

Article

Identification of Psychosocial Risk Profiles Using Machine Learning Clustering Techniques for Human Resource Intervention Planning

Jose E. Naranjo^{1,*}, Mauro Darío Albarracín Álvarez¹

¹ Universidad Técnica de Cotopaxi, UTC, CIYA, Campus La Matriz - Av. Simón Rodríguez, Latacunga, 050102, Ecuador

* Corresponding author: Jose E. Naranjo (J.E.N.), jose.naranjo0463@utc.edu.ec

CITATION

Naranjo JE, Albarracín Álvarez MD. (2026). Identification of psychosocial risk profiles using machine learning clustering techniques for human resource intervention planning. *Human Resources Management and Services*. 8(3): 6083. <https://doi.org/10.18282/hrms6083>

ARTICLE INFO

Received: 29 June 2026

Accepted: 20 July 2026

Available online: 8 August 2026

COPYRIGHT



Copyright © 2026 by author(s).

Human Resources Management and Services is published by PiscoMed Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license.

<https://creativecommons.org/licenses/by/4.0/>

Abstract: Although psychosocial risk assessment is a managerial and regulatory concern, validated tools like FPSICO frequently offer aggregate dimension scores that may mask the co-occurrence of numerous risk variables within employee groups. A replicable unsupervised machine-learning framework for developing interpretable psychological risk profiles and connecting them to occupational health and human resource intervention planning is proposed in this proof-of-concept study. The methodology was illustrated using a wholly synthetic dataset of 180 workers represented by the nine FPSICO variables on a 0–100 risk scale, as access to actual psychosocial evaluation data is limited due to confidentiality concerns. Data normalization, principal component analysis, K-Means clustering, internal cluster evaluation, comparison with Ward hierarchical clustering, and profile interpretation were all part of the analytical workflow. 77.4% of the variation was explained by two main components. With a mean silhouette coefficient of 0.50, a five-cluster solution yielded the best overall structure. With an adjusted Rand index of 0.99, the recovered clusters closely matched the known generative design. The resultant profiles of psychosocial risk were balanced circumstances, psychological demands and workload, inadequate social support and engagement, role conflict and low recognition, and generally high psychosocial risk. The hierarchy of controls was used to relate each profile to an illustrated intervention matrix. The suggested paradigm offers a clear decision-support layer that supplements proven psychological evaluation but does not replace it, making the contribution scientific rather than empirical. The crucial next step is empirical validation using actual organizational data.

Keywords: psychosocial risk; FPSICO; unsupervised machine learning; K-Means clustering; principal component analysis; worker segmentation; hierarchy of controls; human resource management; synthetic data; reproducible methodology; proof of concept

1. Introduction

Work related psychosocial risks, arising from how work is designed, organized, and managed, are now recognized as major determinants of employee health (Lunn et al., 2026), wellbeing, and organizational performance. Sustained exposure to adverse psychosocial conditions such as excessive workload, low autonomy, role ambiguity, weak social support (Alali et al., 2025), and inadequate recognition is associated with stress, burnout, absenteeism, turnover intention, and reduced productivity (Amoadu et al., 2023). Regulatory frameworks across many jurisdictions oblige employers to assess and manage these risks, and international guidance such as ISO 45003 has formalized the management of psychological health and safety at work (Nebbs et al., 2025).

In Spanish speaking and European contexts, the FPSICO method developed by the National Institute for Safety and Health at Work (INSST) is one of the most widely used validated instruments for psychosocial risk assessment (Nebbs et al.,

2025). FPSICO characterizes the psychosocial environment through nine dimensions and classifies the level of risk on each dimension. While information by dimension is crucial for diagnosis and compliance with the law, it presents a practical constraint for managers: results are often aggregated by unit or dimension, which can obscure how various risk factors are grouped within specific sets of workers. Two departments or organizations may have similar average scores, but need very different actions because the underlying risk combinations are different. Occupational health and human resources (HR) teams work with constrained budgets and must set priorities (Neuhaus et al., 2026). In order for interventions to be designed and targeted effectively, they need to know not only which dimensions are higher on average, but also which risk configurations coexist, how common they are, and which employees have a similar risk signature. This is basically a matter of segmentation and is very appropriate for unsupervised machine learning (Donohoe et al., 2024).

This study addresses that methodological gap by proposing a reproducible methodology, rather than reporting on findings concerning a specific organization. To identify latent psychosocial risk profiles from FPSICO dimensions, we used unsupervised learning, principal component analysis (PCA) for structural exploration, and K-Means clustering for segmentation (Neuhaus et al., 2026). These methods were validated through hierarchical groupings. Since true psychosocial assessment data are difficult to acquire and share for reasons of confidentiality, the methodology is shown through a fully synthetic dataset created to mimic the results of FPSICO. The use of simulated data is an intentional methodological decision, as it enables the open disclosure of the entire pipeline and the total dataset for reproduction, and provides a familiar structure with which the method's ability to retrieve profiles can be objectively tested. We propose the contribution as a complement and not as a substitute for the FPSICO assessment (Bennett et al., 2026). The instrument offers a validated measurement, while clustering provides a layer of decision support that can be interpreted. Then, we show how each profile can be converted into an HR intervention plan that is organized according to the hierarchy of controls (Lot Tanko, 2020).

Therefore, the objectives are methodological in nature: (i) to suggest a complete, repeatable analytical workflow to classify workers into psychosocial risk profiles that can be interpreted using FPSICO dimensions; (ii) test its performance on a synthetic dataset with an established profile structure; (iii) verify, using internal indices of validity and recovery of the known design, that the analytical pipeline produces stable and interpretable clusters; and (iv) demonstrate how it is possible to map the resulting profiles to a matrix for practical interventions based on the hierarchy of controls for those responsible for the human resources department (Rivera Armendariz & Banning, 2026). Before concluding, the rest of the article examines related work, presents the materials and methods, including how the data are generated, shows what has been demonstrated, discusses the implications and pathway to empirical validation, describes the framework for planning interventions, and points out limitations (Wang et al., 2026).

2. Materials and Methods

2.1. Research design and framework

This is a methodological study with a demonstrative design: the aim is to propose and exercise a reproducible pipeline that uncovers structure and groups workers by their psychosocial risk signature, not to test causal hypotheses, predict an

outcome, or describe a real organization (Goncalves et al., 2020). Accordingly, the analysis is performed on a synthetic dataset constructed for this purpose, as described in Section 2.3. The overall workflow, from the simulated FPSICO style data through standardization, dimensionality reduction, clustering, validation, profile interpretation, and intervention planning (Rahman et al., 2026), is summarized in **Figure 1**.

