

Toxic and Abusive Leadership in Nursing and Workforce Outcomes: A Bayesian Multilevel Meta-Analysis and Meta-Analytic Structural Equation Modeling of Nurse Wellbeing, Retention, and Exploratory Care-Quality Evidence

Alice P. Lindström 

Sahlgrenska University Hospital, Gothenburg, Sweden

Email: parvaneh_shariat@hotmail.com

How to cite this paper: Lindström, A.P. (2026) Toxic and Abusive Leadership in Nursing and Workforce Outcomes: A Bayesian Multilevel Meta-Analysis and Meta-Analytic Structural Equation Modeling of Nurse Wellbeing, Retention, and Exploratory Care-Quality Evidence. *Open Journal of Nursing*, 16, 545-576. <https://doi.org/10.4236/ojn.2026.1610040>

Received: July 23, 2026

Accepted: October 17, 2026

Published: October 20, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0). <http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

Background: Toxic and abusive leadership in nursing are measurable organizational exposures, but quantitative evidence remains fragmented across constructs, instruments, outcomes, and settings. **Objective:** To estimate the magnitude and heterogeneity of associations between toxic or abusive leadership and nursing workforce outcomes and examine modeled indirect pathways to turnover intention. **Methods:** This PRISMA-guided Bayesian multilevel meta-analysis and meta-analytic structural equation modeling study identified 1,091 records. Twenty-three studies were retained; 20 contributed 48 primary effect sizes and 17 contributed 39 correlations to MASEM. Correlations were harmonized so positive values indicated more adverse outcomes and transformed using Fisher's (z). Robustness was assessed using posterior predictive checks, alternative priors, and sensitivity analyses. **Results:** Toxic or abusive leadership was associated with poorer wellbeing ($r = 0.34$; 95% CrI, 0.23 - 0.45) and retention-related outcomes ($r = 0.33$; 95% CrI, 0.19 - 0.46). Prediction intervals were wide. The just-identified MASEM showed a direct modeled association with turnover intention (0.311; 95% CI, 0.239 - 0.383) and an indirect association through psychological distress (0.035; 95% CI, 0.015 - 0.056). Twenty-two of 23 studies had serious overall risk of bias; care-quality evidence comprised one effect. **Conclusions:** Toxic and abusive leadership are associated with poorer nurse wellbeing and retention-related outcomes. Structural pathways are associational rather than causal, and care-quality evidence remains exploratory.

Keywords

Toxic Leadership, Abusive Supervision, Nursing Workforce, Turnover Intention, Bayesian Meta-Analysis

1. Introduction

Toxic and abusive leadership in nursing are increasingly understood as measurable organizational exposures rather than isolated interpersonal conflicts. Toxic leadership refers to sustained managerial behaviours characterized by intimidation, manipulation, hostility, authoritarian control, narcissistic or self-serving conduct, and erosion of employees' dignity and organizational functioning. Abusive supervision, more specifically, describes nurses' perceptions of persistent hostile verbal or non-verbal supervisory behaviour in the absence of physical contact [1] [2]. Although these constructs overlap in their adverse orientation, they are not interchangeable. Toxic leadership encompasses a broader constellation of destructive managerial behaviours, whereas abusive supervision focuses more specifically on sustained hostile supervisory conduct. Related constructs such as despotic, exploitative, narcissistic, and other destructive leadership forms emphasize different mechanisms, including authoritarianism, self-interest, exploitation, or broader organizational harm.

These behaviours are particularly relevant to nursing because care is delivered in environments characterized by clinical acuity, emotional labour, hierarchical decision-making, interprofessional dependence, and continuous responsibility for patient surveillance. In such settings, leadership may influence nurses' psychological resources, organizational attachment, voice behaviour, and intentions to remain in practice. A growing nursing literature has associated toxic, abusive, despotic, and other destructive leadership behaviours with emotional exhaustion, burnout, psychological distress, workplace deviance, organizational cynicism, job insecurity, workplace bullying, workplace gaslighting, reduced job satisfaction, and diminished organizational commitment [3]-[8]. These workforce consequences are relevant to contemporary nursing shortages because retention is shaped not only by individual and labour-market factors but also by organizational conditions that may be modifiable. Prior syntheses have shown that leadership style, organizational support, horizontal violence, unfinished nursing care, burnout, and job satisfaction are associated with nurses' intentions to leave or remain in their positions [9]-[13]. However, those broader literatures do not establish the pooled magnitude of associations specifically involving toxic, abusive, and closely related destructive leadership exposures, nor do they resolve whether such associations are similar across wellbeing and retention outcomes.

The potential consequences of destructive leadership may extend beyond workforce outcomes. Negative leadership has been associated with nurses' silence and constrained voice, while toxic nurse-manager behaviours have been linked con-

ceptually and empirically with psychological safety, adverse-event reporting, escalation behaviour, and perceived care quality [14] [15]. Broader nursing evidence also links work environments, burnout, and engagement with patient-safety culture, care left undone, patient satisfaction, and quality-of-care indicators [16]-[21]. These findings make a workforce-to-care pathway theoretically plausible, but they do not establish that toxic or abusive leadership directly causes poorer care. In particular, direct quantitative evidence connecting destructive leadership with care-quality outcomes remains sparse. A plausible workforce pathway is that toxic or abusive leadership is associated with greater psychological distress and lower job satisfaction, which may in turn be associated with stronger turnover intention. Such a pathway can be examined statistically using pooled correlation structures, but observational correlations cannot establish temporal sequence or causal mediation. Accordingly, meta-analytic structural equation modeling can be used to summarize whether the available correlation network is consistent with specified direct and indirect associations rather than to demonstrate a causal mechanism.

Despite the growth of this literature, evidence remains fragmented across leadership labels, countries, clinical settings, instruments, study designs, and outcome definitions. Recent reviews have mapped abusive supervision, destructive leadership, cultural variation in negative leadership, toxic-leadership prevalence among nurses outside the United States, and workplace bullying [2] [6] [22]-[25]. However, an integrated quantitative synthesis is still needed to estimate the average association while accounting for multiple effect sizes contributed by the same studies, quantify heterogeneity across the evidence base, and test the robustness of conclusions to alternative analytic assumptions. This study therefore aimed to synthesize quantitative evidence on toxic and abusive leadership in nursing using Bayesian multilevel meta-analysis and, where the available correlation structure permitted, meta-analytic structural equation modeling. Toxic, abusive, and closely related leader-originated destructive constructs were synthesized as a broader adverse-leadership family while retaining their conceptual distinctions through moderator and sensitivity analyses. Nurse wellbeing and retention-related outcomes constituted the primary analytic focus; direct care-quality evidence was treated as secondary and exploratory. The primary research question was: what are the magnitude and heterogeneity of the associations between toxic or abusive leadership and nurse wellbeing and retention-related outcomes? The secondary question was: does the available pooled correlation structure support modeled direct and indirect workforce associations linking toxic or abusive leadership with turnover intention, and what direct quantitative evidence is available for care-quality outcomes?

2. Methods

2.1. Study Design and Search Strategy

This study was designed as a PRISMA-guided quantitative evidence synthesis us-

ing Bayesian multilevel meta-analysis and, where the available correlation structures permitted, meta-analytic structural equation modeling (MASEM). The objective was to synthesize quantitative evidence on toxic and abusive leadership in nursing and its associations with nurse wellbeing, retention-related outcomes, and care-quality outcomes where directly reported. Wellbeing and retention-related outcomes constituted the primary analytic domains, whereas direct care-quality evidence was treated as secondary and exploratory because of its limited availability. A structured search strategy was developed to capture three conceptual domains: the nursing population, destructive leadership exposures, and workforce or care-related outcomes. The search syntax was first developed for PubMed/MEDLINE and then adapted for CINAHL, Scopus, Web of Science, PsycINFO, and Embase. The core search structure combined population terms with leadership-exposure and outcome terms as follows:

(“nurse” OR “nurses” OR “registered nurses” OR “staff nurses” OR “nursing staff” OR “nurse manager”) AND (“toxic leadership” OR “abusive supervision” OR “destructive leadership” OR “despotic leadership” OR “narcissistic leadership” OR “exploitative leadership” OR “tyrannical leadership” OR “negative leadership”) AND (“burnout” OR “emotional exhaustion” OR “psychological distress” OR “stress” OR “job satisfaction” OR “organizational commitment” OR “work engagement” OR “turnover intention” OR “intention to leave” OR “absenteeism” OR “quality of care” OR “patient safety” OR “psychological safety” OR “silence”).

Reference lists of eligible articles and relevant reviews were also screened manually to identify additional records. Database-specific search syntax was adapted to the indexing and field structure of each source, and the corresponding search dates, language and publication-date limits, and protocol-registration information were retained with the search documentation. Complete database-specific strategies should be reported with the supplementary search record to preserve reproducibility without reproducing lengthy search strings in the main text.

2.2. Eligibility Criteria and Study Selection

Studies were eligible if they included nurses or nursing personnel, measured toxic leadership, abusive supervision, or a closely related destructive leadership construct, and reported a quantitative association with at least one eligible workforce or care-related outcome. Toxic leadership and abusive supervision were the principal exposure categories. Destructive, despotic, exploitative, narcissistic, tyrannical, and other leader-originated negative leadership constructs were eligible when they explicitly represented adverse behavior originating from a leader or supervisor. These constructs were treated as conceptually related manifestations of destructive leadership rather than as psychometrically equivalent constructs; exposure category and measurement instrument were therefore examined in moderator and sensitivity analyses. Positive leadership constructs were not reverse-coded and reclassified as destructive leadership. Outcomes included nurse wellbeing,

psychological distress, emotional exhaustion, burnout, stress, job satisfaction, organizational commitment, work engagement, absenteeism, turnover intention, intention to leave, psychological safety, silence, quality of care, or patient-safety-related outcomes. Direct nurse-assessed quality-of-care and patient-safety indicators were eligible for the secondary care-quality objective; organizational or care-adjacent outcomes were not classified as direct care-quality outcomes unless they explicitly measured care quality or patient safety. Studies were excluded if the exposure was not leadership-related, the population was not nursing, the outcome was outside the review scope, the publication type or design was ineligible, or no extractable quantitative effect size was available. Duplicate or overlapping samples were retained only once in the primary dataset.

Titles and abstracts were screened against the eligibility criteria after duplicate removal, followed by full-text eligibility assessment. Screening and eligibility judgments were checked against the same prespecified criteria, and discrepant eligibility decisions were resolved through consensus. The study-selection record retained the decisions and reasons for exclusion used in constructing the final evidence dataset. The study-selection process followed the PRISMA framework and is summarized in (**Diagram 1**). Database searches identified 1,065 records, and 26 additional records were identified through manual searching and reference-list screening. After removal of 312 duplicates, 779 records were screened by title and abstract; 719 were excluded. Sixty reports were sought for full-text retrieval, of which three could not be retrieved. Fifty-seven full-text reports were assessed for eligibility, and 34 were excluded with documented reasons: wrong exposure ($n = 10$), wrong outcome ($n = 5$), wrong population ($n = 3$), wrong publication type or study design ($n = 6$), no extractable quantitative effect size ($n = 6$), duplicate or overlapping sample ($n = 3$), and unclear eligibility after full-text assessment ($n = 1$). Overall, 23 studies were retained in the final evidence dataset. Of these, 20 studies contributed 48 effect sizes to the primary meta-analysis, three studies were retained for sensitivity analyses only, and 17 studies contributed 39 correlations to the primary MASEM analysis.

2.3. Data Extraction and Effect-Size Preparation

All extracted data were organized in an R-ready database before analysis. Each study and effect size was assigned a unique identifier. The primary meta-analytic dataset comprised 48 eligible zero-order correlations from 20 studies: 27 wellbeing effects, 20 retention-related effects, and one direct care-quality effect. When a total toxic or abusive leadership score and corresponding subscale estimates were available for the same outcome, the total-score estimate was preferentially retained in the primary dataset, with eligible subscale estimates reserved for sensitivity analysis. The same principle was applied to total and subscale outcome measures to reduce redundant representation of the same sample. Correlations were directionally harmonized so that positive values represented more adverse outcomes. Accordingly, correlations involving positively oriented outcomes, such as job satis-

faction, organizational commitment, work engagement, personal accomplishment, or care quality, were directionally recoded when required to maintain a common adverse orientation.

Pearson correlations were transformed to Fisher's (z):

$$z = a \tanh(r)$$

with sampling variance:

$$v = \frac{1}{N-3}$$

and standard error:

$$SE = \sqrt{v}$$

Pooled estimates were subsequently back-transformed to the correlation scale using:

$$r = \tanh(z).$$

Data-audit scripts in R were used to verify study identifiers, duplicate entries, missing values, domain and subdomain coding, effect-size eligibility, and consistency between the primary and sensitivity datasets.

