

Volume 2, Issue 2, (2026)

1st June, 2026

DOI.org/10.71079/ASIDE.HS.2.2

ASIDE
HEALTHCARE

**HEALTH
SCIENCES**

**PUBLISHED BY
PUBPORTA**

**ISSN (Print) 3067-8730
ISSN (Online) 3067-8749**



Sacha Krolow e Silva, Larissa Britto de Deus, Leonardo da Silva Bittencourt

1-8

[Effectiveness of Alcohol Gel-Based Antiseptics Against Multidrug-Resistant Bacteria: An in vitro Experimental Study](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.121925210>

Maria Rizzo, Omar Irfan

9-17

[Search Filters to Identify Automation in HEOR: An Umbrella Review of Performance and Overlap](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.122125202>

Wael Alsaqaf, Khaled Alkohlany, Majdi Alshami , Mahmood Shaif , Abdullah Al-Sakkaf, Ahmed Farei, Aziz Shamsan, Nasser Albaddai

18-23

[War Injuries of External Male Genitalia: A Prospective Single-Center Study from the General Military Hospital, Sana'a, Yemen \(2022–2024\)](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.012126461>

Anas Al-Qatefy, Abdelrahman Tolba, Zeyad Eldawy, Mahmoud Ghallab

24-32

[Cross-Cultural Application of the Beck Depression Inventory \(BDI-I/II\) for Assessing Adolescent Depression in Low- and Middle-Income Countries: A Critical Review of Evidence from the Last Quarter-Century](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.012226419>

Uzzal Chondra, Palash C. Roy , Khaled H. Riad , Masud A. Sumon , Md. Zahirul Islam, S.M. Nazmul Islam , Md. Jahid Hasan , Pabitra K. Shill, Mithun Sikder, Shawon S. Joy, Annesha Majumder

33-39

[Selective Attenuation of Acetaminophen-induced Hepatobiliary Dysfunction by Pseudoelephantopus spicatus in Sprague-Dawley Rats](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.040826549>

Elaf Falih Abbas, Mohammed Mousa Imran, Mustafa Yahya Saleh, Huda Omran

40-46

[Ocular Trauma Patterns in the Middle East and North Africa: A Scoping Review with Emphasis on Iraq](#)

DOI: <https://doi.org/10.71079/ASIDE.HS.042826665>



Original Article

Effectiveness of Alcohol Gel-Based Antiseptics Against Multidrug-Resistant Bacteria: An *in vitro* Experimental StudySacha Krolow e Silva¹, Larissa Britto de Deus², Leonardo da Silva Bittencourt^{3, 4, 5}

1-Graduate Program in Biological Sciences: Biochemistry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

2-Cesuca University Center, Porto Alegre, Brazil

3-Department of Morphological and Physiological Sciences, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

4-School of Public Health of Rio Grande do Sul (ESP/RS), Porto Alegre, Brazil

5-State School of Professional Health Education, Hospital de Clínicas de Porto Alegre (EPS/HCPA), Porto Alegre, Brazil

ARTICLE INFO

Article history:

Received 8 Aug. 2025

Received in revised form 5 Oct. 2025

Accepted 10 Oct. 2025

Published 19 Dec. 2025

Keywords:

Antiseptics

Hand hygiene preparations

Hand hygiene

Standard bacterial strains

Healthcare-related infections

ABSTRACT

Background: Since the 19th century, sepsis has revolutionized public health, and hand hygiene remains a cornerstone of infection control, particularly in healthcare settings with high exposure to resistant microorganisms. This study aimed to evaluate the antimicrobial efficacy of four commercial alcohol-gel brands (A, B, C, and D) against standard ATCC strains of *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*, which are prevalent in clinical environments.

Methods: The disk-diffusion method was employed on Mueller-Hinton agar, using sterile disks impregnated with each gel sample. Following a 24-hour incubation at 35°C, the resulting halo zones were measured.

Results: Significant variability was observed among the brands. Brand C exhibited the highest antimicrobial activity, particularly against *P. aeruginosa* (11.10 mm), with an overall mean inhibitory activity of 9.8 mm. Brand B demonstrated broad efficacy (mean 9.2 mm), while Brand A showed moderate activity with higher variability (mean 7.9 mm). Brand D demonstrated minimal to no inhibitory activity in the disk-diffusion assay compared to the other tested products (mean 4.3 mm). These differences were statistically significant across all tested bacteria.