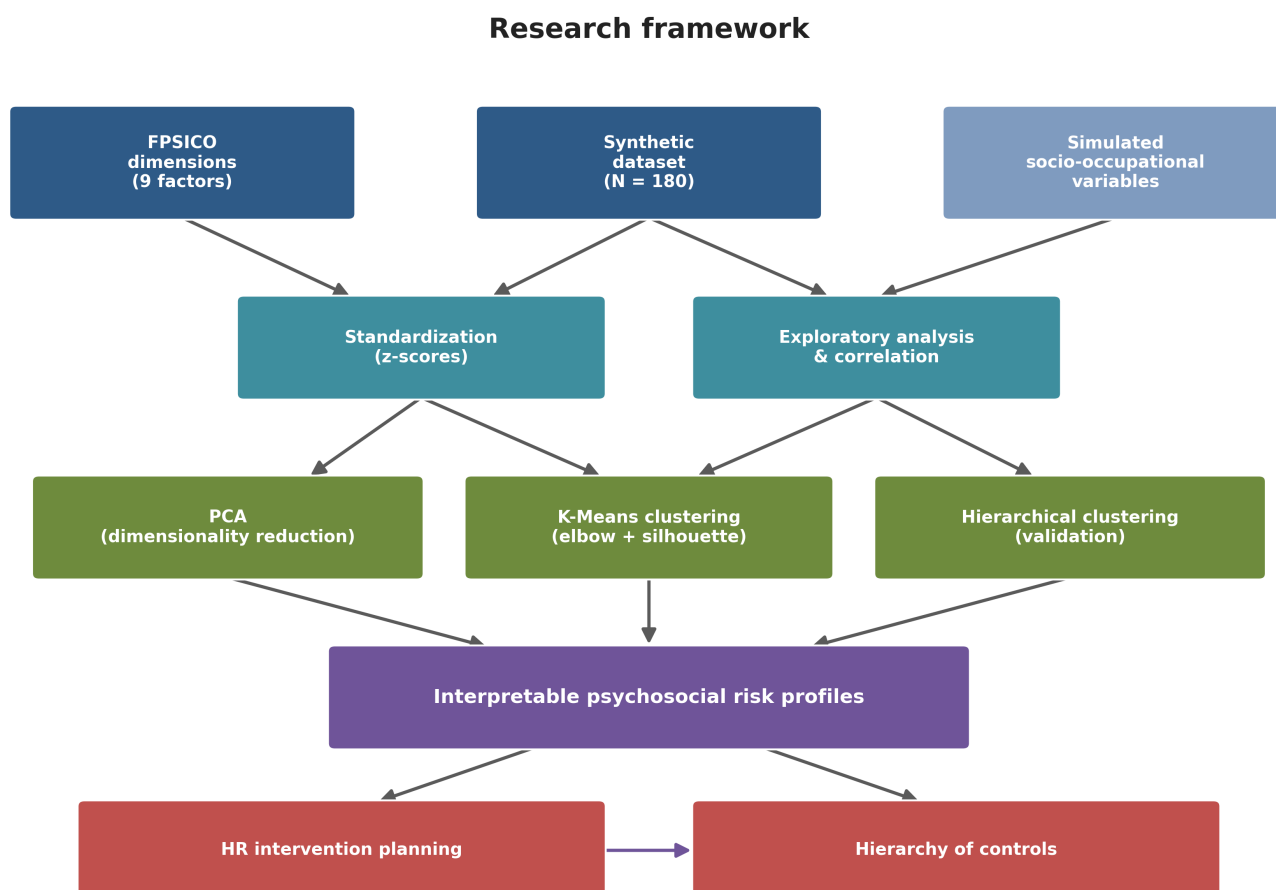


Figure 1. Methodological framework linking simulated FPSICO-style data to clustering and HR intervention planning.

2.2. FPSICO dimensions

The analysis is based on the nine dimensions of the FPSICO method (INSST technical guideline NTP 926), an instrument of 44 questions yielding 89 items whose nine-factor structure has been empirically validated on a sample of 1,718 workers (global Cronbach's $\alpha = 0.895$; confirmatory factor analysis indicating good fit, e.g., CFI = 0.935, RMSEA = 0.035). Each dimension is expressed here as a normalized risk index ranging from 0 to 100, where higher values indicate greater psychosocial risk, consistent with the ascending-risk logic of the FPSICO scoring system (Jachens, 2019). **Table 1** lists the dimensions, their interpretation, and their published internal-consistency coefficients.

Table 1. FPSICO dimensions, interpretation, and published reliability.

Code	Dimension	Interpretation (higher score = higher risk)	$\alpha^\dagger$
TT	Time at work	Demands related to working hours, rest, weekend/holiday work and their effect on personal life.	0.70

Continuation Table:

Code	Dimension	Interpretation (higher score = higher risk)	$\alpha^{\dagger}$
AU	Autonomy	Limited capacity to influence the pace, methods, breaks and scheduling of one's work.	0.87
CT	Workload	Excessive quantity and intensity of work and time pressure to complete tasks.	0.73
DP	Psychological demands	Cognitive and emotional demands required by the job.	0.74
VC	Variety / content of work	Low meaningfulness, variety, and perceived contribution of the work performed.	0.71
PS	Participation / supervision	Imbalance in participation in decisions and in the degree of supervision/control exerted.	0.73
ITC	Interest for the worker / compensation	Low organizational concern for development, stability, and fair compensation.	0.84
DR	Role performance	Role ambiguity and role conflict in job demands and responsibilities.	0.84
RAS	Relationships and social support	Poor quality of relationships and insufficient support from peers and supervisors.	0.72

Following the method's own logic, FPSICO classifies each factor into four risk levels derived from percentiles: adequate (percentile < P65), moderate (P65 to P75), high (P75 to P85), and very high ($\geq$ P85). Because the raw score range differs across factors (for example, time at work ranges from 0 to 37), the dimensions are not directly comparable on their native scales. Expressing each dimension as a 0 to 100 risk index, and subsequently standardizing using z scores, places the dimensions on a common footing so that none dominates the distance computations underlying clustering. Importantly, FPSICO was conceived to produce valuations at the group level for relatively homogeneous groups of workers and to support the identification of problem sources, the prioritization of actions, and comparisons across groups (Jachens, 2019). Clustering is in line with the objective of the instrument as it provides a data-driven way to create these homogeneous groups directly from risk data, without having to rely on predetermined OUs.

2.3. Synthetic dataset design and generation

The analysis uses a completely synthetic dataset of 180 simulated workers; no actual people, responses, or organizations are involved in order to clearly and reproducibly illustrate the process (Tucker et al., 2020). The dataset was created to mimic the output format of the FPSICO instrument and the structure of a medium-sized manufacturing workforce in order to test the pipeline under realistic circumstances while maintaining its shareability and lack of confidentiality restrictions.

Five predetermined latent psychosocial patterns were combined to create the data, each of which was characterized by a mean risk vector over the nine FPSICO dimensions (on a scale from 0 to 100) and a common dispersion. Gaussian noise was applied for each dimension, and the values were limited to the range of 0 to 100. The five motifs were purposefully made to stand out from one another. Elevated workload and psychological demands, for instance, were associated with one pattern; role conflict and poor recognition were associated with another; uniformly low risk

was associated with another; and uniformly high risk was associated with a third. The patterns were assigned the following approximate proportions of the workforce: balanced 28%, workload and demand 22%, low participation and support 20%, role conflict and recognition 18%, and generalized high risk 12%. Crucially, the generative labels were withheld from the entire unsupervised pipeline and retained only to provide a reference benchmark for validating profile recovery (Section 4.4).

Table 2. The synthetic dataset's generative parameters (mean risk index per dimension, 0–100 scale).

Perfil	Prop.	TT	AU	CT	DP	VC	PS	ITC	DR	RAS
Balanced	28%	20	19	22	21	20	21	20	19	18
Workload / psychological demand	22%	70	30	80	78	32	34	36	38	30
Low participation / weak support	20%	34	58	38	40	46	80	58	34	78
Role conflict / low recognition	18%	42	40	44	46	42	52	80	82	44
Generalized high risk	12%	74	70	80	82	68	78	80	76	76

Every dimension was restricted to $[0, 100]$ and derived from N (mean, $\sigma = 11$). 42 is the random seed. Proportions are not experimentally calibrated; they were selected for illustration.