2.4. Quality Appraisal and Risk-of-Bias Assessment

The methodological quality and risk of bias of the included studies were independently assessed by two reviewers using a prespecified, domain-based assessment framework appropriate to the predominantly observational evidence base. The appraisal process was informed by the Joanna Briggs Institute (JBI) Critical Appraisal tools, which provide structured guidance for evaluating the methodological quality and risk of bias across different study designs [26] [27]. The assessment framework was adapted to the methodological characteristics of the included studies while maintaining consistent predefined criteria across studies.

Risk of bias was evaluated using a prespecified domain-based assessment framework appropriate to the predominantly observational evidence base. Seven domains were assessed: confounding, selection, exposure measurement, outcome measurement, missing data, selective reporting, and temporal ambiguity. Each domain was classified as Low, Moderate, Serious, or Unclear according to the predefined coding criteria. Overall judgments were derived from the pattern and severity of the domain-level assessments using the prespecified review rule. Assessments were completed at study level and checked against the same criteria, with discrepant domain or overall judgments reconciled through consensus. Risk-of-bias judgments were used to qualify interpretation of pooled associations and to inform sensitivity analyses where analytically applicable.

2.5. Bayesian Multilevel Meta-Analysis

Bayesian multilevel meta-analysis was performed on harmonized Fisher-(z) effect sizes using the brms package (version 2.23.0) with the rstan backend (version

2.32.7; StanHeaders version 2.32.10). The primary model was specified as:

$$y_i | \mu_i, \sigma \sim N(\mu_i, SE_i^2 + \sigma^2),$$

where y_i represented the harmonized Fisher- z effect estimate and se_i its known sampling standard error. A Gaussian likelihood with identity link was used. Known sampling uncertainty was incorporated through the effect-specific standard errors, while an additional residual effect-level standard deviation (σ) was estimated. Dependence among multiple effect sizes arising from the same study was modeled through a study-level random intercept. The pooled intercept was assigned a Normal (0, 0.5) prior on the Fisher- z scale. Exponential (2) priors were assigned separately to the study-level random-effect standard deviation (τ) and residual standard deviation (σ). Four Markov chain Monte Carlo chains were run for 4000 iterations each, including 1000 warmup iterations per chain, yielding 12,000 post-warmup draws. Thinning was set to 1. Sampling used the No-U-Turn Sampler with `adapt_delta` = 0.99, `max_treedepth` = 15, and seed 20260609.

Convergence was evaluated using split, bulk and tail effective sample sizes, divergent transitions, and NUTS treedepth diagnostics. An $R < 1.01$ was considered indicative of adequate chain mixing. Study-level heterogeneity was summarized by the posterior distributions of τ and τ^2 , whereas residual effect-level heterogeneity was summarized separately using σ and σ^2 . Posterior medians and 95% credible intervals were used for heterogeneity parameters, and prediction intervals were calculated to characterize the expected range of underlying associations in future comparable settings. Domain-specific Bayesian models were estimated for wellbeing and retention-related outcomes. Care-quality evidence was summarized descriptively when the number of primary effects was insufficient for pooled modeling. Because only one primary direct care-quality effect was available, no independent pooled care-quality model was estimated.

2.6. Posterior Predictive Assessment

Posterior predictive adequacy was evaluated using 2000 replicated datasets generated from the posterior predictive distribution of the primary Bayesian model. Model-data correspondence was examined graphically using density-overlay and empirical cumulative distribution function plots and numerically by comparing the observed mean, standard deviation, median, minimum, and maximum with their corresponding posterior predictive distributions. These checks were used to assess model-data compatibility rather than as formal tests of model validity.

2.7. Prior Sensitivity Analysis

Robustness to prior specification was examined by refitting the primary Bayesian model under two alternative prior sets while retaining the same data, likelihood, hierarchical structure, and sampling configuration. A more regularizing specification assigned Normal (0, 0.3) to the pooled intercept and Exponential (4) priors to τ and σ . A more diffuse specification assigned Normal (0, 1) to the intercept

and Exponential (1) priors to both standard deviations. Pooled associations and posterior heterogeneity estimates were compared with those from the baseline specification.

2.8. Moderator and Analytic Sensitivity Analyses

Moderator analyses examined outcome domain and subdomain, exposure category, exposure instrument, and geographical region. Moderator contrasts were interpreted as differences between model-estimated pooled associations across groups rather than as direct within-study head-to-head comparisons. Analytic sensitivity analyses examined the stability of pooled estimates under alternative analytic decisions, including primary effects only, an expanded primary-plus-sensitivity dataset, restriction to toxic leadership, broader destructive-leadership definitions, workforce outcomes only, inclusion of direct care-quality evidence, and restriction to zero-order correlations. Influence diagnostics were also examined to identify individual studies or effects with disproportionate influence on pooled estimates.

2.9. Small-Study-Effect and Funnel-Asymmetry Analyses

Because several studies contributed multiple correlated effects, funnel-asymmetry analyses were not performed directly on the 48 individual primary effects. Instead, effects were first aggregated to one Fisher z estimate per study, yielding 20 study-level estimates. Under the conventional independence assumption, study-level estimates were calculated by inverse-variance weighting:

$$z_j = \frac{\sum_i \frac{z_{ij}}{v_{ij}}}{\sum_i \frac{1}{v_{ij}}}$$

with variance:

$$\text{Var}(z_j) = \frac{1}{\sum_i \frac{1}{v_{ij}}}$$

The resulting study-level effects were analyzed using random-effects models estimated by restricted maximum likelihood. Funnel asymmetry was evaluated using weighted regression with multiplicative dispersion and the study-level standard error as the predictor. The asymmetry test was interpreted separately from the regression-based limit estimate. Funnel plots and trim-and-fill analyses were used as supplementary exploratory diagnostics. Because the precision of an aggregated study estimate depends on the covariance among multiple effects from the same study, within-study dependence was examined using generalized least-squares aggregation under assumed common correlations of ($\rho = 0.30$), 0.50, and 0.70. For two effects (i) and (k) originating from the same study, the off-diagonal sampling covariance was specified as:

$$\text{Cov}(y_i, y_k) = \rho \sqrt{v_i v_k}$$

The generalized least-squares study-level estimate was obtained as:

$$z_j = \frac{1^T V_j^{-1} Z_j}{1^T V_j^{-1} 1}$$

with sampling variance:

$$\text{Var} z_j = \frac{1}{1^T V_j^{-1} 1}$$

The random-effects and funnel-asymmetry analyses were repeated under each assumed within-study correlation. Because funnel asymmetry may arise from genuine heterogeneity, differences in outcomes or measurement, study characteristics, or small-study effects in addition to selective publication, these analyses were interpreted as exploratory diagnostics and not as definitive evidence of publication bias.

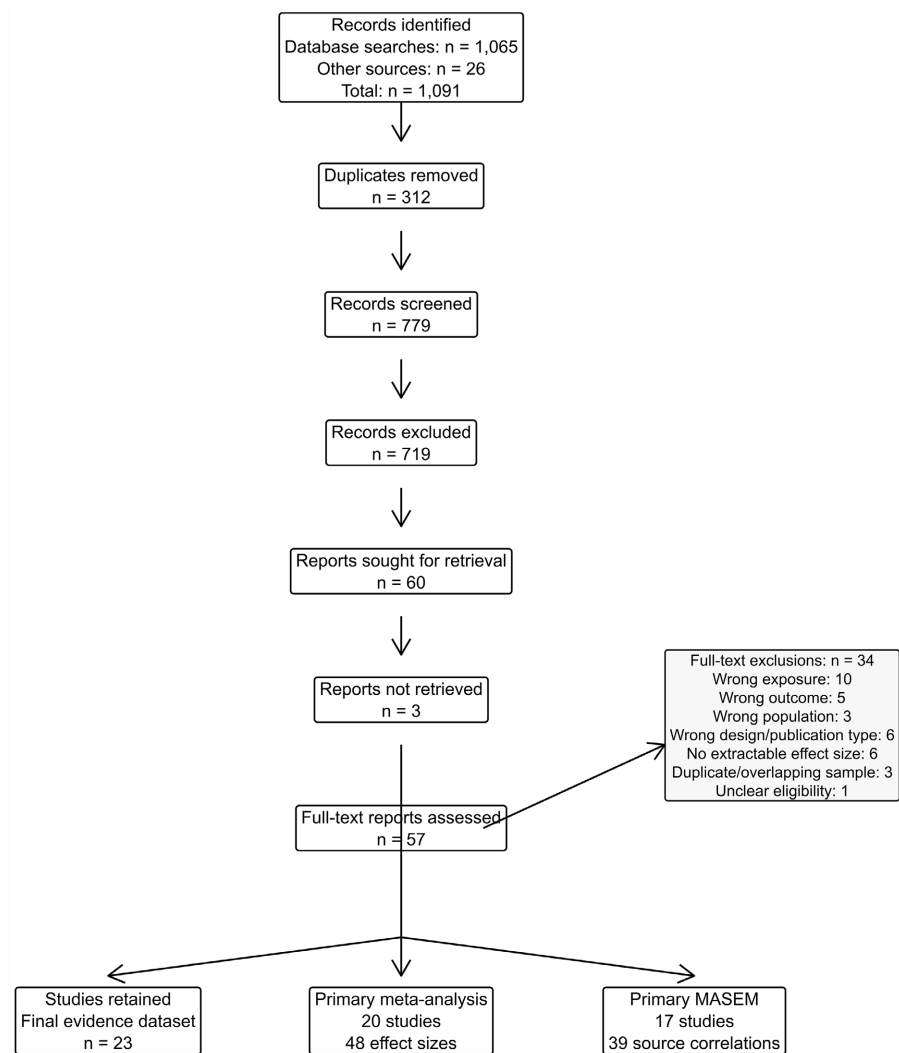
2.10. Meta-Analytic Structural Equation Modeling

Where sufficient pooled correlation structures were available, MASEM was used to examine workforce pathways linking toxic or abusive leadership with turnover intention. The primary MASEM dataset comprised 39 eligible zero-order source correlations from 17 studies, which yielded 18 distinct off-diagonal standardized correlation pairs after conceptual harmonization. The primary workforce model comprised four variables: toxic/abusive leadership, low job satisfaction, psychological distress, and turnover intention. The structural model specified paths from toxic/abusive leadership to low job satisfaction, psychological distress, and turnover intention and paths from low job satisfaction and psychological distress to turnover intention. The covariance between low job satisfaction and psychological distress was freely estimated. Indirect associations were calculated as products of the corresponding component paths. The resulting 4*4 pooled correlation matrix was positive definite, with a minimum eigenvalue of 0.5689, and required no positive-definiteness adjustment. The effective sample size for SEM estimation was set to (N = 768), corresponding to the rounded median sample size of the eight source correlations contributing to the most densely represented toxic/abusive leadership-turnover intention pair. The structural model was estimated in R using the lavaan package. Because the model was just-identified (df = 0), global fit indices such as CFI, TLI, RMSEA, and SRMR were not interpreted as evidence of substantive model fit. Interpretation instead focused on the magnitude, direction, and precision of the structural paths. Because the contributing evidence was predominantly observational and cross-sectional, indirect effects were interpreted as modeled indirect associations rather than evidence of causal mediation. Care-quality correlations were insufficient for confirmatory structural modeling and were therefore retained only as exploratory evidence. Statistical analyses were conducted in R. Bayesian multilevel models used brms version 2.23.0 with rstan ver-

sion 2.32.7 and StanHeaders version 2.32.10; study-level random-effects and funnel-asymmetry analyses were conducted using metafor, and MASEM was estimated using lavaan. Analysis scripts and structured datasets were retained to support reproducibility of the quantitative workflow. To document the structure of the retained evidence base, a randomly selected sample of 10 studies from the final dataset is summarized in **Table 1**. The table is illustrative rather than exhaustive; full bibliographic details for all studies included in the meta-analysis and sensitivity analyses are provided in the reference list and extraction database.

Table 1. Examples of retained studies and study specifications.

