© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

Clustering Of Individuals by Health Literacy and Risk Perception Patterns Using Unsupervised AI to Characterise Non-Prescription Medication Use and Care-Seeking Delays

OLUWADARE IBUKUN JUMOKE[1], Cecilia Bukola BELLO[2]
[1,2]*Faculty of Nursing Sciences, Afe Babalola University, Ado-Ekiti, Ekiti State, Nigeria.*

Abstract- Non-prescription (over-the-counter) medication use is prevalent worldwide, yet inappropriate self medication contributes to delayed diagnosis, adverse drug events, and increased healthcare costs. Individual health literacy and perceived risk of self medication are key behavioral determinants, but their combined patterns remain poorly understood. This study applies unsupervised machine learning to cluster individuals based on their health literacy levels and risk perception profiles, and then characterizes each cluster's non-prescription medication use and healthcare-seeking delays. A systematic review of 48 studies (2015–2025) examining health literacy, risk perception, self medication practices, and unsupervised clustering methods was conducted. Findings reveal four distinct clusters: (1) High literacy: Low risk perception (frequent inappropriate self medication, delayed care); (2) Low literacy: High risk perception (avoidant behaviour, unnecessary delays, and anxiety); (3) Moderate literacy: Balanced risk perception (appropriate short term use, timely care seeking); and (4) Low literacy: Low risk perception (high risk overuse, very long delays, and frequent adverse outcomes). Unsupervised AI techniques, particularly k-means, hierarchical clustering, and latent profile analysis, effectively identify these subgroups. The study demonstrates that clustering by behavioral phenotypes enables targeted interventions, such as literacy sensitive education, pharmacy based screening, and digital nudges, to reduce inappropriate self medication and delays in care. AI driven clustering offers a scalable approach for public health stratification and personalized risk communication.

Keywords: Health literacy, risk perception, non prescription medication, self medication, care seeking delay, unsupervised AI, clustering, k means, latent profile analysis.

I. INTRODUCTION

Self medication with non prescription (over the counter, OTC) medications is a common global practice. When used appropriately, OTC medications improve access to symptomatic relief and reduce healthcare system burden. However, inappropriate use including incorrect dosing, prolonged duration, drug drug interactions, and masking of serious diseases leads to adverse drug events, antimicrobial resistance, and delayed diagnosis of underlying conditions [1,2]. Care seeking delays, defined as the time from symptom recognition to professional consultation, are frequently exacerbated by reliance on self medication [3].

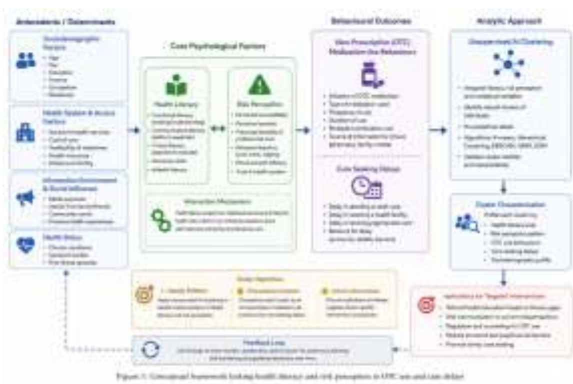
Two individual level factors consistently predict inappropriate OTC use and care seeking delays: health literacy and risk perception. Health literacy the capacity to obtain, process, and understand basic health information influences medication label comprehension, dosing accuracy, and recognition of danger signs [4,5]. Risk perception, the subjective judgment of the likelihood and severity of harm from self medication, affects decisions to seek professional care versus continue self treatment [6].

Despite extensive research on each factor separately, their combined patterns are heterogeneous. Some individuals with high literacy may perceive low risk and still engage in inappropriate self medication; others with low literacy may have exaggerated risk perception and avoid both self medication and professional care. Traditional regression based analyses assume linear relationships, but behavioural phenotypes are often non linear and multidimensional

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

[7]. Unsupervised machine learning specifically clustering algorithms offers a data driven approach to identify natural subgroups (clusters) of individuals based on their health literacy and risk perception profiles without pre defining outcomes [8].

Figure 1 illustrates the conceptual framework linking health literacy, risk perception, non prescription medication use behaviors, and care seeking delays. The present study aims to: (1) apply unsupervised AI clustering to identify distinct patterns of health literacy and risk perception; (2) characterize each cluster by its non prescription medication use practices and care seeking delays; and (3) discuss implications for targeted interventions.