Sociodemographic and occupational variables, including age, gender, department, job position, seniority, work shift, and contract type, were simulated alongside the dimension scores to improve descriptive realism (Dankar et al., 2022). These variables were not used in any clustering step; they serve only to illustrate how, in a real deployment, profiles could subsequently be compared across the organization. The dispersion parameter was tuned so that the recovered structure exhibited realistic, rather than perfect, separability (mean silhouette of 0.50 rather than near perfect values), thereby avoiding an artificially easy clustering task. A fixed random seed makes the dataset and all results exactly reproducible from the accompanying code.

2.4. Dataset and variables

Two sets of variables were employed. Age, gender, department, job position, seniority, work shift, and contract type are examples of simulated sociodemographic and occupational variables that were kept just for descriptive purposes and excluded from the clustering procedure. For all segmentation analyses, the analytical feature set consisted of the nine FPSICO dimension risk indices. This division is intentional since it guarantees that profiles are solely based on psychosocial risk content, after which the descriptive factors can be analyzed to show how each profile might focus inside an organization in a genuine deployment (Yapi et al., 2026). **Table 3** summarizes the variables:

Table 3. Dataset variables (all values simulated).

Variable	Type	Role in analysis
WorkerID	Identifier	Synthetic record key
Age	Numeric	Descriptive only
Gender	Categorical	Descriptive only

Continuation Table:

Variable	Type	Role in analysis
Department	Categorical	Descriptive only
Job position	Categorical	Descriptive only
Seniority (years)	Numeric	Descriptive only
Work shift	Categorical	Descriptive only
Contract type	Categorical	Descriptive only
9 FPSICO dimensions (TT, AU, CT, DP, VC, PS, ITC, DR, RAS)	Numeric (0–100)	Clustering features

2.5. Machine learning methods

NumPy, pandas, and scikit-learn were used in all Python studies. The following steps made up the pipeline. (1) Standardization: To ensure that each dimension contributed equally to distance calculations, the nine dimensions were transformed to z-scores (Jolliffe & Cadima, 2016). (2) Exploratory analysis: the data was described using a Pearson correlation matrix, score distributions, and descriptive statistics. (3) Dimensionality reduction: PCA was used to visualize the data and identify the prominent axes of variation; components were kept using the cumulative-variance rule and the Kaiser criterion (eigenvalue ≥ 1). (4) Clustering: K-Means (k ranging from 2 to 8, with 50 initializations) was the primary method (Sarker, 2021). The number of clusters was ascertained using the Calinski Harabasz index, the Davies Bouldin index, the average silhouette coefficient, the elbow technique (within cluster sum of squares), and interpretability as a tiebreaker. (5) Validation: Using the adjusted Rand index (ARI), Ward hierarchical clustering was carried out separately and contrasted with the K-Means split. (6) Profiling: Cluster centroids were examined using the original 0–100 scale, profiles were named based on their risk signatures, and a one-way ANOVA for each dimension verified that the dimensions varied among clusters. A fixed random seed was used to guarantee reproducibility (Raschka et al., 2022).

3. Results.

3.1. Descriptive statistics and correlations

With significant dispersion across all dimensions (standard deviations ranging from roughly 18 to 26), the mean dimension risk indices for the simulated workforce ranged from 36.7 (variety and content) to 49.8 (psychological demands) on the 0 to 100 scale, reflecting the heterogeneous exposure incorporated into the generative design. **Table 4** reports the descriptive statistics.

Table 4. Descriptive statistics of FPSICO dimension risk indices (0–100).

Dimension	Mean	SD	Min	Q1	Median	Q3	Max
TT – Time at work	45.75	22.91	0.0	25.96	41.84	66.77	92.41
AU – Autonomy	39.46	20.82	0.0	21.67	37.83	57.64	93.12
CT – Workload	48.98	25.49	4.58	30.20	42.19	71.61	100.0

Continuation Table:

Dimension	Mean	SD	Min	Q1	Median	Q3	Max
DP – Psychological demands	49.77	24.46	3.07	29.18	49.14	68.58	100.0
VC – Variety / content	36.71	18.41	0.0	24.37	36.21	48.53	83.86
PS – Participation / supervision	49.59	26.41	0.13	29.14	46.01	72.85	100.0
ITC – Interest / compensation	49.42	25.92	1.26	28.80	47.32	69.42	100.0
DR – Role performance	44.51	26.30	1.62	24.62	38.41	67.13	100.0
RAS – Relationships / support	44.53	26.46	0.0	23.60	38.05	68.22	100.0

Three coherent blocks were found in the correlation matrix (**Figure 2**). Time spent at work, workload, and psychological demands were all connected by a demand block (pairwise correlations of 0.81–0.84). A control-and-support block linked autonomy, participation/supervision, and relationships/social support (e.g., participation–support $r = 0.85$, autonomy–support $r = 0.78$). A recognition–role block linked interest/compensation with role performance ($r = 0.80$). These blocks are in line with the theoretical division between needs, resources, and incentives and hint at the structure that was subsequently retrieved by PCA and clustering.

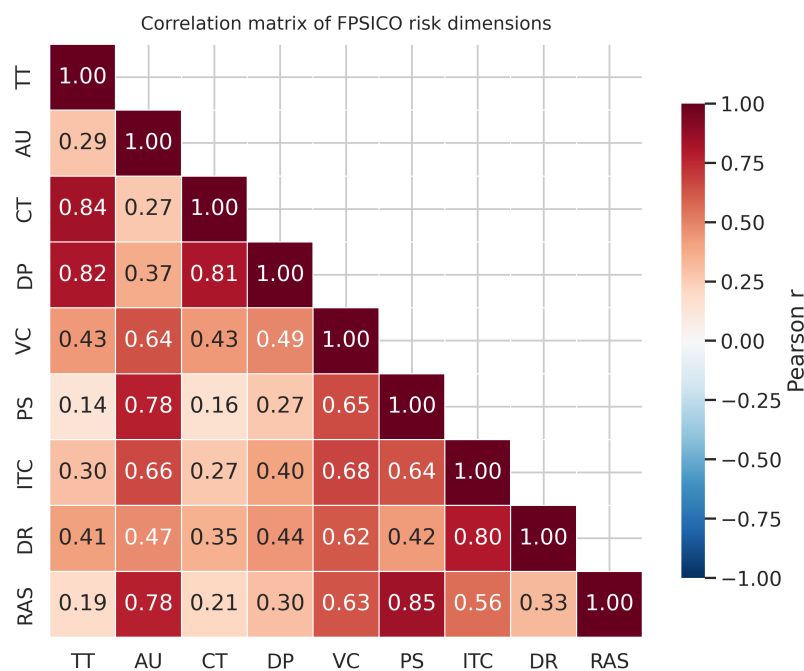


Figure 2. Pearson correlation matrix of the nine FPSICO risk dimensions.

3.2. Principal component analysis

Two components with eigenvalues larger than 1 were found using PCA of the nine standardized FPSICO dimensions. These factors collectively accounted for 77.4% of the variation (**Table 5**, **Figure 3**). An overall psychosocial-severity axis is represented by the first component (55.2%), which loaded positively on all nine dimensions. Employees with high PC1 scores are generally at higher risk. In order to capture a demand-versus-support axis, the second component (22.2%) compared

the demand dimensions (time at work, workload, psychological demands; positive loadings) against the control-and-support dimensions (participation, relationships/support, autonomy; negative loadings). A third, lesser component was related to recognition and role performance. In addition to making visualization easier, this low-dimensional structure verifies that the nine dimensions are not independent.