Study source	Country and setting	Design and sample	Leadership exposure/ instrument	Eligible outcomes or contribution
[28]	China; four tertiary general hospitals	Cross-sectional survey; 1127 registered nurses	Abusive supervision; Tepper Abusive Supervision Questionnaire	Turnover intention; primary meta-analysis and MASEM correlation data
[29]	Türkiye; public and university hospitals	Cross-sectional online survey; 168 nurses	Toxic leadership; Toxic Leadership Scale	Turnover intention; primary meta-analysis and sensitivity effects
[30]	Saudi Arabia; emergency departments in public hospitals	Cross-sectional study; 387 emergency nurses	Abusive supervision; Toxic Leadership Assessment	Organizational commitment; primary meta-analysis
[31]	Ghana; 12 public hospitals	Multicentre cross-sectional study; 943 nurses	Toxic leadership behaviour; ToxBH-NM scale	Job satisfaction and turnover intention; primary meta-analysis and MASEM
[32]	China; two public hospitals	Cross-sectional survey; 227 nurses	Abusive supervision; Tepper 15-item scale	Psychological distress; primary meta-analysis
[33]	Australia; five public acute hospitals	Cross-sectional survey with SEM; 250 nurses	Abusive supervision dimensions: task attacks, personal attacks, and isolation	Job satisfaction, psychological distress, and turnover intention; primary meta-analysis
[34]	China; four hospitals in Harbin	Cross-sectional quantitative survey; 702 registered nurses	Abusive supervision; Chinese version of Tepper scale	Work engagement; primary meta-analysis and MASEM
[35]	Pakistan; 15 government and private hospitals	Time-lagged self-report survey; 231 nurses	Exploitative leadership; Schmid Exploitative Leadership Scale	Psychological distress; primary meta-analysis
[36]	Philippines; 15 hospitals	Cross-sectional self-report survey; 770 nurses	Toxic leadership behaviour; ToxBH-NM scale	Job satisfaction, stress, absenteeism, and turnover intention; primary meta-analysis and MASEM
[37]	Philippines; seven acute-care hospitals	Cross-sectional descriptive survey; 770 clinical nurses	Abusive supervision; Abusive Supervision Scale	Stress, work engagement, absenteeism, and turnover intention; primary meta-analysis and MASEM

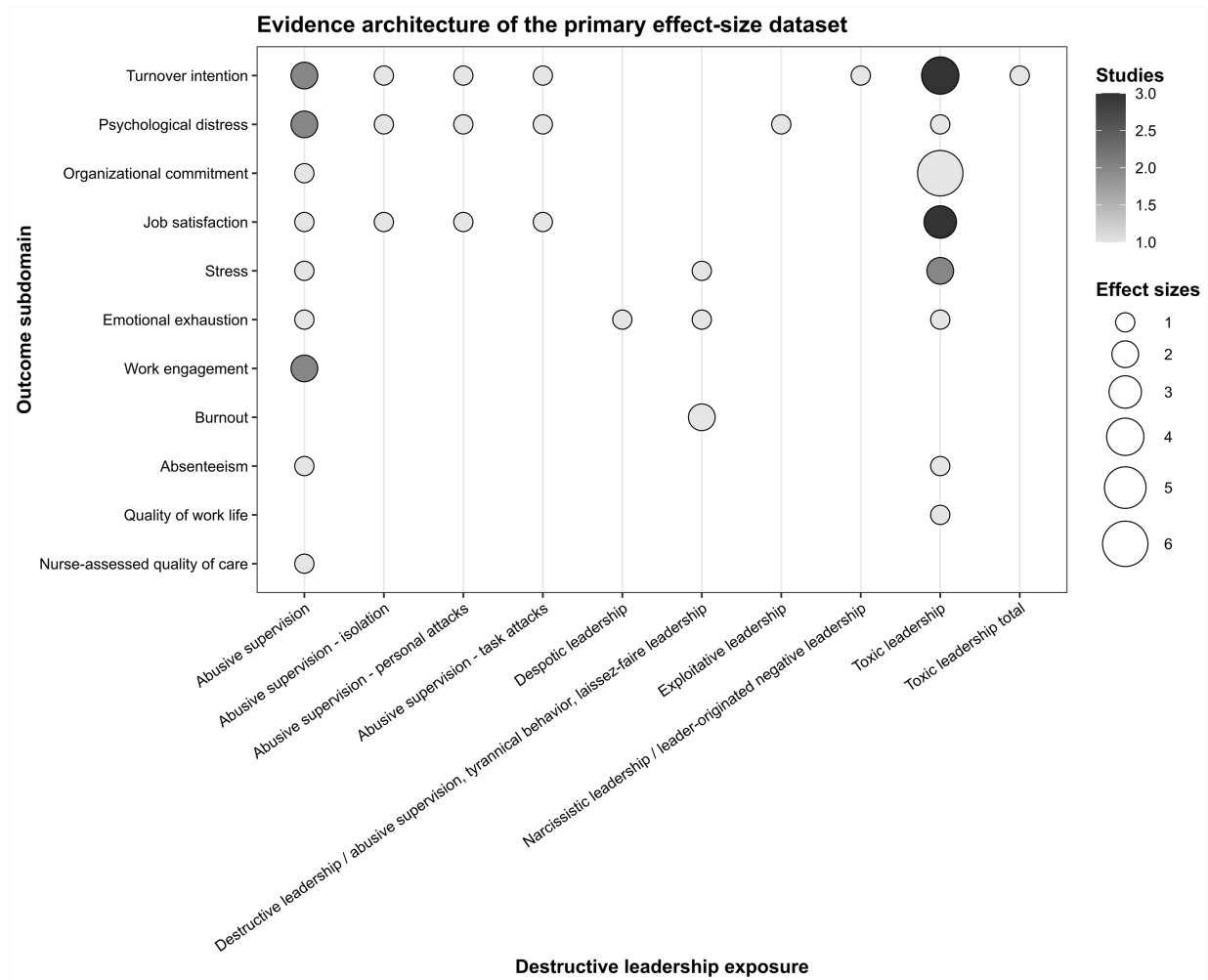


The diagram summarizes record identification, duplicate removal, title-and-abstract screening, full-text eligibility assessment, reasons for exclusion, and final inclusion in the evidence dataset, primary meta-analysis, sensitivity analyses, and primary MASEM analysis.

Diagram 1. PRISMA: flowchart of study selection.

3. Results

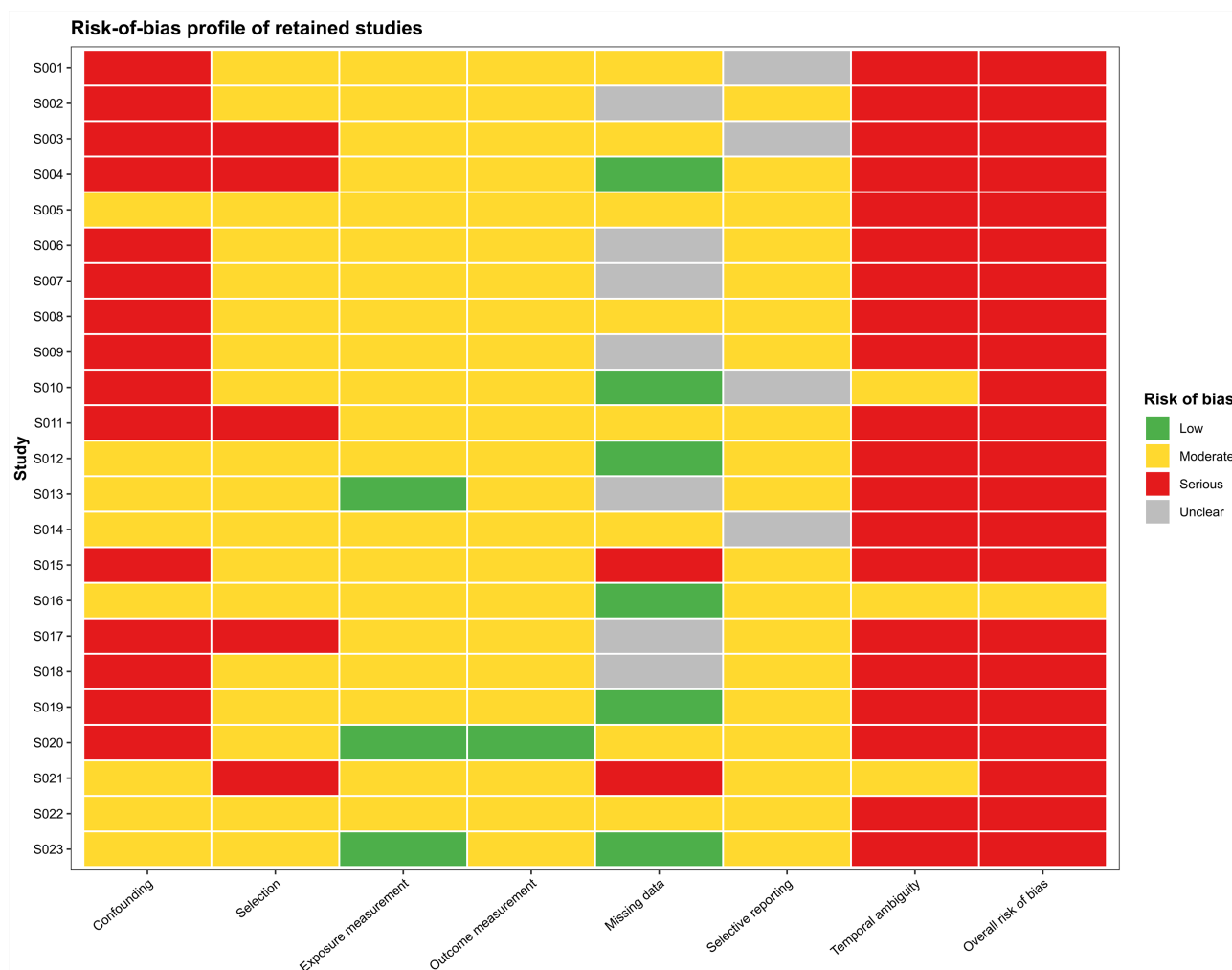
The primary analysis dataset contained 48 effect sizes from 20 studies, comprising 27 wellbeing effects, 20 retention-related effects, and one direct care-quality effect. As shown in **Figure 1**, the evidence was therefore concentrated overwhelmingly in workforce outcomes, with toxic leadership and abusive supervision representing the principal exposure categories. Turnover intention was the most frequently represented outcome, followed by job satisfaction, psychological distress, organizational commitment, stress, and emotional exhaustion; evidence for burnout, work engagement, absenteeism, quality of work life, and direct care quality was comparatively limited. Accordingly, quantitative inference focused primarily on wellbeing and retention, whereas care-quality evidence was retained as exploratory rather than treated as a separately poolable domain.



Bubble size indicates the number of effect sizes, and shading indicates the number of contributing studies across exposure constructs and outcome subdomains.

Figure 1. Evidence architecture of the primary effect-size dataset.

The risk-of-bias profile indicated substantial limitations in the underlying evidence base. Twenty-two of the 23 retained studies (95.7%) were rated as having serious overall risk of bias, and one study (4.3%) was rated as moderate. Temporal ambiguity was the most prominent domain-level concern, with 20 studies rated as serious and three as moderate, followed by confounding, for which 15 studies were rated as serious and eight as moderate. Selection was rated as serious in five studies and moderate in 18. Exposure measurement was rated as low in three studies and moderate in 20, whereas outcome measurement was rated as low in one study and moderate in 22. Missing-data judgments were more heterogeneous (six low, eight moderate, two serious, and seven unclear), while selective reporting was rated as moderate in 19 studies and unclear in four. Taken together, these judgments indicate that the pooled results should be interpreted as associations rather than causal effects, with particular caution regarding temporal ordering and residual confounding (Figure 2).



Traffic-light plot of study-level risk-of-bias judgments across methodological domains. Each row represents one retained study; side annotations summarize sample size, design, outcome domain, and exposure construct.

Figure 2. Risk-of-bias profile of retained studies.

Figure 3 and **Figure 4** present the domain-specific Bayesian multilevel estimates after harmonization so that positive correlations consistently represented more adverse outcomes. Individual primary effects were predominantly on the adverse side of the null, although their magnitudes varied substantially across studies and outcomes. The single direct care-quality effect was retained descriptively and was not modeled as an independent pooled domain.