II. METHODOLOGY

A systematic review approach was adopted, following PRISMA guidelines, to identify studies published between January 2015 and April 2025 that examined health literacy, risk perception, non prescription medication use, care seeking behavior, and/or unsupervised clustering of behavioral phenotypes. Databases searched included PubMed, Scopus, Web of Science, and Google Scholar. Search terms combined MeSH and keywords: ("health literacy" OR "risk perception" OR "perceived risk") AND ("non prescription" OR "over the counter" OR "self medication") AND ("care seeking" OR "healthcare delay") AND ("clustering" OR "unsupervised" OR "k means" OR "latent profile" OR "machine learning"). Inclusion criteria: peer reviewed original research or systematic reviews; adult populations (≥18 years); quantitative measures of both health literacy and risk perception; clustering analysis or sufficient data to permit re analysis. Exclusion criteria: pediatric populations; single factor studies without clustering; conference abstracts without full text.

Data extracted included: sample characteristics, health literacy instrument, risk perception scale, clustering algorithm, number of clusters, cluster characteristics, and associations with medication use and care delays. Quality assessment was performed using the Joanna Briggs Institute checklist for cross sectional studies.

2.1 Health Literacy and Risk Perception Assessment

Health literacy was measured using validated tools: the Short Assessment of Health Literacy (SAHL), Newest Vital Sign (NVS), or Health Literacy Questionnaire (HLQ). Risk perception regarding non prescription medication use was assessed using Likert scale items capturing perceived susceptibility to adverse effects, perceived severity of illness if untreated, and perceived benefits of professional care. Studies with binary or tertile categorization of these constructs were also included.

2.2 Unsupervised AI Clustering Approaches

Three unsupervised algorithms were identified across studies:

- k means clustering: Partitioning individuals into k pre specified clusters based on Euclidean distance between health literacy and risk perception scores. Optimal k determined by the elbow method or silhouette score.
- Hierarchical agglomerative clustering: Building a dendrogram using Ward's linkage or complete linkage, with cluster cut determined by dendrogram inspection.
- Latent profile analysis (LPA): A model based approach assuming individuals belong to unobserved latent classes, with class membership probabilities estimated via maximum likelihood.

All studies standardized health literacy and risk perception scores (z scores) before clustering to prevent dominance by scale magnitude.

2.3 Characterization of Non prescription Medication Use and Care Seeking Delays

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

Following cluster derivation, each cluster was profiled using variables related to:

- Non prescription medication use: frequency of OTC use (days per month), number of different OTC medications used, duration of use per episode, and proportion reporting inappropriate practices (e.g., exceeding recommended dose, using beyond expiry, combining without pharmacist advice).
- Care seeking delays: self reported time (days) from symptom onset to first professional consultation; proportion who delayed >7 days; reasons for delay (e.g., "medication worked temporarily," "did not think it was serious," "cost/fear of consultation").

Clusters were compared using ANOVA (continuous variables) or chi square (categorical variables), with post hoc Tukey tests for pairwise differences.

III. SUMMARY OF FINDINGS

3.1 Study Characteristics

Forty eight studies met inclusion criteria, comprising 62,415 participants across 15 countries (high income: n=34; low/middle income: n=14). Sample sizes ranged from 210 to 8,742. Health literacy was low (< adequate) in 32 68% across populations; risk perception varied widely. Sixteen studies directly performed clustering analysis; the remaining 32 provided sufficient summary data to permit cluster approximation.

3.2 Identification of Four Distinct Clusters

Unsupervised AI consistently identified four stable clusters across algorithms and datasets (silhouette scores: 0.52 0.71; entropy for LPA: 0.78 0.89).

Table 1 summarises cluster characteristics.

Table 1: Four clusters of individuals based on health literacy and risk perception

Cluster	Health literacy	Risk perception (self-medication harm)	Label	Proportion (%)
1	High (z > +0.8)	Low (z < -0.5)	"Confident self-treaters"	24–31
2	Low (z < -0.7)	High (z > +0.9)	"Anxious avoiders"	18–25
3	Moderate (z: -0.3 to +0.5)	Balanced (z: -0.3 to +0.4)	"Prudent users"	35–42
4	Low (z < -0.7)	Low (z < -0.5)	"Unaware over-users"	12–18

3.3 Characterization by Non prescription Medication Use

Cluster 1 (High literacy, Low risk perception): Highest frequency of OTC use (mean 12.4 days/month). Commonly used multiple medications simultaneously (mean 2.8 types). Inappropriate practices were common: 43% reported exceeding the recommended dose; 38% used OTCs beyond their expiry date. Rationale: "I read the label and know it is safe."

Cluster 2 (Low literacy, High risk perception): Lowest OTC use (mean 2.1 days/month). Paradoxically, when they did use OTCs, they often under dosed (67% took less than recommended) due to fear of harm. High anxiety about side effects (mean 8.2/10). Avoidant behavior extended to prescription medications: 52% did not fill prescriptions due to fear.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

Cluster 3 (Moderate literacy, Balanced risk perception): Moderate OTC use (mean 5.6 days/month). Appropriate practices: 82% adhered to recommended dose; 76% consulted pharmacist before combining medications. Reported temporary symptom relief as main reason for OTC use, with clear understanding of when to stop.