Table 5. PCA explained variance.

Component	Eigenvalue	Explained variance (%)	Cumulative (%)
PC1	4.999	55.24	55.24
PC2	2.009	22.19	77.43
PC3	0.866	9.57	87.01
PC4	0.333	3.68	90.69
PC5	0.228	2.52	93.21
PC6	0.182	2.01	95.22
PC7	0.160	1.76	96.98
PC8	0.141	1.56	98.54
PC9	0.132	1.46	100.00

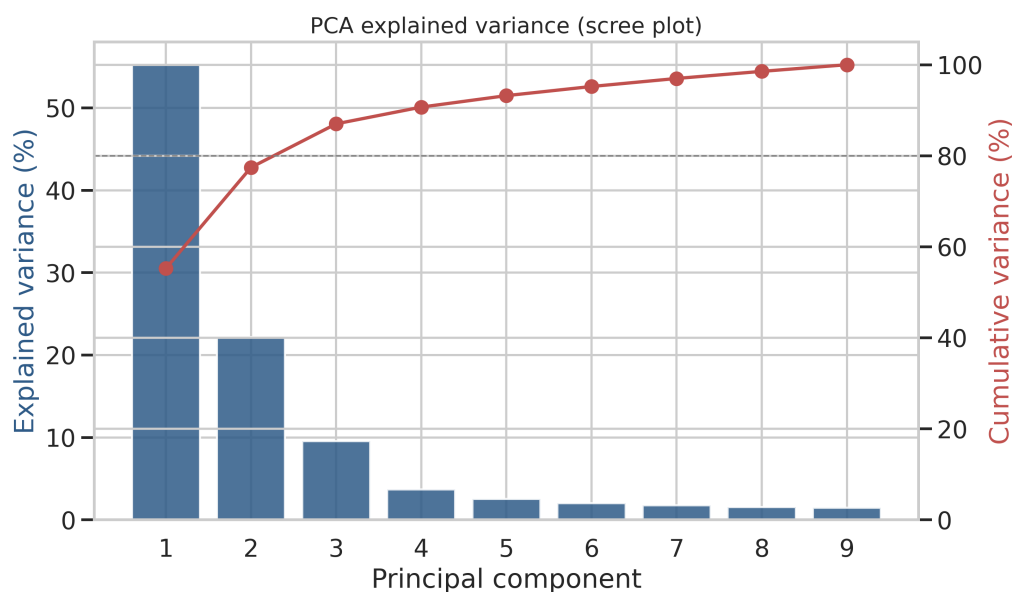


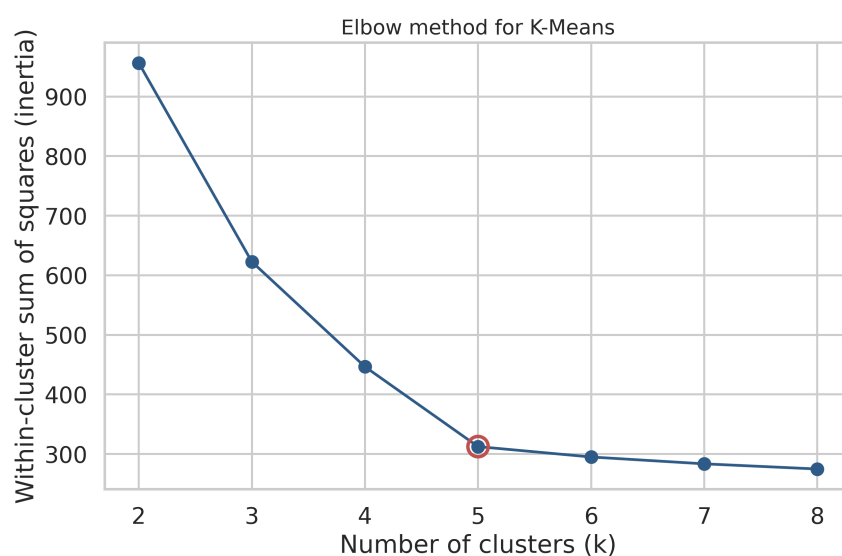
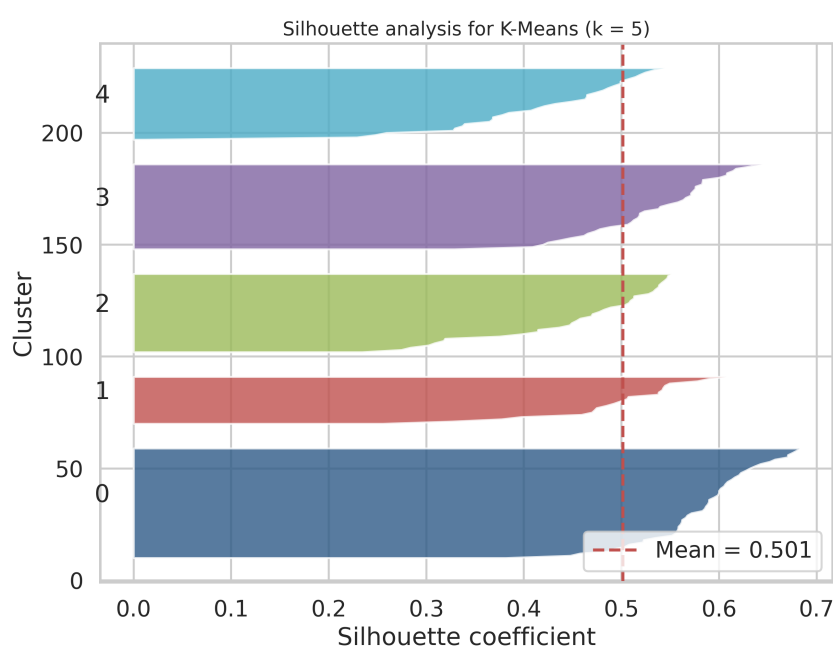
Figure 3. PCA scree plot: individual and cumulative explained variance.

3.3. Selecting the number of clusters

K-Means was run for $k = 2$ to 8. The within-cluster sum of squares decreased markedly up to $k = 5$ and then flattened, indicating an elbow at five clusters (**Figure 4**). The average silhouette coefficient peaked at $k = 5$ (0.50), the Calinski Harabasz index was maximal at $k = 5$ (183.1), and the Davies–Bouldin index was minimal at $k = 5$ (0.79). All four criteria converged on a five-cluster solution (**Table 6**). The per-cluster silhouette diagram (**Figure 5**) showed positive silhouettes across all clusters with few misassigned points, supporting a clean partition.

Table 6. Clustering performance across candidate values of k (K-Means).

k	Inertia (WCSS)	Silhouette	Calinski–Harabasz	Davies–Bouldin
2	956.46	0.389	123.49	0.889
3	622.68	0.429	141.75	0.908
4	446.52	0.468	154.18	0.902
5	312.45	0.501	183.09	0.794
6	295.02	0.386	156.29	1.185
7	283.54	0.300	135.91	1.516
8	274.85	0.222	120.25	1.806

**Figure 4.** Elbow method: within-cluster sum of squares versus k.**Figure 5.** Silhouette analysis for the five-cluster K-Means solution (mean = 0.50).