The domain-specific models showed closely similar pooled associations for wellbeing and retention (**Table 2**). For wellbeing, the pooled adverse correlation was $r = 0.34$ (95% credible interval [CrI], 0.23 to 0.45), based on 27 effects from 14 studies. For retention, the corresponding pooled correlation was $r = 0.33$ (95% CrI, 0.19 to 0.46), based on 20 effects from 10 studies. In both models, the posterior probability of an adverse pooled association was approximately 1.00. The 95% prediction intervals were considerably wider than the credible intervals, extending from -0.13 to 0.69 for wellbeing and from -0.10 to 0.66 for retention. Thus, alt-

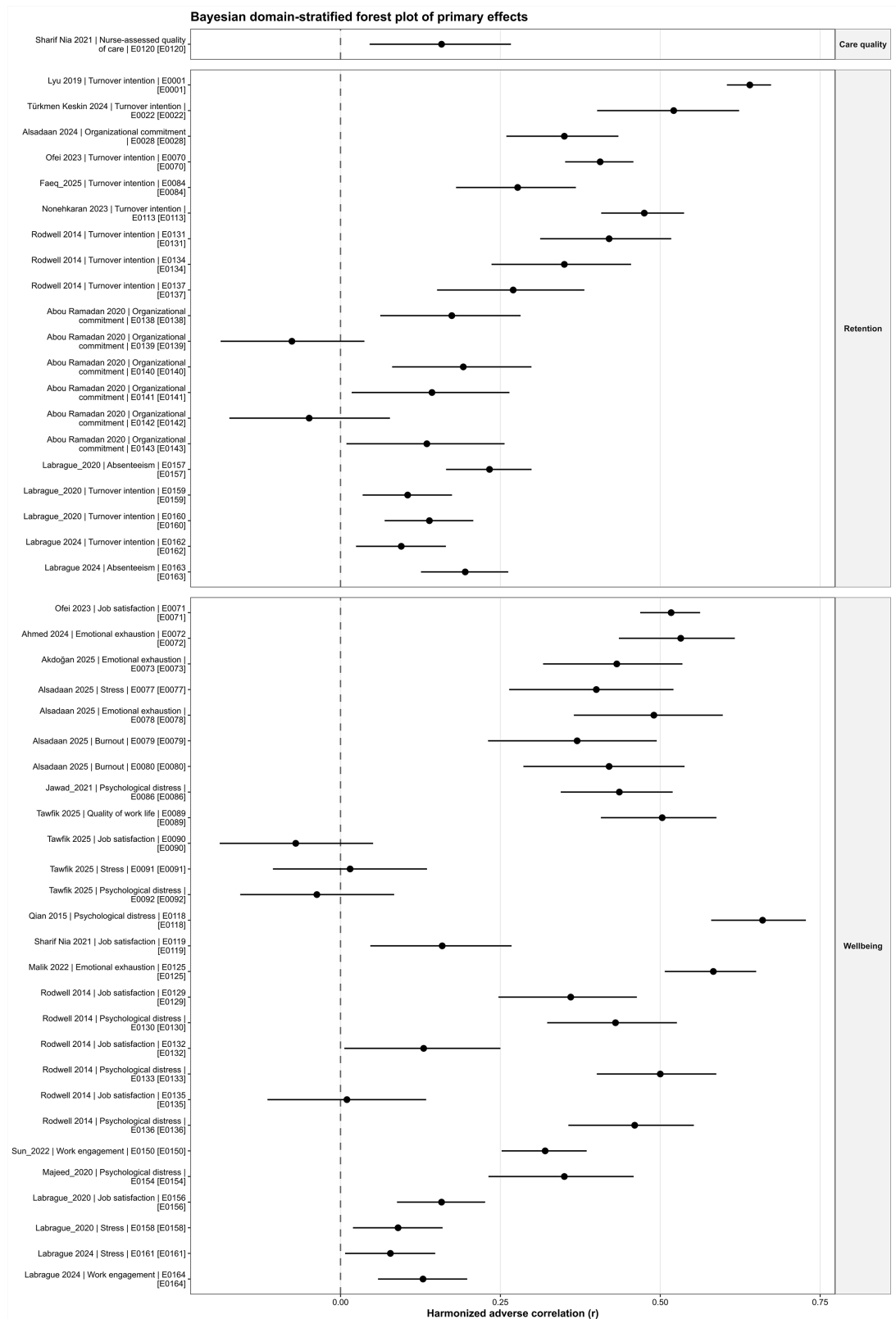
though the pooled mean associations were credibly adverse in both domains, the underlying association expected in a new comparable setting remains substantially heterogeneous and may be small or close to the null in some settings (Table 3).

Table 2. Bayesian pooled effects by primary outcome domain.

Outcome domain	Studies	Effects	Pooled r	95% credible interval	95% prediction interval
Wellbeing	14	27	0.34	0.23 to 0.45	−0.13 to 0.69
Retention	10	20	0.33	0.19 to 0.46	−0.10 to 0.66

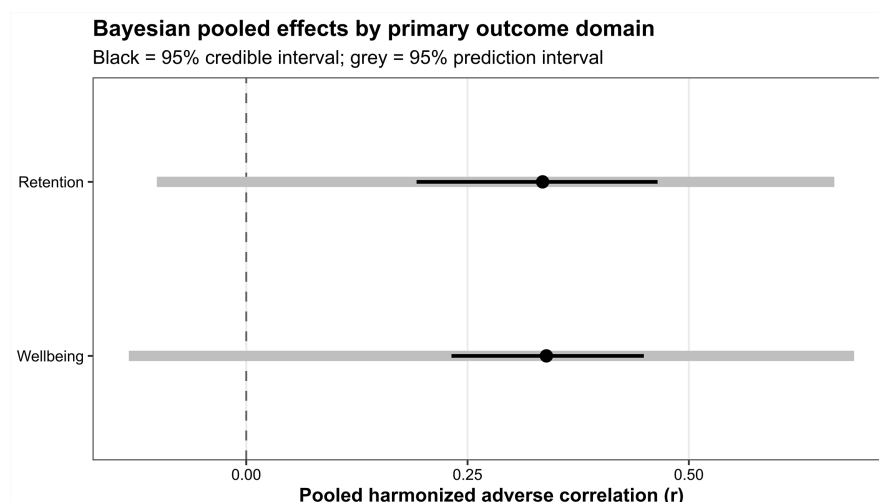
Note. Positive correlations indicate more adverse outcomes after directional harmonization. The heterogeneity estimates reported below refer to the overall 48-effect primary Bayesian model and are therefore not assigned to the separate wellbeing and retention rows of (Table 2).

In the overall primary Bayesian model containing all 48 harmonized effects, the pooled association was $r = 0.357$ (95% CrI, 0.268 to 0.447), with a posterior probability of a positive adverse association of approximately 1.00. Between-study heterogeneity was $\tau = 0.187$ (95% CrI, 0.111 to 0.296), corresponding to $\tau^2 = 0.035$ (95% CrI, 0.012 to 0.088). The model also identified residual effect-level heterogeneity of $\sigma = 0.138$ (95% CrI, 0.101 to 0.191), corresponding to $\sigma^2 = 0.019$ (95% CrI, 0.010 to 0.037). These components were estimated separately and indicate variation both between studies and among effects beyond their known sampling uncertainty. Computational diagnostics for the stored primary model indicated stable posterior sampling. The maximum $R^{\wedge}$ was 1.0014, minimum bulk and tail effective sample sizes were 2861 and 4708, respectively, no divergent transitions occurred, and no transitions reached the configured maximum treedepth; the maximum observed treedepth was 6. Posterior predictive checks likewise did not identify substantial model-data misfit. Across 2000 replicated 48-effect datasets, the observed mean Fisher-z value (0.3049) closely matched the predictive median (0.3043), and the observed standard deviation (0.2242) closely matched the predictive median standard deviation (0.2226). The observed median (0.3080), minimum (−0.0761), and maximum (0.7928) also fell within their respective 95% posterior predictive ranges. These results indicate that the fitted model reproduced the principal distributional characteristics of the observed effect-size data adequately. The overall pooled association was also insensitive to reasonable changes in prior specification. The baseline model yielded $r = 0.357$ (95% CrI, 0.268 to 0.447); the more regularizing specification yielded $r = 0.351$ (95% CrI, 0.260 to 0.435); and the more diffuse specification yielded $r = 0.363$ (95% CrI, 0.271 to 0.454). The posterior probability of an adverse association was approximately 1.00 under all three specifications. Study-level heterogeneity remained similar across models ($\tau = 0.181 - 0.189$), as did residual heterogeneity ($\sigma \approx 0.137 - 0.138$), indicating that the substantive conclusion was not materially dependent on the baseline prior specification. All prior-sensitivity models also converged satisfactorily, with maximum $R^{\wedge}$ values between 1.0011 and 1.0015, minimum bulk effective sample sizes exceeding 2000, minimum tail effective sample sizes exceeding 3400, and no divergent transitions or maximum-treedepth hits.



Individual harmonized adverse correlations are displayed by outcome domain. Points represent primary effect-size estimates and horizontal lines indicate their approximate 95% intervals.

Figure 3. Bayesian domain-stratified forest plot of primary effects.

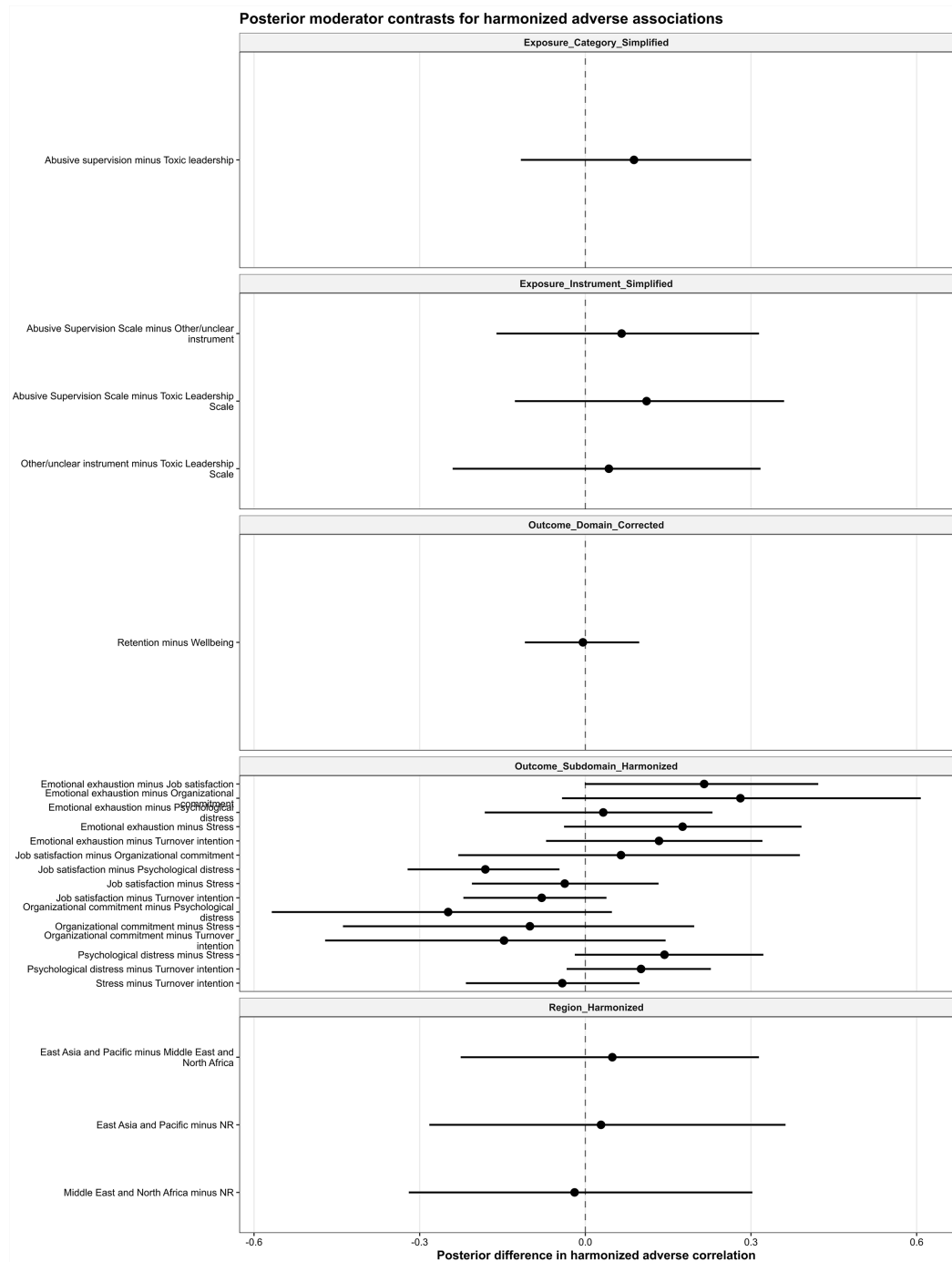


Black intervals indicate 95% credible intervals around pooled domain estimates; grey intervals indicate 95% prediction intervals.

Figure 4. Bayesian pooled effects by primary outcome domain.