Conclusions: These findings indicate that alcohol concentration is not the sole determinant of antimicrobial performance. Formulation factors—such as humectants, polymers, and viscosity—substantially impact efficacy. The composition of an alcohol-based hand sanitizer directly governs its antimicrobial power. The study reinforces the importance of selecting validated products, adhering to hygiene protocols, and conducting microbiological surveillance to prevent healthcare-associated infections.

1. Introduction

Asepsis, the absence of infectious agents in environments through preventive actions that inhibit the entry and spread of pathogens, began to change the reality of healthcare in the 19th century: in 1846, Ignatius Philip Semmelweis reduced mortality from puerperal sepsis through hand washing; in 1865, Joseph Lister introduced phenol as an antiseptic to reduce post-operative infections; and during the Crimean War (1854-1855), Florence Nightingale promoted hygiene reforms and nurse training to reduce hospital infections [1–3]. Today, hand hygiene remains a fundamental measure to prevent the transmission of diseases, especially in healthcare environments with high exposure to resistant microorganisms. Human skin harbors both a resident microbiota, which is difficult to

remove and generally harmless, and a transient microbiota, which is easily eliminated with proper hygiene [4]. The use of soap and water reduces the microbial load and stops transmission, while alcohol-based antiseptics are often used in the healthcare sector due to their bactericidal and bacteriostatic properties [5, 6].

Sanitizers, especially 70% alcohol gel, which were already being studied for infection control, became extremely popular during the coronavirus pandemic (SARS-CoV-2), where the demand for hand sanitizers grew dramatically, by more than 60% in the first year and more than 120% in the second year [7, 8]. Even so, studies carried out in hospitals during the pandemic have shown that pedal-operated hand sanitizer dispensers, combined with behavioral nudges, in hand hygiene practices had low visitor adherence, especially after critical contacts, such as exposure to respiratory fluids as well as many health professionals having insufficient knowledge about asepsis before invasive procedures [9, 10].

Aqueous solutions of ethanol and n-propanol, at concentrations comparable to those found in commercial disinfectants (60-90% alcohol) [11], destabilize viral membranes, such as SARS-CoV-2, and can cause structural disintegration and inactivation of the respiratory syncytial virus in suspension tests, reaching reductions of more than 3log10 in 30 seconds of contact, regardless of the

*Corresponding author: Sacha Krolow e Silva, Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil. Email: sachasilva@hcpa.edu.br

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Silva SK e, de Deus LB, Bittencourt L da S. Effectiveness of Alcohol Gel-Based Antiseptics Against Multidrug-Resistant Bacteria: An *in vitro* Experimental Study. ASIDE Health Sci. 2025;2(2):1-8, doi:10.71079/ASIDE.HS.121925210

format, whether gel or foam [12]. These results reinforce the use of alcohol-based sanitizers as a preventative measure for virus transmission, especially in the absence of soap and water.

Staphylococcus aureus (ATCC 25923) is a coagulase-positive, hemolytic bacterium from the Micrococcaceae family, which produces various cytolytic toxins (α , β , γ , δ , and leukocidin), enterotoxins (A to E) associated with foodborne infections, as well as enzymes related to its virulence. It is an opportunistic pathogen in community and hospital environments, often resistant to penicillin and, in some strains, to methicillin (oxacillin), as is the case with methicillin-resistant *Staphylococcus aureus* (MRSA), whose mortality rate in Brazil can vary between 40% and 80%, especially in intensive care units (ICUs) [13]. *Enterococcus faecalis* (ATCC 29212), from the Streptococcaceae family, is found in the human intestine and female genital tract, and is responsible for urinary tract infections, intra-abdominal abscesses, endocarditis, and meningitis. It represents between 85% and 90% of *Enterococcus* species. Although generally less resistant, strains with resistance to vancomycin have been detected in Brazil since 1996 due to alterations in the cell wall precursors that prevent the antibiotic from binding [13].

Escherichia coli (ATCC 25922), belonging to the Enterobacteriaceae family, is a common fermenting bacterium in the human intestine and is generally non-pathogenic. However, it is often associated with septicemia, meningitis, and enteritis, which gives it clinical relevance (Rey, 2003). Since 2011, there has been an increase in cases of *E. coli* infection in Europe. Between 2007 and 2016, it was responsible, along with *Salmonella* spp. and *S. aureus*, for around 90.5% of cases of food-borne illness, with new records as recently as 2024 [14–16]. *Pseudomonas aeruginosa* (ATCC 27853), of the Pseudomonadaceae family, is often found in immunocompromised or hospitalized patients, due to its low nutritional requirements and ability to colonize hospital environments. Although it is not the most common microorganism in these places, its clinical importance lies in its high resistance to various antimicrobials [17, 18].