3.4. Cluster structure and validation

Projected onto the first two principal components, the five clusters occupied clearly separated regions (**Figure 6**): the balanced cluster sat at the low-severity end of PC1, the generalized-high cluster at the opposite end, while the demand and support profiles separated vertically along PC2, exactly as the loadings predict. When compared to the K-Means solution, Ward hierarchical clustering yielded an essentially comparable partition (silhouette = 0.50) with an adjusted Rand index of 0.987. This agreement shows that the recovered partition for the current synthetic dataset is widely reproduced by Ward hierarchical clustering and is not exclusive to the K-Means implementation. Since the dataset was created using a known five-profile design, it provides an additional validation reference that is only possible because the underlying synthetic structure is known. An adjusted Rand index of 0.99 was obtained by comparing the recovered clusters with the generating labels, indicating that the pipeline nearly precisely recovers the known structure. The main methodological outcome is this retrieval of the known structure. The pipeline retrieved the planted structure with great fidelity under the current simulation's settings.

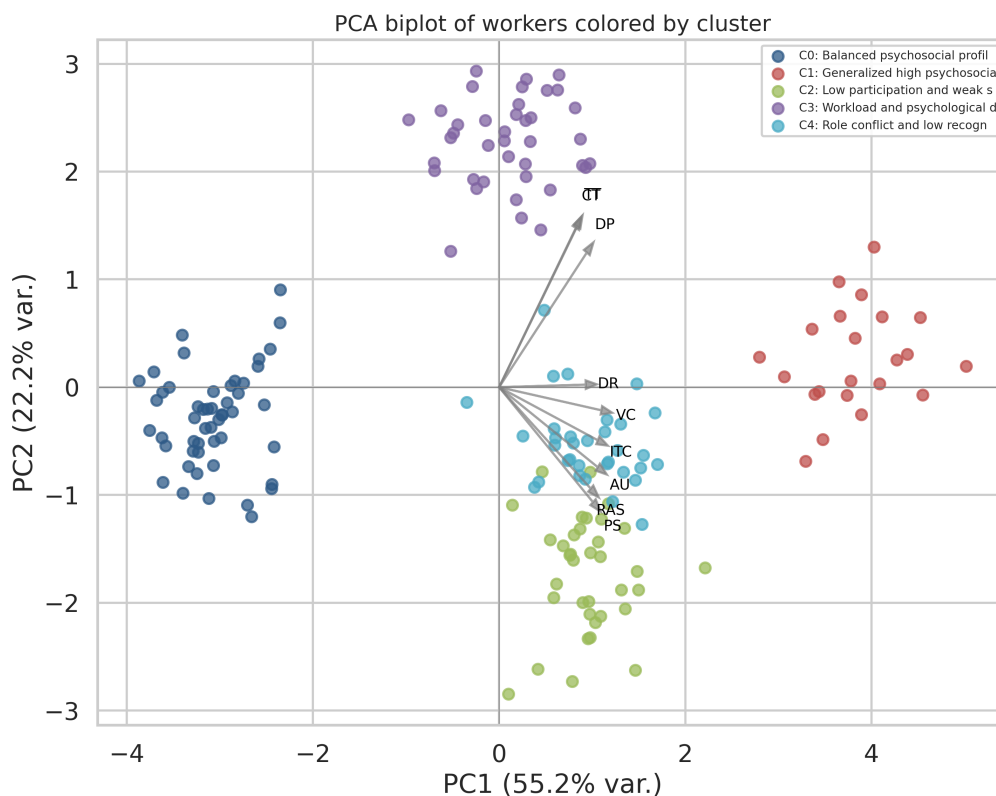


Figure 6. PCA biplot of workers colored by cluster, with dimension loading vectors.

3.5. Psychosocial risk profiles

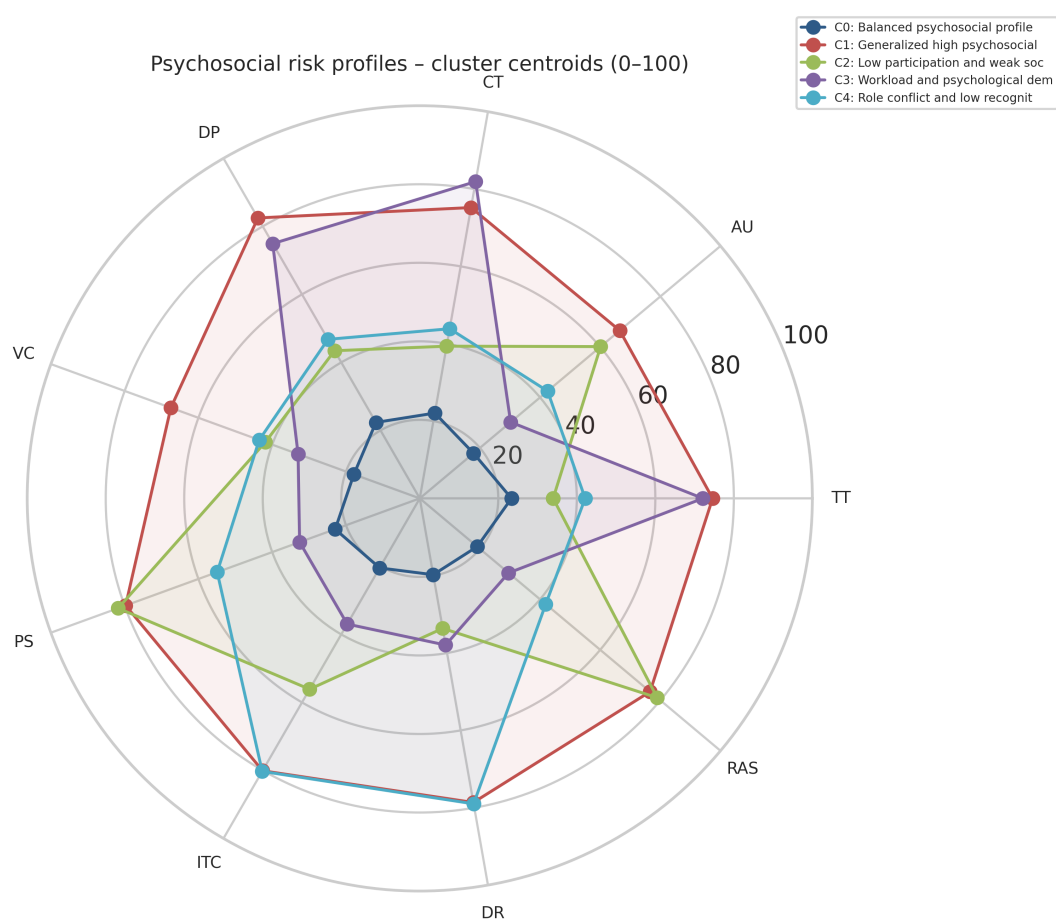
The five clusters were interpreted from their centroids on the original 0–100 scale (**Table 7**) and named accordingly (**Table 8**). As expected, when groups are defined by a clustering algorithm, a one-way ANOVA found that every dimension differed across clusters (all $F \geq 95$, $p < 0.001$); this serves as a descriptive confirmation of separation rather than an independent hypothesis test. The radar chart (**Figure 7**) visualizes the distinct risk signatures.

Table 7. Cluster centroids on the 0–100 risk scale (and cluster size).