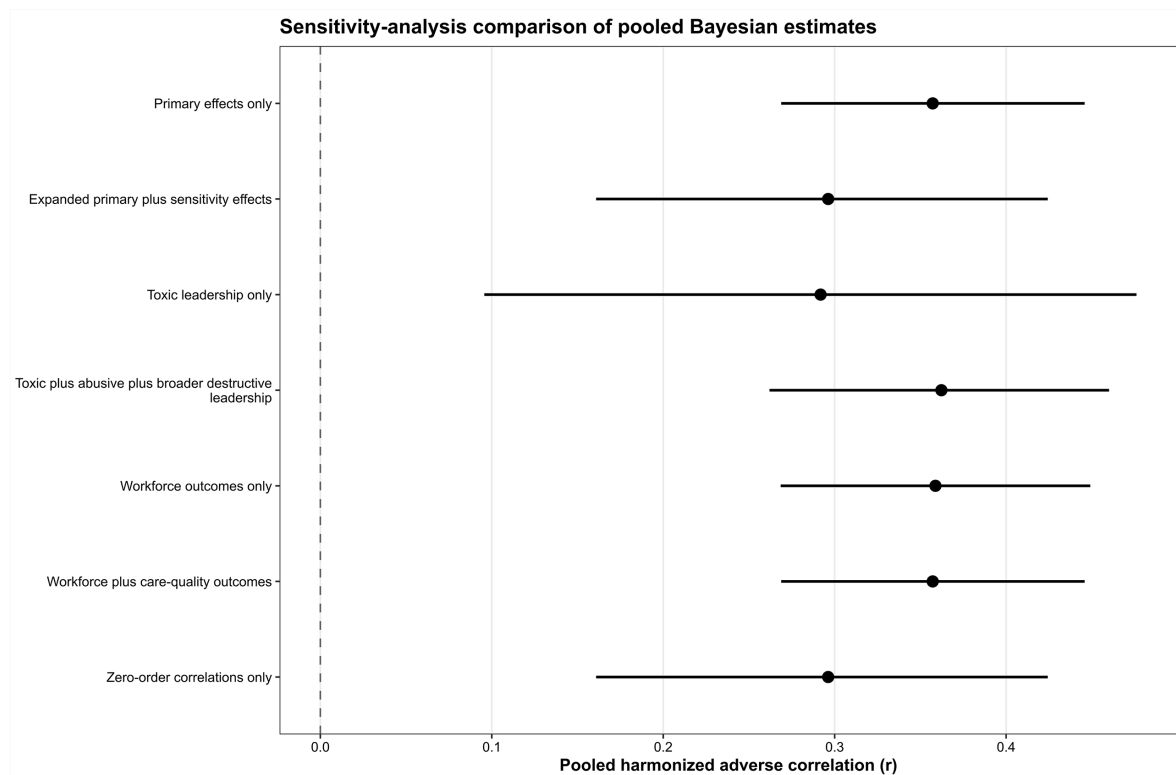
Figure 5 examined whether the pooled harmonized adverse association differed across prespecified moderators, including outcome subdomain, exposure category, exposure instrument, and region. The contrasts should be interpreted as differences between model-estimated pooled correlations, not as direct head-to-head comparisons within the same studies. The moderator analyses did not provide sufficiently precise evidence of systematic effect modification. Emotional exhaustion showed the largest positive contrasts relative to organizational commitment, job satisfaction, stress, and turnover intention, but the corresponding posterior uncertainty intervals included zero. The retention-versus-wellbeing contrast was centered near zero, consistent with the closely similar pooled estimates observed for the two primary workforce domains. Exposure-related contrasts were likewise uncertain: abusive supervision and the Abusive Supervision Scale showed directionally larger adverse estimates than toxic leadership or other instruments, but their uncertainty intervals did not establish reliable between-group differences. Regional contrasts were also imprecise. Thus, the available moderator analyses identified possible sources of heterogeneity but did not establish that the magnitude of the adverse association differed consistently by outcome subdomain, leadership label, measurement instrument, or geographical region. **Figure 6** assessed the robustness of the pooled association under alternative analytic decisions. The pooled harmonized association remained adverse across all examined sensitivity scenarios. Estimates were broadly comparable when analyses were restricted to workforce outcomes, when the single direct care-quality effect was included, and when only primary effects were retained. Broadening the exposure definition to include related destructive-leadership constructs produced a comparable, slightly larger estimate, whereas restriction to toxic leadership alone retained an adverse association but with greater uncertainty because fewer data were available. The zero-order-only and expanded primary-plus-sensitivity analyses yielded some-

what smaller estimates, but both remained on the adverse side of the null. These analyses therefore supported the stability of the direction of the primary association across major inclusion and coding decisions without implying equivalence among the underlying leadership constructs.



Points represent posterior differences in pooled harmonized adverse correlations between moderator levels; horizontal lines represent posterior uncertainty intervals. Positive values indicate a stronger adverse association for the first level relative to the comparator.

Figure 5. Posterior moderator contrasts for harmonized adverse associations.



Points represent pooled harmonized adverse correlations under alternative analytic specifications; horizontal lines represent posterior uncertainty intervals.

Figure 6. Sensitivity-analysis comparison of pooled Bayesian estimates.

3.1. Meta-Analytic Structural Equation Modeling

Figure 7 presents the primary MASEM workforce-pathway model linking toxic or abusive leadership with turnover intention through low job satisfaction and psychological distress. The primary MASEM evidence base comprised 39 eligible source correlations from 17 studies, yielding 18 distinct off-diagonal pooled correlation pairs after conceptual harmonization. The four-variable workforce model included toxic/abusive leadership, low job satisfaction, psychological distress, and turnover intention. The pooled 4×4 correlation matrix was positive definite (minimum eigenvalue = 0.5689) and required no matrix adjustment. The effective SEM sample size was 768. The model was just-identified ($df = 0$); consequently, the nominally perfect global fit indices were mathematical consequences of saturation and were not interpreted as evidence of model fit. Toxic or abusive leadership showed positive modeled associations with both intermediate workforce variables. The standardized association with low job satisfaction was 0.351 (95% CI, 0.284 to 0.417), and the association with psychological distress was 0.268 (95% CI, 0.200 to 0.336). Toxic or abusive leadership also retained a direct modeled association with turnover intention after both intermediate variables were included (0.311; 95% CI, 0.239 to 0.383). Psychological distress was independently associated with turnover intention (0.132; 95% CI, 0.063 to 0.200), whereas low job satisfaction was not independently associated with turnover intention in the same

model (0.001; 95% CI, -0.070 to 0.072).

As shown in (Table 3), the modeled indirect association through psychological distress was 0.035 (95% CI, 0.015 to 0.056), whereas the indirect association through low job satisfaction was essentially null (0.0004; 95% CI, -0.024 to 0.025). The total modeled indirect association was 0.036 (95% CI, 0.005 to 0.066), and the total modeled association between toxic or abusive leadership and turnover intention was 0.347 (95% CI, 0.281 to 0.413). These estimates indicate that the modeled workforce pathway was characterized primarily by the direct leadership-turnover association, accompanied by a smaller indirect association through psychological distress. Because the underlying correlations were predominantly observational and cross-sectional, these indirect estimates represent modeled associational pathways and should not be interpreted as evidence of causal mediation. Evidence coverage across the six correlations required for the four-variable workforce matrix was markedly uneven. Toxic/abusive leadership and turnover intention had the strongest coverage, with seven studies contributing eight source correlations. Toxic/abusive leadership-low job satisfaction was supported by two studies and two correlations; toxic/abusive leadership-psychological distress by three studies and three correlations; low job satisfaction-turnover intention by three studies and four correlations; psychological distress-turnover intention by two studies and three correlations; and low job satisfaction-psychological distress by two studies and two correlations. Thus, although 17 studies contributed to the broader primary MASEM evidence base, individual cells of the workforce matrix were often informed by only two or three studies. Figure 8 therefore documents an important limitation of the structural analysis: the path estimates are model-based summaries of an uneven correlation network rather than estimates derived from a common complete correlation matrix observed in every study. Only one study contributed a direct care-quality correlation. Care-quality evidence was therefore not incorporated into a confirmatory structural pathway and remained descriptive and exploratory.

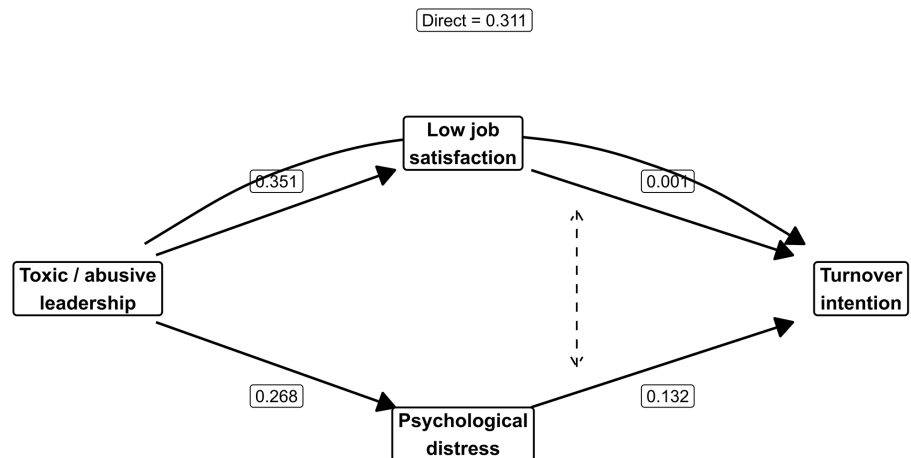
Table 3. Primary MASEM direct, indirect, and total modeled associations.

Pathway	Estimate	95% CI	Interpretation
Toxic/abusive leadership → low job satisfaction	0.351	0.284 to 0.417	Modeled positive association
Toxic/abusive leadership → psychological distress	0.268	0.200 to 0.336	Modeled positive association
Toxic/abusive leadership → turnover intention	0.311	0.239 to 0.383	Direct modeled association
Low job satisfaction → turnover intention	0.001	-0.070 to 0.072	No independently supported association
Psychological distress → turnover intention	0.132	0.063 to 0.200	Modeled positive association
Indirect association via low job satisfaction	0.0004	-0.024 to 0.025	No supported modeled indirect association
Indirect association via psychological distress	0.035	0.015 to 0.056	Modest modeled indirect association
Total indirect association	0.036	0.005 to 0.066	Total modeled indirect association
Total modeled association	0.347	0.281 to 0.413	Overall modeled workforce-pathway association

Note. Indirect associations are model-based associational quantities and are not interpreted as causal mediation effects. The structural model was just-identified ($df = 0$); global fit indices therefore do not provide an independent test of model fit.

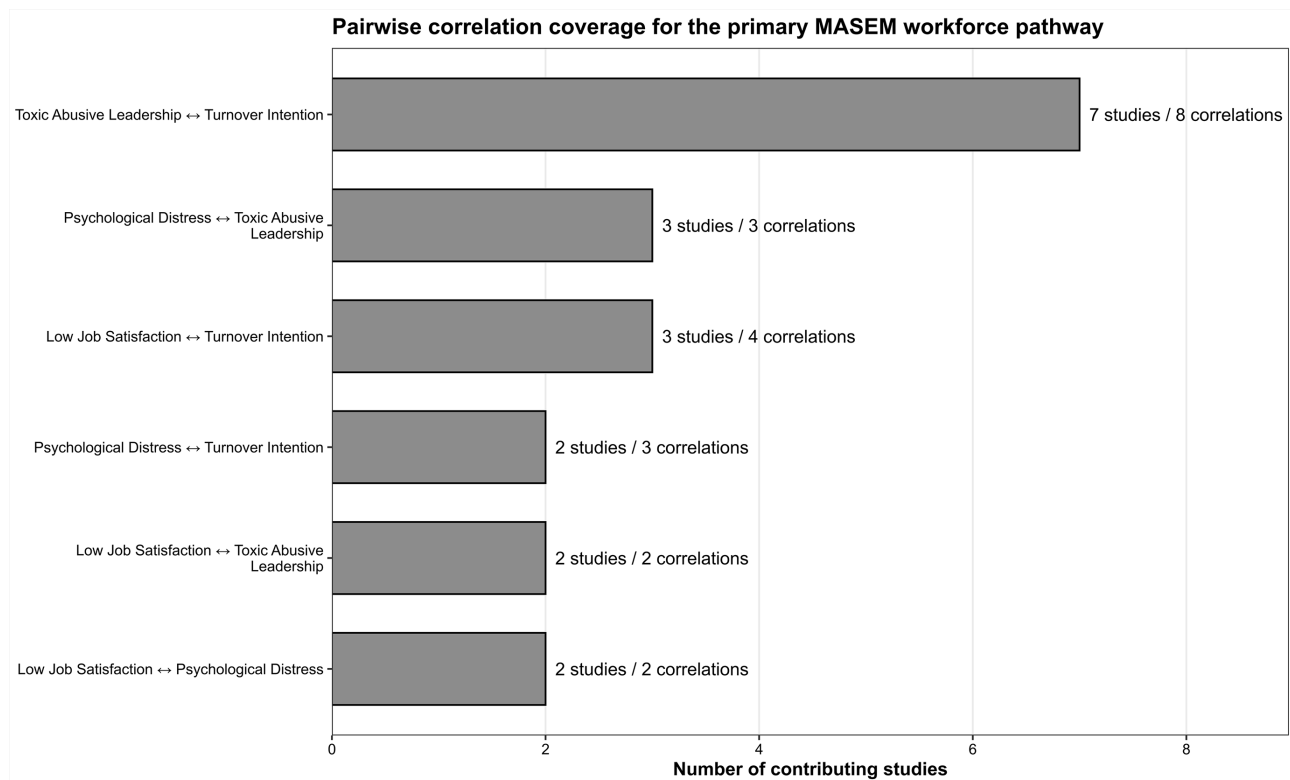
Primary MASEM workforce-pathway model

Standardized modeled associations; paths are associational, not causal



Structural model of the direct and modeled indirect associations linking toxic or abusive leadership with turnover intention through low job satisfaction and psychological distress. Paths represent standardized modeled associations and should not be interpreted causally.