Considering the findings and the importance of access to sanitized products for the general public, this study sought to test the efficacy of antiseptics purchased in the regular trade and used by health professionals in hospitals and laboratories, as well as by ordinary people, against strains of *S. aureus* (ATCC 25923), *E. coli* (ATCC 25922), *P. aeruginosa* (ATCC 27853), and *E. faecalis* (ATCC 29212), which are considered common in hospital and laboratory environments.

2. Methods

2.1. Antiseptics

The antiseptics were purchased from a pharmacy and are referred to as brands A, B, C, and D. Their chemical compositions are described below as presented on the label.

- Brand A: ethyl alcohol 70%, water, bidistilled glycerin, acrylic polymer - TR1, propylene glycol USP, triethanolamine;
- Brand B: water, alcohol 70%, C10-C30 alkyl propenoate, methylparaben, propylene glycol, disodium EDTA, aminomethyl propanol, perfume, amylcinnamic, linalool.
- Brand C: ethyl alcohol 70%, carbopol, neutralizing base, glycerin, purified water.
- Brand D: alcohol 70%, water, carbomer, aminomethyl propanol, propylene glycol.

2.2. Antimicrobial activity

In the experiments, 48 plates containing NEWPROV 140x15mm Müeller-Hinton Agar, described for testing antimicrobial susceptibility by the disk-diffusion method, were used to seed the bacteria. The different alcohol brands were placed on the plates using sterile 6.30mm antibiotic-free disks from Cefar Diagnóstica Ltda. Laboratório de Pesquisa e Desenvolvimento de Produtos para Diagnósticos.

Standard strains identified for quality control, as those from the American Type Culture Collection (ATCC), were used in this study. The reference strains were obtained from a reliable source. The bacterial strains used in this work were: *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*. The inoculum was prepared using the Direct Colony Suspension Method in a 0.9% saline solution from pre-selected isolated colonies on a 24-hour agar plate. The suspension was adjusted so that its turbidity was on the McFarland scale of 0.5 (1×10^8 CFU/mL) and checked on a densitometer.

The bacterial suspension was seeded according to the Kirby & Bauer methodology described in the Clinical and Laboratory Standards Institute (CLSI 2013, as described in the Antibioqram Manual prepared by Laborclin's Technical Sector team in 2013 [19, 20]:

1. Soak a sterile cotton swab in the bacterial suspension, pressing it against the walls of the tube to remove excess suspension;
2. Sow gently in all directions on the plate (four directions), rotating the plate approximately 90° and repeating the procedure to ensure even distribution of the inoculum.
3. Swab the edge of the agar plate and put the lid on to prevent moisture from being absorbed before applying the alcohol-gel-impregnated disks.

The sterile disks were placed on a sterilized watch glass, where 1ml of each brand of alcohol-gel was pipetted above the disk and allowed to absorb [21]. After the disks absorbed the alcohol-gel by capillary action, ensuring complete wetting without oversaturation, they were carefully positioned in the appropriately labelled sections of the inoculated agar plates. The lids were promptly replaced to reduce the risk of contamination and prevent evaporation. The same procedure was carried out for the four brands for each of the bacteria evaluated.

Each Müeller-Hinton agar plate measured 140×15 mm (2100 mm^2) and was used to test a single bacterial strain. On each plate, four sterile disks of 6.30 mm diameter containing the alcohol-gel samples (one per brand) were placed at the vertices of a square, approximately 2.0 cm from the plate edge, with a minimum spacing of 24 mm between disks. A sterile disk with water was positioned at the center as a negative control. Each disk was surrounded by roughly 389 mm^2 of agar free of antibacterial agents, allowing uniform diffusion and minimizing edge effects. Since it is conducted in a laboratory environment, we use routine hospital antibiograms as a positive control to confirm the effectiveness of ATCC strains.