Cluster (n, %)	TT	AU	CT	DP	VC	PS	ITC	DR	RAS
C0 (50, 27.8%)	23.4	17.9	22.1	22.3	17.9	22.9	20.5	19.7	19.1
C1 (22, 12.2%)	74.6	66.5	75.2	82.5	67.5	79.7	80.1	78.6	76.6
C2 (36, 20.0%)	34.0	60.1	39.4	43.4	41.9	81.8	56.1	33.6	79.0
C3 (39, 21.7%)	72.1	30.2	81.9	74.9	33.0	32.6	37.0	37.9	29.5
C4 (33, 18.3%)	42.1	42.5	43.9	46.8	43.5	54.9	80.3	79.0	41.9

Table 8. Psychosocial risk profiles.

Cluster	Profile name	Defining (high) dimensions	Prevalence
C0	Balanced psychosocial profile	None elevated; uniformly low risk	27.8%
C3	Workload and psychological demand profile	Workload, psychological demands, time at work	21.7%
C2	Low participation and weak social support profile	Participation/supervision, relationships/support, low autonomy	20.0%
C4	Role conflict and low recognition profile	Role performance, interest/compensation	18.3%
C1	Generalized high psychosocial risk profile	All nine dimensions elevated	12.2%

**Figure 7.** Radar chart of cluster centroids across the nine FPSICO dimensions.

In substantive terms, the balanced profile (27.8% of workers) shows uniformly low risk and requires maintenance rather than remediation. The workload and psychological demand profile (21.7%) is defined by high workload, psychological demands, and time pressure with otherwise moderate support—classic high-demand work. The low participation and weak social support profile (20.0%) combines weak participation, poor relationships/support, and limited autonomy, while demands remain moderate—a resource-deficit pattern. The role conflict and low recognition profile (18.3%) is marked by role ambiguity/conflict and low recognition and compensation—an effort–reward imbalance signature. The generalized high-risk profile (12.2%), though the smallest group, is elevated on every dimension and represents the priority population for intervention.

4. Discussion

The demonstration shows that, when applied to data with a genuine latent structure, the proposed pipeline organizes the nine FPSICO dimensions into a small number of coherent and interpretable psychosocial risk profiles and recovers that structure with high fidelity. The PCA structure based on two components, consisting of an overall severity axis and a contrast between demands and support, was deliberately incorporated into the generative design to reflect established theoretical frameworks. The severity axis corresponds to cumulative exposure, while the contrast between demands and support reflects the distinctions proposed by the demand control model and the job demands resources model. Likewise, the role and recognition pattern reflect the effort reward imbalance perspective. The convergence of four complementing criteria for cluster selection on a five-cluster solution, together with the highly similar hierarchical clustering solution and the nearly perfect recovery of the generative labels, shows that, under the given generating conditions, the pipeline operates as expected.

When applied to actual data, this kind of segmentation has the practical benefit of turning evaluation results into a foundation for prioritization. While profiles show which combinations of components occur together, how common each configuration is, and which employees share a similar risk signature, average dimension scores tell HR teams which factors are high. There may be a significant role conflict group and a distinct workload group in a department with moderate average scores, which call for rather different solutions. Because profiles highlight these variations, they facilitate focused and economical responses. Because they are based on simulated profiles, these implications are illustrative in the current study; their usefulness depends on using the pipeline on actual organizational data.

Crucially, clustering enhances FPSICO rather than replacing it. The unsupervised analytical layer rearranges the verified measurement and the regulatory diagnostic provided by the instrument into an actionable framework. The methodology is descriptive and exploratory. The profiles should be viewed as a decision support tool rather than as a clinical or diagnostic categorization of persons, and clusters characterize patterns of co-occurrence rather than causative linkages.

5. Pathway to Empirical Validation

Future empirical validation should apply the proposed pipeline to actual organizational FPSICO datasets collected under appropriate ethical and data-protection conditions. Before clustering, the data should undergo systematic quality checks, including verification of completeness, detection of inconsistent

responses, treatment of missing variables, and normalization of the psychosocial dimensions. The objective of this stage is to ensure that the resulting profiles reflect meaningful organizational patterns rather than artifacts produced by data-quality problems.

No universal minimum sample size guarantees a stable clustering solution. Instead, adequacy should be evaluated empirically through multiple initializations, bootstrap resampling, sensitivity analyses, and assessment of profile stability across alternative analytical specifications. Future applications should examine whether the number, composition, and characteristics of the identified clusters remain consistent under reasonable variations in preprocessing decisions and clustering parameters.

Missing values should be evaluated before clustering, considering appropriate imputation procedures and sensitivity analyses to determine whether missingness affects profile stability and interpretation. Real organizational implementations also require anonymization or pseudonymization of worker information, secure data governance procedures, and compliance with applicable ethical and data-protection requirements. These safeguards are essential to preserve confidentiality and prevent the misuse of psychosocial assessment results.

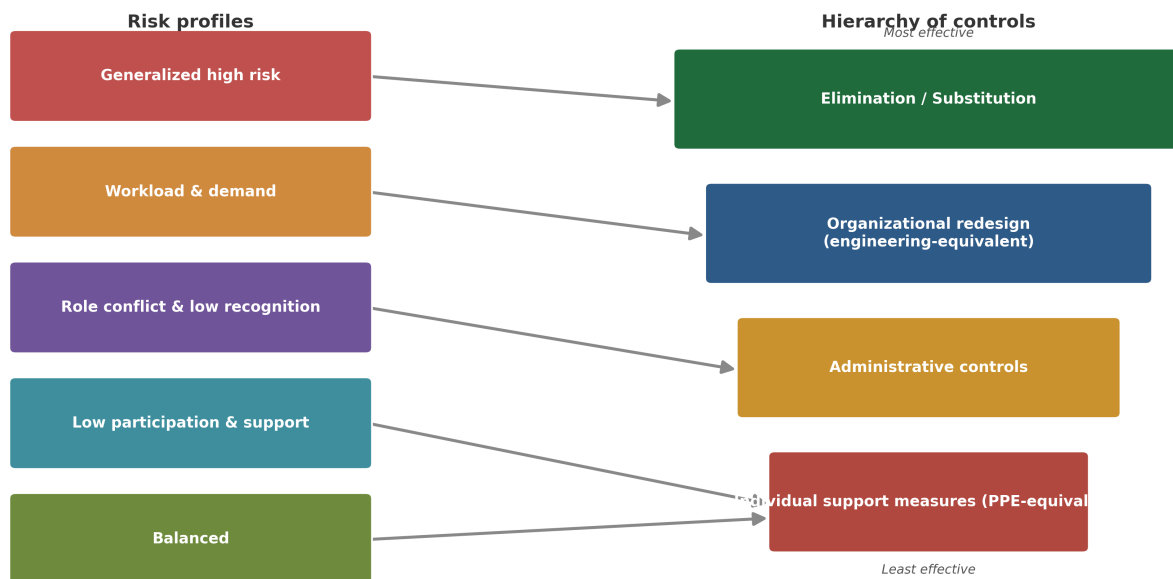
The interpretation and use of the generated profiles should involve HR professionals, occupational health specialists, and employee representatives. The identified clusters should be understood as decision-support patterns that help recognize combinations of psychosocial exposures rather than as diagnostic labels assigned to individual workers. Adequate training is therefore required to ensure that the profiles support preventive organizational actions and participatory intervention planning.

The five profiles identified in this synthetic demonstration should not be regarded as universal categories. Their number, structure, and practical meaning may differ across industries, occupational groups, countries, and organizational environments. Therefore, external validation with real organizational data is required before generalizing these profiles, and future applications should adapt the interpretation of clusters according to the specific social and operational context in which the assessment is conducted.