Figure 7. Primary MASEM workforce-pathway model.



Bars indicate the number of studies contributing to each pooled correlation pair. Coverage differed substantially across matrix cells, with the leadership-turnover pair most strongly represented and several intermediate-variable pairs supported by only two or three studies.

Figure 8. Pairwise correlation coverage for the primary MASEM workforce pathway.

3.2. Small-Study Effects and Funnel Asymmetry

Figure 9 and **Figure 10** summarize the study-level funnel-asymmetry assessment for the overall primary analysis. Because several studies contributed multiple primary effect sizes, formal small-study-effect analyses were not conducted by treating all 48 effects as independent observations. Instead, the primary effects were aggregated to one Fisher-z estimate per study, yielding 20 study-level estimates. The conventional analysis initially assumed independence among multiple sampling errors within each study, while subsequent analyses explicitly examined the sensitivity of the findings to plausible within-study correlations. Under the conventional independence assumption ($\rho = 0$), the study-level random-effects estimate was $r = 0.381$ (95% CI, 0.299 to 0.458), with substantial heterogeneity ($\tau^2 = 0.0434$, $I^2 = 97.6\%$). Egger-type weighted regression indicated funnel asymmetry under this assumption ($t = 2.55$, $df = 18$, $P = 0.020$). Importantly, the value previously described as an “Egger intercept” is more accurately the regression-based limit estimate: at $\rho = 0$, this estimate was $z = 0.091$, corresponding to $r = 0.090$, with an interval that included the null. The asymmetry finding was not robust to plausible positive dependence among effects originating from the same study. When generalized least-squares aggregation assumed $\rho = 0.30$, the asymmetry test was no longer statistically evident ($P = 0.179$); corresponding values were $P = 0.455$ at $\rho = 0.50$ and $P = 0.808$ at $\rho = 0.70$. In contrast, the pooled study-level association was essentially unchanged across these assumptions, ranging only from $r = 0.381$ to $r = 0.384$, and τ^2 remained approximately 0.042 - 0.043. Thus, assumptions about within-study dependence had little effect on the estimated average study-level association but had a substantial effect on the evidence for funnel asymmetry.

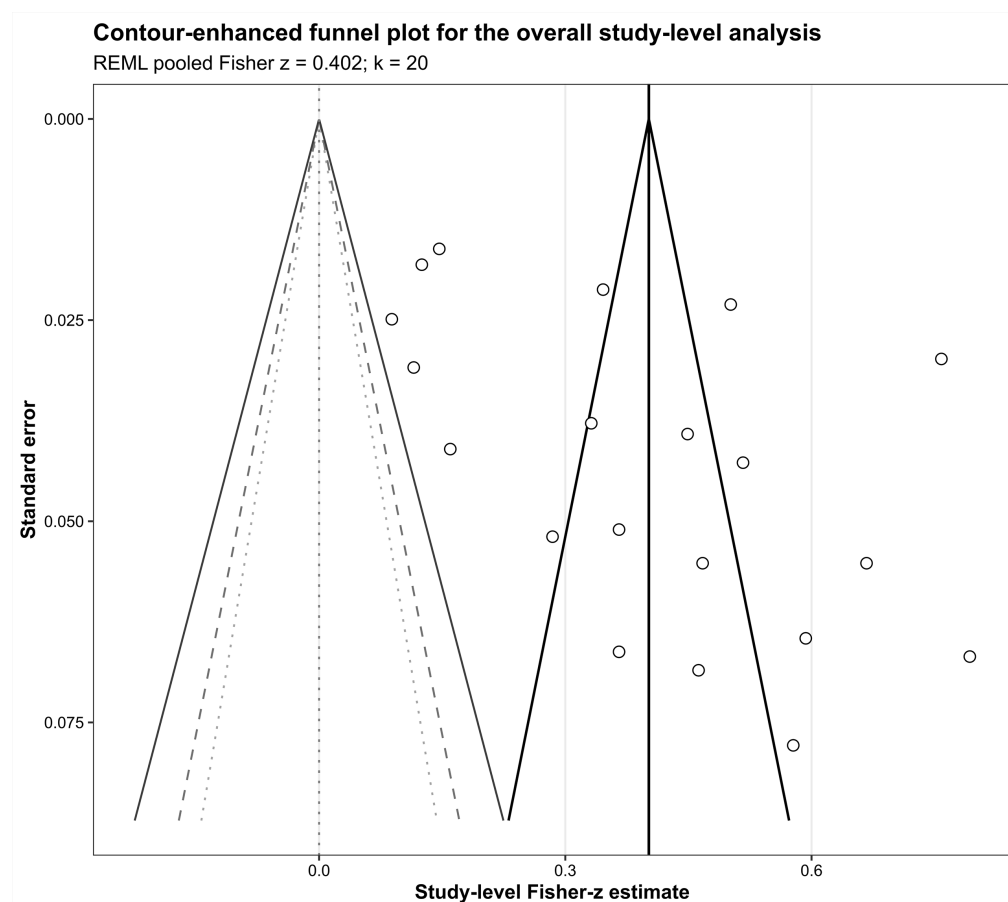
The regression-based limit estimate was also assumption-sensitive, increasing from $r = 0.090$ under $\rho = 0$ to approximately 0.164, 0.252, and 0.338 under $\rho = 0.30$, 0.50, and 0.70, respectively; its uncertainty interval included zero under every dependence assumption. Accordingly, neither the conventional asymmetry test nor the regression-based limit estimate provides a stable basis for concluding that publication bias is present or for replacing the primary Bayesian pooled estimate with a bias-adjusted alternative. Trim-and-fill did not impute additional studies in the original overall analysis; however, the absence of imputed studies was not interpreted as evidence that selective publication was absent. More generally, funnel asymmetry in this evidence base may reflect genuine between-study heterogeneity, differences in leadership constructs and outcome domains, measurement variation, clinical setting, or other small-study phenomena in addition to selective publication. Care quality was not assessable for funnel asymmetry because only one primary direct effect was available. Taken together, the small-study-effect analyses provide a more cautious interpretation than the conventional Egger result alone. A funnel-asymmetry signal was observed only under the assumption of independent within-study sampling errors and disappeared when moderate-to-strong positive within-study dependence was allowed. The direction and magni-

tude of the study-level pooled association, however, were highly stable across these dependence assumptions. These study-level REML estimates were used specifically for funnel-asymmetry sensitivity analyses and do not replace the primary multilevel Bayesian estimates reported above.

Table 4. Study-level funnel-asymmetry and within-study dependence sensitivity analyses.

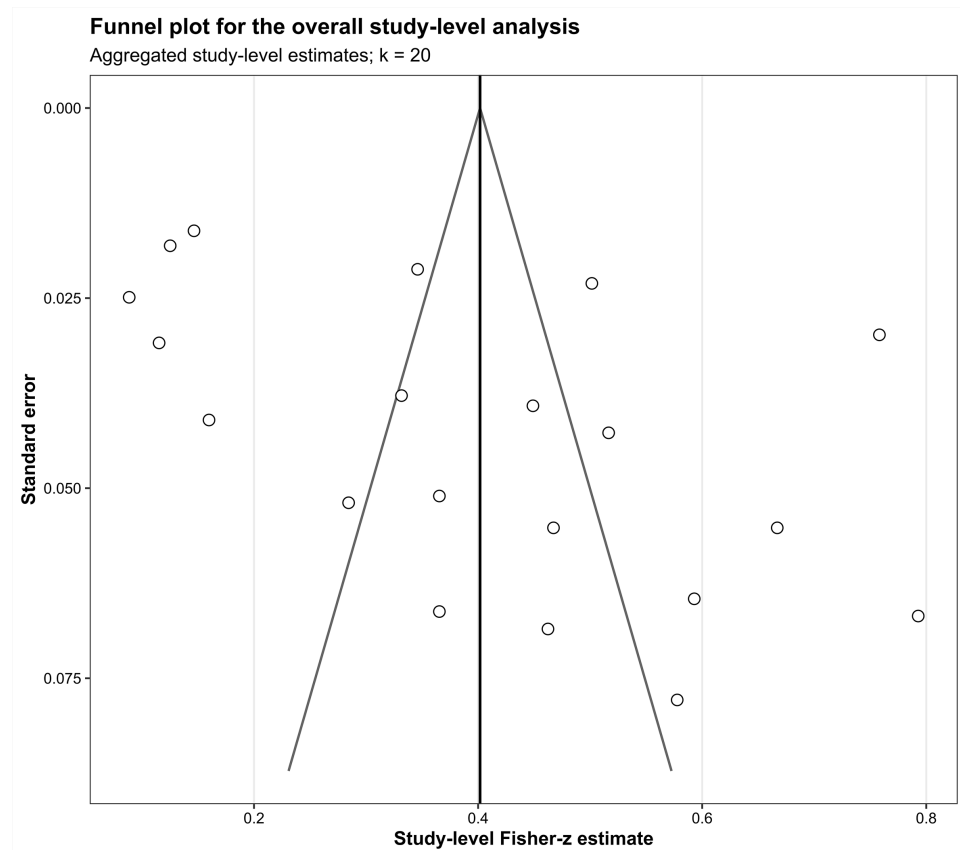
Assumed within-study ρ	Studies	Pooled r (95% CI)	τ^2	I^2 , %	Egger t (df = 18)	Egger asymmetry P	Regression-based limit r
0.00	20	0.381 (0.299 to 0.458)	0.0434	97.6	2.548	0.020	0.090
0.30	20	0.382 (0.300 to 0.459)	0.0430	96.2	1.398	0.179	0.164
0.50	20	0.383 (0.300 to 0.460)	0.0428	95.7	0.763	0.455	0.252
0.70	20	0.384 (0.301 to 0.460)	0.0424	95.2	0.246	0.808	0.338

Note. Effects were aggregated to one study-level estimate before funnel-asymmetry testing. ρ denotes the assumed common within-study correlation among sampling errors for multiple effects contributed by the same study. The pooled r values in this table arise from study-level REML sensitivity models and are not the primary Bayesian pooled estimates. Egger P values assess funnel asymmetry; the regression-based limit estimate is a separate model quantity and should not be described as the asymmetry-test intercept.



The 20 aggregated study-level Fisher- z estimates are plotted against their standard errors. The plot is interpreted descriptively because funnel asymmetry may arise from several mechanisms in a heterogeneous evidence base.

Figure 9. Contour-enhanced funnel plot for the overall study-level analysis.



Aggregated study-level Fisher-z estimates are displayed around the overall study-level random-effects estimate. Formal asymmetry results should be interpreted together with the within-study dependence sensitivity analyses reported in (Table 4).

Figure 10. Funnel plot for the overall study-level analysis.

4. Discussion

This PRISMA-guided quantitative evidence synthesis found that toxic and abusive leadership in nursing were associated with adverse workforce outcomes, particularly nurse wellbeing and retention-related outcomes. The primary Bayesian multilevel models showed closely similar pooled adverse correlations for wellbeing and retention, indicating that destructive leadership was linked both with psychological strain and with withdrawal-related responses. However, these findings should be interpreted as associations, not causal effects, because the retained evidence base was dominated by observational and cross-sectional designs, with substantial risk of bias, heterogeneity, and limited direct evidence for care-quality outcomes. The consistency of the findings across the principal analyses is important. In addition to the similar domain-specific estimates for wellbeing and retention, the overall primary Bayesian association remained essentially unchanged under more regularizing and more diffuse prior specifications. Posterior predictive checks also showed no substantial discrepancy between the observed effect-size distribution and data generated from the fitted model. These findings strengthen confidence in the numerical stability and model-based robustness of the pooled

association, but they do not reduce the methodological limitations of the underlying observational literature. In particular, good MCMC convergence, prior robustness, and adequate posterior predictive performance cannot remove confounding, temporal ambiguity, common-method bias, or selective reporting in the primary studies.

