Management Institute. (2019). *The standard for risk management in portfolios, programs, and projects*. Project Management Institute.

Ram, J., Corkindale, D., & Wu, M.-L. (2013). Implementation critical success factors (CSFs) for ERP: Do they contribute to implementation success and post-implementation performance? *International Journal of Production Economics, 144*(1), 157–174. https://doi.org/10.1016/j.ijpe.2013.01.032

SAP SE. (2024). *Transition to SAP S/4HANA: SAP Activate roadmap* (Version 24Q3) [Roadmap and project plan]. SAP Activate Roadmap Viewer.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. John Wiley & Sons. https://doi.org/10.1002/9780470316849

U.S. Government Accountability Office. (2015). *Schedule assessment guide: Best practices for project schedules* (GAO-16-89G). https://www.gao.gov/products/gao-16-89g

U.S. Government Accountability Office. (2020). *Cost estimating and assessment guide: Best practices for developing and managing program costs* (GAO-20-195G). https://www.gao.gov/products/gao-20-195g

Vanhoucke, M. (2010). Using activity sensitivity and network topology information to monitor project time performance. *Omega, 38*(5), 359–370. https://doi.org/10.1016/j.omega.2009.09.001

Vose, D. (2008). *Risk analysis: A quantitative guide* (3rd ed.). John Wiley & Sons.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80–83. https://doi.org/10.2307/3001968