6. HR Intervention Planning

Each recovered profile is converted into an illustrative, prioritized intervention plan using the hierarchy of controls adapted to psychosocial risk, where elimination and substitution rank highest, followed by organizational/work redesign (the engineering control equivalent), administrative controls, and individual support measures, to demonstrate how the methodology informs decision-making (**Figure 8**). A caveat specific to psychosocial factors must be stated: unlike physical hazards, these factors are not simply to be removed but optimized, since well-designed demands, autonomy, and social relationships are beneficial. Accordingly, the elimination and substitution level are interpreted in this context as removing the adverse situation or its source, for instance, by eliminating competing mandates or unnecessary time pressure, while maintaining and optimizing the underlying element. The guiding principle is that higher risk profiles require upstream organizational action, and individual measures should complement, rather than replace, structural change. **Table 9** presents the intervention matrix as a reusable template that an organization would populate from its own profile results.

Intervention planning framework: profiles → hierarchy of controls



Higher-risk profiles prioritize upstream, organization-level controls; individual measures complement but never replace them.

Figure 8. Intervention planning framework: mapping profiles to the hierarchy of controls.

Table 9. HR intervention matrix based on the hierarchy of controls.

Profile	Elimination / Substitution	Organizational redesign	Administrative controls	Individual support
Balanced	Preserve protective conditions	Maintain current job design; share good practices	Periodic reassessment; monitor for drift	Voluntary well-being resources
Workload & psychological demand	Remove redundant tasks; eliminate unnecessary deadlines	Rebalance workload; adjust staffing/pace; review shift design	Workload monitoring; realistic planning; recovery breaks	Stress-management and time-management support
Low participation & weak social support	Remove barriers to participation in decisions	Redesign for greater autonomy; strengthen teamwork and supervisor support structures	Participatory committees; communication channels; supervisor accountability	Coaching; team-building; peer support networks
Role conflict & low recognition	Eliminate conflicting/overlapping mandates	Clarify roles and responsibilities; redesign recognition and compensation systems	Clear job descriptions; transparent appraisal and reward criteria	Career development and feedback for affected staff
Generalized high risk	Prioritize and tackle the most severe hazards first	Comprehensive redesign across demands, control, support and rewards	Action plan with owners, deadlines and follow-up audits	Targeted health surveillance and support; referral pathways

This matrix would give HR managers a justifiable foundation for allocating limited resources when applied to actual data. While the support and recognition profiles necessitate the reform of participation and incentive systems, the generalized high risk and demand-related profiles usually support prompt structural action. Monitoring is mostly necessary for the balanced group in order to identify any possible decline. The scale of intervention is also influenced by profile predominance. About one in eight of the simulated employees in this example fit the broad high-risk profile, a sign that would require priority investment in a real business.

7. Limitations

This study's main drawback stems from its proof-of-concept nature. The profiles, their proportions, and the descriptive correlations derived from them are products of the generative model, and the dataset is completely synthetic. As a result, there is no empirical assertion made by the data regarding the configuration or prevalence of psychosocial risk in any actual workforce. Therefore, rather than being proof of organizational psychosocial factors, the study should be seen as a methodical demonstration of the behavior and repeatability of the suggested pipeline.

The fact that the simulated profiles were produced using preset mean vectors and a similar dispersion value is a related drawback. The results are nonetheless better controlled than actual organizational evaluations, even if Gaussian variation was included and the profile sizes were purposefully uneven. Measurement inaccuracy, missing values, response biases, unusual observations, unbalanced group sizes, non-normal distributions, and significantly overlapping patterns of psychosocial exposure are all possible in real FPSICO data. As a result, it is not appropriate to interpret the silhouette coefficient, modified Rand index, and other cluster-selection values presented in this work as anticipated performance levels in practical applications.

More border observations and less distinct clusters would result from greater overlap in empirical data. Cluster membership should not be used as a precise, long-term, or diagnostic categorization in such circumstances. Employees that work close to cluster borders may have traits in common with many profiles. Instead of relying just on a hard cluster label, their interpretation should take into account the initial FPSICO dimension scores, distances to cluster centroids, sensitivity studies, and, when relevant, probabilistic membership estimations. Therefore, equivocal circumstances should be treated by HR and occupational health specialists as mixed exposure patterns rather than putting people into strict classifications.

There are more methodological restrictions when using K-Means. When clusters are roughly spherical and compact in the standardized feature space, separated by Euclidean distance, and have about equal dispersion, the method works well. Instead, real psychosocial profiles might be extensive, unevenly distributed, layered, non-convex, or impacted by outliers. Additionally, K-Means allocates each observation to a single cluster, even in cases when the separation is poor, and demands that the number of clusters be predetermined. Multiple initializations, four complementing cluster-selection criteria, and comparison with Ward hierarchical clustering were used in the current work to partially solve these constraints. However, global robustness is not established by agreement amongst approaches on synthetic data.

When the observed data structure indicates that the K-Means assumptions might not be met, future empirical applications should examine other clustering techniques. When probabilistic membership and uneven covariance structures are important, Gaussian mixture models could be suitable. In addition to revealing

layered correlations between profiles, hierarchical clustering may facilitate exploratory interpretation. Although the effectiveness of density-based techniques like DBSCAN is highly dependent on parameter selection and may be restricted in moderate-dimensional data, they may be able to detect irregular structures and noise observations. When several approaches yield conflicting conclusions, consensus or ensemble clustering may offer further proof of stability. Instead of being chosen arbitrarily, these options should be assessed based on cluster stability, substantive interpretability, theoretical coherence, and practical use.

Another drawback is interpretation. Clusters may not match theoretically identifiable profiles as well in actual organizational data as they do in this simulation. While two other clusters may differ primarily in severity rather than exposure type, a cluster may combine high demands, poor support, and low recognition. As a result, profile names should continue to be descriptive rather than diagnostic. Occupational health specialists, HR specialists, managers, employee representatives, and, where applicable, employees themselves should all participate in the interpretation process. Before choosing treatments, it is important to take into account the cluster centroids, within-cluster variability, subgroup composition, and qualitative organizational evidence.

Lastly, rather than describing causal linkages, the cross-sectional study reveals patterns of co-occurrence. FPSICO ratings would also be self-reported and susceptible to response and common-method biases in an actual application. Empirical validation utilizing ethically gathered organizational data is a crucial next step. Future research should look at how stable the profiles are across samples, time periods, industries, and nations; connect them to pertinent outcomes like performance, absenteeism, turnover intention, and health indicators; and assess whether profile-guided interventions result in significant organizational improvements.

A further limitation concerns the assumptions underlying K-Means clustering. K-Means performs best when clusters are relatively compact, approximately spherical, and have comparable variance. These assumptions may not be fully satisfied in real psychosocial datasets, where risk profiles are likely to overlap and workers may occupy intermediate positions between groups. Therefore, future applications should evaluate the stability of the solutions using alternative clustering approaches, such as Gaussian mixture models, DBSCAN, hierarchical clustering, or ensemble clustering strategies. Ambiguous cases should not be interpreted as definitive categories, but rather as transitional profiles requiring contextual interpretation by HR professionals.