The comparable magnitude of the wellbeing and retention estimates suggests that toxic or abusive leadership is associated not only with how nurses experience their work psychologically but also with withdrawal-related responses relevant to workforce stability. Nursing practice occurs in settings characterized by clinical acuity, emotional labour, hierarchical authority, and interprofessional dependence. In such environments, sustained hostile, exploitative, or self-serving leadership behaviours may plausibly be associated with emotional exhaustion, psychological distress, reduced job satisfaction, weakened organizational attachment, and intention to leave. This interpretation is consistent with prior nursing evidence linking negative leadership behaviours with burnout, psychological distress, workplace deviance, job insecurity, workplace gaslighting, reduced job satisfaction, and diminished organizational commitment [3]-[8] [30] [38]. The findings also extend prior evidence on nurse retention. Earlier reviews have shown that leadership style, organizational support, burnout, job satisfaction, horizontal violence, and unfinished nursing care are associated with nurses' intention to leave or remain in practice [9]-[13]. The present synthesis narrows this broader literature to leader-originated destructive exposures and shows that their association with retention-related outcomes persists when multiple effect sizes contributed by the same studies are modeled hierarchically. The findings therefore support treating destructive leadership as one potentially modifiable organizational correlate of workforce retention rather than reducing nurses' intentions to leave solely to individual preferences or external labour-market conditions.

4.1. Heterogeneity and Construct Interpretation

The wide prediction intervals qualify the pooled findings. Although the average associations for wellbeing and retention were credibly adverse, future studies may observe smaller, null, or stronger associations. The overall Bayesian model likewise identified variability both between studies and among effects within studies beyond known sampling error, reinforcing that the pooled correlation represents an average across a heterogeneous evidence base rather than a universal effect size. An important source of this heterogeneity is conceptual. Toxic leadership, abusive supervision, exploitative leadership, despotic leadership, narcissistic leadership, and related leader-originated negative behaviours share a higher-order destructive orientation, but they are not interchangeable constructs. Abusive supervision emphasizes sustained hostile supervisory behaviour, whereas other constructs place greater emphasis on authoritarianism, exploitation, self-interest, narcissism, or broader organizational harm. Pooling them therefore serves the purpose of estimating a broader adverse-leadership association while preserving their concep-

tual differences through moderator and sensitivity analyses. The absence of reliable moderator contrasts should not be interpreted as evidence that these constructs are psychometrically equivalent; rather, the available evidence was insufficiently precise to establish systematic differences among them. Moderator analyses did not identify a reliable source of heterogeneity. Emotional exhaustion showed directionally larger contrasts than several other subdomains, but uncertainty intervals included zero. Exposure-label, instrument, and regional contrasts were similarly imprecise. Sensitivity analyses nevertheless showed that the adverse direction of the pooled association persisted when the leadership definition was narrowed or broadened and under alternative effect-inclusion decisions. This combination of findings supports robustness of the general adverse association while arguing against strong claims about differential effects across specific leadership constructs, instruments, regions, or workforce subdomains.

4.2. Workforce Pathways in the MASEM

The MASEM results refined the workforce-pathway question. Toxic or abusive leadership was associated with low job satisfaction and psychological distress and retained a direct modeled association with turnover intention. The indirect association through psychological distress was modest but supported, whereas the indirect association through low job satisfaction was essentially null after psychological distress and leadership exposure were included in the same model. Within the specified correlation structure, psychological distress therefore represented the more informative modeled indirect pathway between destructive leadership and turnover intention. This pattern is compatible with broader nursing evidence connecting psychological strain, burnout, adverse work environments, and retention-related outcomes [16]-[18] [20] [21]. The structural findings require important qualification. Although 17 studies contributed 39 source correlations to the primary MASEM evidence base, coverage of the individual correlations required for the four-variable workforce matrix was uneven. Several mediator-related correlations were informed by only two or three studies. Moreover, the structural model was just-identified ($df = 0$); consequently, nominally perfect fit indices are an algebraic consequence of saturation rather than evidence that the proposed workforce pathway uniquely or exceptionally fits the data. The model should therefore be interpreted as a structured summary of the available pooled correlation network, not as an independently validated causal model. For the same reason, the indirect association through psychological distress should not be described as demonstrating mediation. The contributing evidence was predominantly cross-sectional or observational and did not establish the temporal sequence required for causal mediation. Alternative directions of association and unmeasured common causes remain plausible. Longitudinal studies measuring leadership exposure before changes in psychological distress and subsequent turnover behaviour are needed before the proposed sequence can be evaluated causally.

4.3. Care-Quality Evidence

Care-quality findings require the greatest caution. The introduction identified a plausible workforce-to-care pathway through psychological safety, silence, reporting behaviour, escalation, and perceived care quality. This possibility is consistent with broader evidence linking nursing work environments, burnout, engagement, safety culture, missed care, and quality indicators [16] [18]-[21]. Previous studies link toxic and abusive leadership with organizational silence, psychological distress, emotional exhaustion, job neglect, and turnover intention among nurses [15] [28] [35] [39] [40]. These effects may extend to care quality through impaired communication, psychological safety, and team climate. Supporting this pathway [18] [41] [42] linked leadership practices with care quality, burnout, and intention to leave. However, the present synthesis contained only one primary direct care-quality effect, and only one study contributed a care-quality correlation to the broader MASEM evidence architecture. These data are insufficient for either an independent pooled care-quality estimate or a confirmatory structural care-quality pathway. Accordingly, the present study provides quantitative evidence primarily for workforce outcomes. Any downstream relationship with patient safety or care quality remains a hypothesis-generating research question rather than an established finding of this meta-analysis.

4.4. Small-Study Effects and Reporting Bias

The small-study-effect analyses also require a more nuanced interpretation than the conventional Egger result alone would suggest. When the 48 primary effects were aggregated to 20 study-level estimates under an independence assumption, the conventional Egger-type regression indicated funnel asymmetry ($P = 0.020$). However, this signal was highly sensitive to the assumed dependence among multiple effects originating from the same studies: it was no longer evident when within-study correlations of $\rho = 0.30$, 0.50 , or 0.70 were imposed. By contrast, the corresponding study-level pooled association changed negligibly across these assumptions. This pattern indicates that the evidence for funnel asymmetry is assumption-dependent, whereas the direction and approximate magnitude of the average adverse association are substantially more stable. The analysis therefore does not justify concluding either that publication bias is present or that it is absent. Funnel asymmetry may also arise from genuine heterogeneity in leadership constructs, outcome definitions, measurement instruments, sample characteristics, or study settings. Similarly, the absence of trim-and-fill imputations should not be interpreted as proof that selective reporting did not occur. Small-study-effect diagnostics are therefore best regarded as sensitivity analyses that qualify confidence in the exact magnitude of the pooled association rather than as evidence that overturns its adverse direction.

4.5. Strengths and Limitations

Several features strengthen the quantitative synthesis. The multilevel Bayesian ap-

proach accommodated multiple effect sizes contributed by the same studies rather than treating them as independent observations. The statistical model explicitly incorporated known sampling uncertainty together with study-level and residual effect-level heterogeneity. Computational diagnostics showed stable sampling, posterior predictive checks did not indicate substantial model-data misfit, and the primary overall association remained stable under alternative prior specifications. Sensitivity analyses also examined alternative exposure definitions and analytic inclusion decisions, while the funnel-asymmetry analysis explicitly evaluated plausible within-study dependence. These strengths should be considered alongside important limitations. First, 22 of the 23 retained studies were judged to have serious overall risk of bias, with temporal ambiguity and confounding particularly prominent. Most studies used cross-sectional or same-wave observational designs, preventing firm conclusions regarding directionality or causation. Second, many exposure and outcome measures were self-reported, creating the possibility of common-method inflation. Third, the review combined related but non-identical destructive-leadership constructs and heterogeneous workforce outcomes; although moderator and sensitivity analyses were undertaken, several subgroups contained limited information. Fourth, the MASEM correlation network was unevenly populated and the workforce model was saturated, limiting the degree to which structural estimates can discriminate among alternative theoretical models. Fifth, the conventional overall funnel-asymmetry signal was sensitive to assumptions concerning within-study dependence, leaving selective-publication effects uncertain. Finally, only one primary effect directly assessed care quality, precluding confirmatory conclusions concerning downstream patient-care consequences.

4.6. Practical Implications

For nurse managers and hospital administrators, the findings support treating toxic and abusive leadership as organizational workforce concerns that warrant systematic attention alongside staffing, workload, and other conditions influencing nurse wellbeing and retention. Health systems may consider incorporating validated assessments of destructive leadership behaviours into leadership evaluation, staff-climate monitoring, confidential reporting processes, and organizational risk-management procedures. Leadership-development programs should address hostile supervision, intimidation, exploitative or self-serving behaviour, psychological safety, constructive feedback, and accountability rather than focusing exclusively on the promotion of desirable leadership styles. The results do not establish that any specific leadership intervention will causally improve wellbeing or reduce turnover. Nevertheless, the consistency of the adverse associations supports monitoring destructive leadership together with indicators such as psychological distress, job satisfaction, engagement, absenteeism, and intention to leave. Such monitoring may help organizations identify settings in which leadership-related workforce risk warrants further assessment or intervention. Evaluation of leadership programs should ideally use prospective designs and measure both leadership behaviour and subsequent workforce outcomes.

5. Conclusion

This PRISMA-guided Bayesian multilevel meta-analysis and meta-analytic structural equation modeling study found that toxic and abusive leadership in nursing were associated with poorer workforce outcomes. The primary models showed similar adverse pooled associations for nurse wellbeing and retention-related outcomes. The overall association was stable across alternative prior specifications and major analytic sensitivity decisions, while substantial heterogeneity indicated that the magnitude of the association is likely to vary across settings, constructs, instruments, and populations. These findings support a consistent average adverse workforce association but not a single universal effect size. The MASEM identified a direct modeled association between toxic or abusive leadership and turnover intention and a smaller modeled indirect association through psychological distress. Low job satisfaction did not independently account for turnover intention after the other modeled variables were considered. Because the structural model was saturated and based predominantly on observational correlations with uneven pairwise coverage, these pathways should be interpreted as associational rather than causal mediation mechanisms. Evidence regarding publication-related small-study effects was also uncertain: the conventional overall asymmetry signal was not robust to plausible assumptions about within-study dependence. This uncertainty affects confidence in the exact magnitude of the pooled association but does not materially alter its adverse direction. Direct care-quality evidence was sparse and insufficient for confirmatory modeling. The principal conclusion is therefore workforce-focused: toxic and abusive leadership are associated with poorer nurse wellbeing and retention-related outcomes, whereas downstream effects on patient safety and care quality remain insufficiently established. Longitudinal, multi-source, and care-linked studies are needed to distinguish destructive leadership constructs more precisely, establish temporal ordering, test the proposed psychological-distress pathway, and determine whether reducing destructive leadership produces measurable improvements in nurse and patient outcomes.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Tepper, B.J. (2000) Consequences of Abusive Supervision. *Academy of Management Journal*, **43**, 178-190. <https://doi.org/10.2307/1556375>
- [2] Labrague, L.J. (2024) Abusive Supervision and Its Relationship with Nursing Workforce and Patient Safety Outcomes: A Systematic Review. *Western Journal of Nursing Research*, **46**, 52-63. <https://doi.org/10.1177/01939459231212402>
- [3] Ahmed, A.K., Atta, M.H.R., El-Monshed, A.H. and Mohamed, A.I. (2024) The Effect of Toxic Leadership on Workplace Deviance: The Mediating Effect of Emotional Ex-