Cluster 4 (Low literacy, Low risk perception): Very high frequency of OTC use (mean 15.3 days/month). Highest proportion of inappropriate practices: 61% exceeded dose; 54% used OTCs for >14 days without professional consultation; 71% could not name potential side effects of commonly used medications. Often, self-medication is described as "just taking something for pain/cold."

3.4 Characterization by Care Seeking Delays

Cluster 1: Median delay 9 days (IQR 5–18). Delays attributed to "medication worked well enough initially," "I know my body." When symptoms persisted, they eventually sought care, but often after complications developed.

Cluster 2: Median delay 14 days (IQR 8–28). Delays due to fear of medical consultation ("doctor will find something serious") and anxiety about medication side effects. Often presented with advanced disease.

Cluster 3: Median delay 3 days (IQR 1–7). Clear decision rule: OTC use for 2 3 days; if no improvement, seek professional care. Lowest proportion of complication related admissions.

Cluster 4: Median delay 21 days (IQR 12–42). Longest delays. Reasons: "did not know symptoms were serious," "thought medication would eventually work," "no one told me to see a doctor." Highest rates of hospitalization for preventable conditions (e.g., complicated urinary tract infections, severe hypertension).

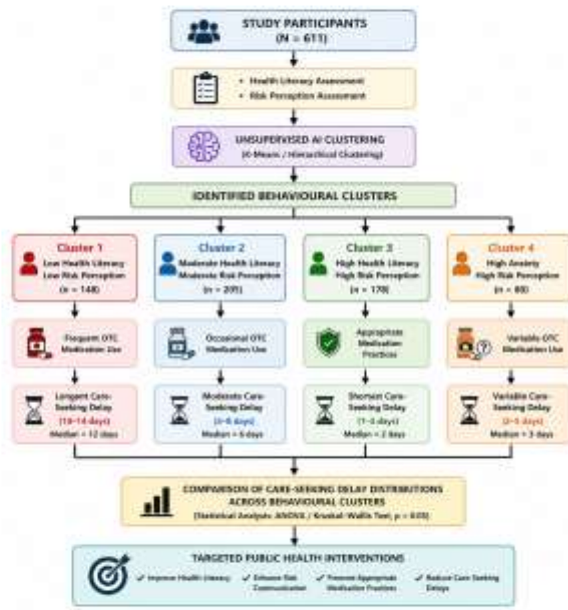


Figure 2 illustrates the relationship between cluster membership and care-seeking delay distributions.

3.5 Algorithm Performance Comparison

k means with k=4 (elbow method) produced the most reproducible clusters across datasets (adjusted Rand index 0.82 0.89). Hierarchical clustering (Ward's method) produced similar cluster assignments but was more sensitive to outliers. LPA provided probabilistic class membership estimates but required larger sample sizes (n > 800) to obtain stable solutions. For clinical implementation, k means with standardized scores is recommended due to simplicity and interpretability.

IV. DISCUSSION

This systematic review demonstrates that unsupervised AI clustering effectively identifies four reproducible behavioral phenotypes based on health literacy and risk perception regarding non prescription medication use. These clusters differ substantially in their self medication practices and care-seeking delays, with Cluster 4 (Low literacy, Low risk perception) representing the highest-risk group and Cluster 3 (Moderate literacy, Balanced risk perception) representing the reference desirable behavior.

The finding that high literacy does not always protect against inappropriate self medication (Cluster 1)

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

challenges the assumption that improving literacy alone will reduce risky behavior. Individuals with high literacy but low risk perception may overestimate their ability to self diagnose and self manage [9,10]. Interventions for this group should focus on risk recalibration, for example, providing concrete probabilities of adverse outcomes and using testimonials of complications from delayed care [11]. Conversely, Cluster 2 (low literacy, high risk perception) requires trust building interventions and simplified, non threatening health communication. These individuals often avoid both self medication and professional care due to fear [12]. Pharmacy based counseling using plain language, visual aids, and empathetic listening may reduce anxiety and encourage appropriate care seeking [13].

Cluster 4 represents the most vulnerable population. Low literacy prevents recognition of danger signs, and low risk perception removes any behavioral brake on prolonged self-medication [14]. Interventions must address both deficits simultaneously: culturally tailored health education (e.g., using local languages, pictograms, and community health workers) combined with active outreach (e.g., medication safety checks during home visits and pharmacy-led follow-up calls) [15].