8. Conclusions

This paper proposed a reproducible, unsupervised machine-learning methodology for transforming FPSICO-style psychosocial assessment data into an interpretable, decision-ready map of worker risk, and demonstrated its behavior on a fully synthetic dataset. Under the specified generative conditions, PCA recovered an overall-severity axis and a demand-versus-support contrast, and K-Means—validated by hierarchical clustering, four complementary cluster-selection criteria, and near-perfect recovery of the generative design (adjusted Rand index = 0.99)—identified five interpretable psychosocial risk profiles: balanced, workload and psychological demand, low participation and weak social support, role conflict and low recognition, and generalized high risk. Mapping each profile to the hierarchy of controls produced a reusable intervention-planning template that would help HR and occupational health teams prioritize upstream, organization-level action and target individual support

appropriately. The contribution is clearly methodological and complementary: the pipeline adds an analytical layer to FPSICO rather than replacing or predicting it, and because both the data and the code are synthetic and publicly available, the entire study can be replicated and audited in its entirety. Applying the same pipeline to actual, ethically collected organizational data is a logical and essential next step, after which the strategy can transition from proven viability to validated practice.

Author contributions: Conceptualization, J.E.N. and M.D.A.A.; methodology, J.E.N.; software, J.E.N.; validation, J.E.N. and M.D.A.A.; formal analysis, J.E.N.; investigation, J.E.N.; resources, J.E.N. and M.D.A.A.; data curation, J.E.N.; writing—original draft preparation, J.E.N.; writing—review and editing, J.E.N. and M.D.A.A.; visualization, J.E.N.; supervision, J.E.N.; project administration, J.E.N.; funding acquisition, M.D.A.A. All authors have read and agreed to the published version of the manuscript.

Acknowledgment: The authors would like to express their sincere gratitude to the Universidad Técnica de Cotopaxi (UTC) for its institutional support. The authors also acknowledge the Research Group AIOLOS of the Universidad Técnica de Cotopaxi and the INTELIA Research Network for their valuable academic collaboration, scientific exchange, and support in the development of this research.

Conflict of interest: The authors declare no conflict of interest.

Data Availability Statement: The dataset analyzed in this study is entirely synthetic and contains no real personal or organizational data.

Ethics Statement: This study used exclusively synthetic, computer-generated data and involved no human participants or processing of personal information.

References

1. Alali, M., Masood, M. A., Khaled, R., AlQaydi, S., Semerjian, L., Ahmed, F. R., Maharani, C., & Abass, K. (2025). Psychosocial risks and occupational challenges across workplace sectors in Gulf cooperation council countries. *Discover Social Science and Health*. <https://doi.org/10.1007/s44155-025-00343-w>
2. Amoadu, M., Ansah, E. W., & Sarfo, J. O. (2023). Influence of psychosocial safety climate on occupational health and safety: a scoping review. *BMC Public Health*, 23(1). <https://doi.org/10.1186/s12889-023-16246-x>
3. Bennett, A., Dodd, W., Yeung, Y., & Elliott, S. J. (2026). Characterizing the health impacts of artisanal and small-scale gold mining (ASGM) across the life course: an umbrella review. In *Environmental Research Letters* (Vol. 21, Number 11). Institute of Physics. <https://doi.org/10.1088/1748-9326/ae7130>
4. Dankar, F. K., Ibrahim, M. K., & Ismail, L. (2022). A Multi-Dimensional Evaluation of Synthetic Data Generators. *IEEE Access*, 10, 11147–11158. <https://doi.org/10.1109/ACCESS.2022.3144765>
5. Donohoe, E., Camonita, F., Tagueo -Wise, V., Guey, A. C., Zejerman, I., Todaro -Ey, L. L., Godderis, A., & Boone -Ku, L. (2024). Mental health in the construction sector: preventing and managing psychosocial risks in the workplace Report European Agency for Safety and Health at Work Mental health in the construction sector: preventing and managing psychosocial risks in the workplace European Agency for Safety and Health at Work-EU-OSHA 2. *Journal: European Agency for Safety and Health at Work (EU-OSHA)*. <https://doi.org/10.2802/7183661>
6. Goncalves, A., Ray, P., Soper, B., Stevens, J., Coyle, L., & Sales, A. P. (2020). Generation and evaluation of synthetic patient data. *BMC Medical Research Methodology*, 20(1). <https://doi.org/10.1186/s12874-020-00977-1>
7. Jachens, L. (2019). Humanitarian aid workers' mental health and duty of care. In *Europe's Journal of Psychology* (Vol. 15, Number 4, pp. 650–655). PsychOpen. <https://doi.org/10.5964/ejop.v15i4.2221>
8. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 374, Number 2065). Royal

- Society of London. <https://doi.org/10.1098/rsta.2015.0202>
9. Lot Tanko, B. (2020). *International Journal of Real Estate Studies INTREST Compliance with the Use of Personal Protective Equipment (PPE) on Construction Sites in Johor, Malaysia*. <https://www.researchgate.net/publication/365426368>
 10. Lunn, R. M., Helmick, K. R., Cooney, G., Clemons, M., Polansky, M., Snow, S. J., Tracy, W., & Dixon, D. (2026). CVD Environmental Health Disparities Tool: systematic evidence mapping of psychosocial stressors, environmental exposures, and cardiovascular diseases to inform disparities research and action. *Environmental Health*, 25(1), 28. <https://doi.org/10.1186/s12940-026-01269-9>
 11. Nebbs, A., Martin, A., Neil, A. L., Broughton, K., Dawkins, S., & Roydhouse, J. (2025). Implementing the ISO 45003 standard: An approach combining implementation teams and a facilitated planning process. *Safety Science*, 191. <https://doi.org/10.1016/j.ssci.2025.106961>
 12. Neuhaus, C. A., Bellingrath, S., Müller, A., Taibi, Y., & Metzler, Y. A. (2026). Towards integrated hazard assessment: Convergence of objective and subjective work environment assessments in manufacturing. *Journal of Safety Research*, 97, 74–85. <https://doi.org/10.1016/j.jsr.2026.02.003>
 13. Rahman, M. I., Hossain, M. R., Sayem, S. M., & Emdad, F. Bin. (2026). Exploring the role of synthetic data in the future of AI in healthcare: A scoping review of frameworks, challenges, and implications. In *Intelligence-Based Medicine* (Vol. 13). Elsevier B.V. <https://doi.org/10.1016/j.ibmed.2025.100342>
 14. Raschka, Sebastian., Liu, Yuxi., Mirjalili, Vahid., & Dzhuulgakov, Dmytro. (2022). *Machine learning with Pytorch and Scikit-Learn : develop machine learning and deep learning models with scikit-learn and PyTorch*. Packt Publishing, Limited.
 15. Rivera Armendariz, C. A., & Banning, A. (2026). Integrating LA–ICP–MS into environmental geosciences: Fundamentals, emerging applications, technological developments, and future outlook. In *Spectrochimica Acta - Part B Atomic Spectroscopy* (Vol. 242). Elsevier B.V. <https://doi.org/10.1016/j.sab.2026.107559>
 16. Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. In *SN Computer Science* (Vol. 2, Number 3). Springer. <https://doi.org/10.1007/s42979-021-00592-x>
 17. Tucker, A., Wang, Z., Rotalinti, Y., & Myles, P. (2020). Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *Npj Digital Medicine*, 3. <https://doi.org/10.1038/S41746-020-00353-9>
 18. Wang, Y., Wilchek, M., & Batarseh, F. A. (2026). From data to policy: a systematic review of AI in water regulations and compliance. *Npj Clean Water*, 9(1). <https://doi.org/10.1038/s41545-026-00555-w>
 19. Yapi, A., Latouche, P., Guillin, A., & Bailly, Y. (2026). A new machine learning framework for occupational accidents forecasting with safety inspections integration. *Journal of Safety Research*, 97, 244–260. <https://doi.org/10.1016/j.jsr.2026.03.004>