- haustion, and the Moderating Effect of Organizational Cynicism. *BMC Nursing*, **23**, Article No. 669. <https://doi.org/10.1186/s12912-024-02308-x>
- [4] Abdou, A.H. and Albohnayh, A.S.M. (2026) When Nursing Leadership Turns Toxic: Mediating, Amplifying, and Buffering Effects of Emotional Exhaustion, Job Insecurity, and Coworker Support. *BMC Nursing*, **25**, Article No. 382. <https://doi.org/10.1186/s12912-026-04545-8>
- [5] Austin, R.L. (2026) Toxic Leadership Behaviors and Employee Turnover Intention: Mediated by Job Satisfaction. ProQuest Dissertations. <https://www.proquest.com/dissertations-theses/toxic-leadership-behaviors-employee-turnover/docview/3326968448/se-2>
- [6] Chen, S., Zhang, X., Chen, H., Zhang, F., Tung, T. and Chen, H. (2026) Cultural Differences in the Relationship between Negative Leadership Behaviours and Nurses' Emotional Exhaustion: A Systematic Review. *Journal of Nursing Management*, **2026**, e9748169. <https://doi.org/10.1155/jonm/9748169>
- [7] Moisoglou, I., Katsiourmpa, A., Konstantakopoulou, O., Papathanasiou, I.V., Katsapi, A., Prasini, I., *et al.* (2025) Workplace Gaslighting Is Associated with Nurses' Job Burnout and Turnover Intention in Greece. *Healthcare*, **13**, Article 1574. <https://doi.org/10.3390/healthcare13131574>
- [8] Mrayyan, M.T. and Abu-Hantash, D.W. (2026) The Impact of Perceived Clinical Nurse Managers' Despotic Leadership on Nurse Turnover Intention: A Cross-Sectional Study. *Journal of Advanced Nursing*, **82**, 6107-6127. <https://doi.org/10.1111/jan.70259>
- [9] Cho, Y., Jeong, S.H., Kim, H.S. and Kim, Y.M. (2022) Effects of Leadership Styles of Nursing Managers on Turnover Intention of Hospital Nurses: A Systematic Review and Meta-Analysis. *Journal of Korean Academy of Nursing*, **52**, 479-498. <https://doi.org/10.4040/jkan.22039>
- [10] Galanis, P., Moisoglou, I., Papathanasiou, I.V., Malliarou, M., Katsiourmpa, A., Vraha, I., *et al.* (2024) Association between Organizational Support and Turnover Intention in Nurses: A Systematic Review and Meta-Analysis. *Healthcare*, **12**, Article 291. <https://doi.org/10.3390/healthcare12030291>
- [11] Kim, H. and Kim, E.G. (2021) A Meta-Analysis on Predictors of Turnover Intention of Hospital Nurses in South Korea (2000-2020). *Nursing Open*, **8**, 2406-2418. <https://doi.org/10.1002/nop2.872>
- [12] Stemmer, R., Bassi, E., Ezra, S., Harvey, C., Jojo, N., Meyer, G., *et al.* (2022) A Systematic Review: Unfinished Nursing Care and the Impact on the Nurse Outcomes of Job Satisfaction, Burnout, Intention-to-Leave and Turnover. *Journal of Advanced Nursing*, **78**, 2290-2303. <https://doi.org/10.1111/jan.15286>
- [13] Zhang, Y., Yin, R., Lu, J., Cai, J., Wang, H., Shi, X., *et al.* (2022) Association between Horizontal Violence and Turnover Intention in Nurses: A Systematic Review and Meta-Analysis. *Frontiers in Public Health*, **10**, Article 964629. <https://doi.org/10.3389/fpubh.2022.964629>
- [14] Alshmemri, M. (2026) The Dark Side of Nurse-Manager Leadership: Toxic Leadership Behaviours, Adverse-Event Reporting, and Care Quality—An Interpretive Descriptive Study. *BMC Nursing*, **25**, Article No. 329. <https://doi.org/10.1186/s12912-026-04506-1>
- [15] Li, Z.Y., Yang, Y.P., Wang, Q. and Tung, T.H. (2024) The Relationship between Negative Leadership Behaviours and Silence among Nurses. *Journal of Nursing Management*, **2024**, Article 4561005. <https://doi.org/10.1155/2024/4561005>
- [16] Garcia, C.L., Abreu, L.C., Ramos, J.L.S. and Castro, C.F.D. (2019) Influence of Burn-

- out on Patient Safety: Systematic Review and Meta-Analysis. *Medicina*, **55**, Article 553. <https://doi.org/10.3390/medicina55090553>
- [17] Lai, Y.H., Chang, M.Y. and Wang, C.C. (2025) *Healthcare*, **13**, Article 2718. <https://doi.org/10.3390/healthcare13212718>
- [18] Laschinger, H.K.S. and Leiter, M.P. (2006) The Impact of Nursing Work Environments on Patient Safety Outcomes: The Mediating Role of Burnout Engagement. *JONA: The Journal of Nursing Administration*, **36**, 259-267. https://journals.lww.com/jonajournal/abstract/2006/05000/the_impact_of_nursing_work_environments_on_patient.19.aspx
- [19] Li, L.Z., Yang, P., Singer, S.J., Pfeffer, J. and Mathur, M.B. (2024) Nurse Burnout and Patient Safety, Satisfaction, and Quality of Care: A Systematic Review and Meta-Analysis. *JAMA Network Open*, **7**, e2443059. <https://doi.org/10.1001/jamanetworkopen.2024.43059>
- [20] Liu, X., Zheng, J., Liu, K., Baggs, J.G., Liu, J., Wu, Y., *et al.* (2018) Hospital Nursing Organizational Factors, Nursing Care Left Undone, and Nurse Burnout as Predictors of Patient Safety: A Structural Equation Modeling Analysis. *International Journal of Nursing Studies*, **86**, 82-89. <https://doi.org/10.1016/j.ijnurstu.2018.05.005>
- [21] Mossburg, S.E. and Himmelfarb, C.D. (2021) The Association between Professional Burnout and Engagement with Patient Safety Culture and Outcomes: A Systematic Review. *Journal of Patient Safety*, **17**, e1307-e1319. <https://doi.org/10.1097/pts.0000000000000519>
- [22] Ahumada, M., Ansoleaga, E., Ramírez-Vielma, R., Mera-Lemp, M.J., Soto-Contreras, E. and Brito-Placencia, D. (2025) Relationship between Destructive Leadership Styles and Health Workers' Mental Health: A Systematic Review. *Revista Brasileira de Medicina do Trabalho*, **23**, 1-11. <https://doi.org/10.47626/1679-4435-2025-1378>
- [23] Fischer, T., Tian, A.W., Lee, A. and Hughes, D.J. (2021) Abusive Supervision: A Systematic Review and Fundamental Rethink. *The Leadership Quarterly*, **32**, Article 101540. <https://doi.org/10.1016/j.leaqua.2021.101540>
- [24] Pan, L., Shang, J., Xue, K., Yang, C., Shi, H., Yu, G., *et al.* (2026) The Levels, Prevalence, and Related Factors of Toxic Leadership among Nurses outside the United States: A Systematic Review and Meta-Analysis. *JONA: The Journal of Nursing Administration*, **56**, 266-273. <https://doi.org/10.1097/nnn.0000000000001725>
- [25] Sun, S., Chen, H., He, Y., Yu, F., Yang, Y., Chen, H., *et al.* (2025) Workplace Bullying and Turnover Intentions among Workers: A Systematic Review and Meta-Analysis. *BMC Public Health*, **25**, Article No. 2394. <https://doi.org/10.1186/s12889-025-23339-2>
- [26] Aromataris, E., Fernandez, R., Godfrey, C.M., Holly, C., Khalil, H. and Tungpunkom, P. (2015) Summarizing Systematic Reviews: Methodological Development, Conduct and Reporting of an Umbrella Review Approach. *International Journal of Evidence-Based Healthcare*, **13**, 132-140. <https://doi.org/10.1097/xe.000000000000055>
- [27] Porritt, K., Evans, C., Bennett, C., Loveday, H., Bjerrum, M., Salmond, S., *et al.* (2024) Systematic Reviews of Qualitative Evidence. JBI Manual for Evidence Synthesis. JBI. <https://synthesismanual.jbi.global>
- [28] Lyu, D., Ji, L., Zheng, Q., Yu, B. and Fan, Y. (2019) Abusive Supervision and Turnover Intention: Mediating Effects of Psychological Empowerment of Nurses. *International Journal of Nursing Sciences*, **6**, 198-203. <https://doi.org/10.1016/j.ijnss.2018.12.005>
- [29] Türkmen Keskin, S. and Özduyan Kiliç, M. (2024) Investigation of the Relationship between Nurses' Perception of Toxic Leadership and Their Organizational Trust Lev-

- els and Turnover Intentions. *Journal of Advanced Nursing*, **80**, 1859-1867.
<https://doi.org/10.1111/jan.15951>
- [30] Alsadaan, N. and Alqahtani, M. (2024) Toxic Leadership in Emergency Nurses: Assessing Abusive Supervision and Its Team-Level Impacts on Conflict Management and Organizational Commitment. *Journal of Nursing Management*, **2024**, 1-11.
<https://doi.org/10.1155/2024/4271602>
- [31] Ofei, A.M.A., Poku, C.A., Paarima, Y., Barnes, T. and Kwashie, A.A. (2023) Toxic Leadership Behaviour of Nurse Managers and Turnover Intentions: The Mediating Role of Job Satisfaction. *BMC Nursing*, **22**, Article No. 374.
<https://doi.org/10.1186/s12912-023-01539-8>
- [32] Qian, J., Wang, H., Han, Z.R., Wang, J. and Wang, H. (2015) Mental Health Risks among Nurses under Abusive Supervision: The Moderating Roles of Job Role Ambiguity and Patients' Lack of Reciprocity. *International Journal of Mental Health Systems*, **9**, Article 22. <https://doi.org/10.1186/s13033-015-0014-x>
- [33] Rodwell, J., Brunetto, Y., Demir, D., Shacklock, K. and Farr-Wharton, R. (2014) Abusive Supervision and Links to Nurse Intentions to Quit. *Journal of Nursing Scholarship*, **46**, 357-365. <https://doi.org/10.1111/jnu.12089>
- [34] Sun, N., Zheng, Q., Li, L., Zhu, H., Liu, X., Zhou, S., *et al.* (2022) A Model of Abusive Supervision, Self-Efficacy, and Work Engagement among Chinese Registered Nurses: The Mediating Role of Self-Efficacy. *Frontiers in Psychology*, **13**, Article 962403.
<https://doi.org/10.3389/fpsyg.2022.962403>
- [35] Majeed, M. and Fatima, T. (2020) Impact of Exploitative Leadership on Psychological Distress: A Study of Nurses. *Journal of Nursing Management*, **28**, 1713-1724.
<https://doi.org/10.1111/jonm.13127>
- [36] Labrague, L.J., Nwafor, C.E. and Tsaras, K. (2020) Influence of Toxic and Transformational Leadership Practices on Nurses' Job Satisfaction, Job Stress, Absenteeism and Turnover Intention: A Cross-Sectional Study. *Journal of Nursing Management*, **28**, 1104-1113. <https://doi.org/10.1111/jonm.13053>
- [37] Labrague, L.J. (2024) Stress as a Mediator between Abusive Supervision and Clinical Nurses' Work Outcomes. *International Nursing Review*, **71**, 997-1004.
<https://doi.org/10.1111/inr.12961>
- [38] Abou Ramdan, A.H. and Eid, W.M. (2020) Toxic Leadership: Conflict Management Style and Organizational Commitment among Intensive Care Nursing Staff. *Evidence-Based Nursing Research*, **2**, 12. <https://doi.org/10.47104/ebnrojs3.v2i4.160>
- [39] Jawad, N., Malik, O.F. and Shahzad, A. (2021) Cross-Lagged Relationship between Abusive Supervision and Psychological Health among Nurses. *Research Journal of Social Sciences and Economics Review*, **2**, 66-74.
[https://doi.org/10.36902/rjsser-vol2-iss3-2021\(66-74\)](https://doi.org/10.36902/rjsser-vol2-iss3-2021(66-74))
- [40] Malik, O.F., Jawad, N., Shahzad, A. and Waheed, A. (2022) Longitudinal Relations between Abusive Supervision, Subordinates' Emotional Exhaustion, and Job Neglect among Pakistani Nurses: The Moderating Role of Self-Compassion. *Current Psychology*, **42**, 26945-26965. <https://doi.org/10.1007/s12144-022-03817-1>
- [41] Sharif Nia, H., Pahlevan Sharif, S., She, L., Sivarajan Froelicher, E., Kaveh, O. and Rahmatpour, P. (2022) The Relationship between Abusive Supervision, Psychological Ownership, and Quality of Nursing Care: The Mediating Role of Job Satisfaction. *Perspectives in Psychiatric Care*, **58**, 1223-1231.
<https://doi.org/10.1111/ppc.12919>
- [42] Lavoie-Tremblay, M., Fernet, C., Lavigne, G.L. and Austin, S. (2016) Transforma-

tional and Abusive Leadership Practices: Impacts on Novice Nurses, Quality of Care and Intention to Leave. *Journal of Advanced Nursing*, **72**, 582-592.
<https://doi.org/10.1111/jan.12860>