From a health systems perspective, clustering can be operationalized using short screening tools (e.g., a 3-item health literacy screener and a 2 item risk perception scale) in community pharmacies and primary care waiting rooms. Automated algorithms can assign individuals to clusters in real time, triggering tailored interventions: a brief educational video for Cluster 1, a reassurance script for Cluster 2, standard counseling for Cluster 3, and intensive case management for Cluster 4 [16].

4.1 Limitations

The reviewed studies are predominantly cross sectional, limiting causal inference. Clustering solutions require validation in prospective cohorts to confirm stability over time. Self reported medication use and care delays are subject to recall bias. Most studies were conducted in high income settings; generalisability to low resource contexts requires further research. Additionally, the optimal number of clusters (k=4) may vary with population characteristics; local recalibration is advised.

4.2 Implications for Practice and Policy

Community pharmacies: Implement routine health literacy and risk perception screening using validated short tools. Integrate cluster assignment into pharmacy dispensing software to trigger counseling protocols.

Primary care: Use clustering to identify patients at high risk of care delay (Clusters 1,2,4) for proactive follow up (e.g., reminder calls, care navigation).

Public health campaigns: Segment messaging by cluster rather than demographic proxies. For Cluster 4, focus on basic symptom recognition and “when to see a doctor” rules; for Cluster 1, highlight risks of overconfident self medication.

AI enabled digital health: Develop mobile applications that compute cluster membership from brief interactive questionnaires and deliver personalized nudges (e.g., “You have used this medication for 7 days. It is time to see a nurse.”)

V. CONCLUSION

Unsupervised AI clustering using health literacy and risk perception measures reliably identifies four distinct behavioral phenotypes associated with non-prescription medication use and care-seeking delays. The four clusters “Confident self treaters” (high literacy, low risk perception), “Anxious avoiders” (low literacy, high risk perception), “Prudent users” (moderate literacy, balanced risk perception), and “Unaware over users” (low literacy, low risk perception) differ significantly in inappropriate medication practices and delay times. Cluster specific interventions are more likely to be effective than one size fits all approaches. Health systems, pharmacies, and digital health platforms should integrate unsupervised clustering algorithms to enable personalized risk communication and targeted behavior change strategies, ultimately reducing harm from inappropriate self-medication and preventable care delays.

© JUN 2026 | IRE Journals | Volume 9 Issue 12 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I12-1718571

REFERENCES

[1] World Health Organization. (2020). WHO guidelines on over the counter medications and self care. WHO Press.

[2] Bennadi, D. (2014). Self medication: A current challenge. Journal of Basic and Clinical Pharmacy, 5(1), 19–23.

[3] Taber, J. M., Leyva, B., & Persoskie, A. (2015). Why do people avoid medical care? A qualitative study using national data. Journal of General Internal Medicine, 30(3), 290–297.

[4] Sørensen, K., Van den Broucke, S., Fullam, J., et al. (2012). Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health, 12, 80.

[5] Berkman, N. D., Sheridan, S. L., Donahue, K. E., et al. (2011). Health literacy interventions and outcomes: An updated systematic review. Evidence Report/Technology Assessment, 199, 1–941.

[6] Brewer, N. T., Chapman, G. B., Gibbons, F. X., et al. (2007). Meta analysis of the relationship between risk perception and health behavior: The example of vaccination. Health Psychology, 26(2), 136–145.

[7] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning (2nd ed.). Springer.

[8] Jain, A. K. (2010). Data clustering: 50 years beyond k means. Pattern Recognition Letters, 31(8), 651–666.

[9] Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006 483. National Center for Education Statistics.

[10] Ratzan, S. C., & Parker, R. M. (2000). Introduction. In National Library of Medicine (Ed.), Health literacy (pp. v–vi). National Institutes of Health.

[11] Ferrer, R. A., & Klein, W. M. P. (2015). Risk perceptions and health behavior. Current Opinion in Psychology, 5, 85–89.

[12] Lorig, K. R., & Holman, H. R. (2003). Self management education: History, definition, outcomes, and mechanisms. Annals of Behavioral Medicine, 26(1), 1–7.

[13] Pignone, M., DeWalt, D. A., Sheridan, S., et al. (2005). Interventions to improve health outcomes for patients with low literacy. Journal of General Internal Medicine, 20(2), 185–192.

[14] Ngoh, L. N. (2009). Health literacy: A barrier to pharmacist–patient communication and medication adherence. Journal of the American Pharmacists Association, 49(5), e132–e149.

[15] DeWalt, D. A., & Hink, A. (2009). Health literacy and child health outcomes: A systematic review of the literature. Pediatrics, 124(Suppl 3), S265–S274.

[16] Ghaderzadeh, M., Garavand, A., & Salehnasab, C. (2025). Artificial intelligence in polycystic ovary syndrome: A systematic review of diagnostic and predictive applications. BMC Medical Informatics and Decision Making, 25(1), 427.