After the procedure, the plates were identified with the name of the bacteria and letters representing the brands of each antiseptic, then they were incubated at 35°C for 24 hours. After the incubation period, the halos were read as a measure of antibacterial activity using a 15.0 cm graduated ruler. Measurements were performed independently and separately by two authors [SKS and LBD], and any discrepancies were discussed and resolved to ensure consistency with a third author [LSB]. Positive readings were those

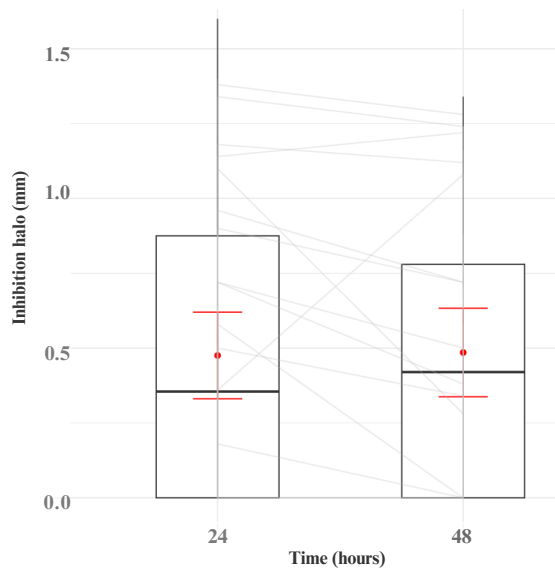


Figure 1: Temporal analysis of response by time (24h and 48h).

Boxplots by time with mean $\pm$ 95% CI and paired lines.

in which a halo of inhibited bacterial growth could be observed around the disks submerged in the antiseptic.

In this study, sterile water applied to a disk at the assay center was used as the negative control, ensuring that any observed halo of inhibition zones could be attributed solely to the antiseptic gels rather than to procedural artifacts. As a positive control, routine hospital antibiograms were used to verify the expected susceptibility profile of ATCC strains. They served as a biological quality control to confirm the reliability of the experimental system, although they did not represent a conventional antiseptic reference. Nonetheless, for the scope of the present comparative analysis, the sterile water disk was considered adequate to establish the antimicrobial activity of the tested products. The experiment sought to demonstrate the initial effect/immediate activity of alcohol-gel, being performed in triplicate at the critical time (24h). And for exploratory purposes, it was observed after 48 hours to see if the effect of the described activity of alcohol-gel remained. Results are expressed as mean values $\pm$ standard errors, reflecting the average antimicrobial activity observed across the replicates for each treatment and bacterial strain.

2.3. Statistical analysis

Statistical analyses were performed in R. For repeated measurements on the same disks/plates, linear mixed-effects models will be used with time as a fixed effect and plate/disk as a random effect. For independent measurements, ANOVA or non-parametric tests (Kruskal-Wallis/Mann-Whitney) will be applied. Results are expressed as mean $\pm$ standard error (SE) or 95% CI (when $n > 1$), with p-values adjusted for multiple comparisons (Bonferroni or Holm) for three independent experiments. Statistical significance was set at $p < 0.05$.

3. Results

After incubation for 24 hours, the diameters of halos were measured and then interpreted in relation to the positive and negative controls. The experiments were repeated three times, and each seeded plate

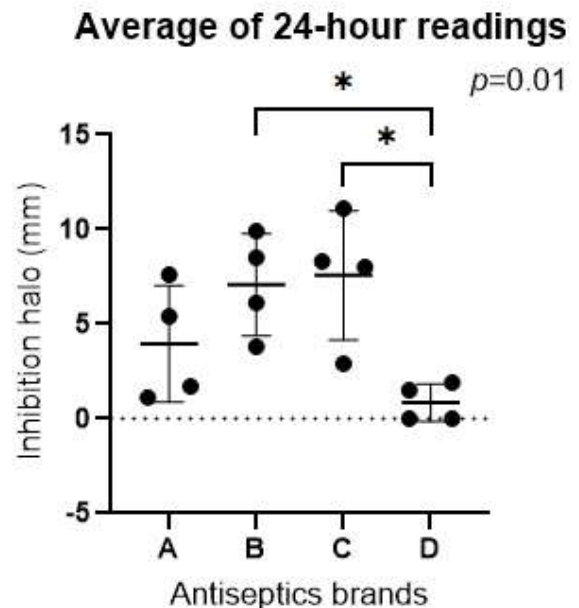


Figure 2: Antimicrobial efficacy of antiseptic brands based on average inhibition halo diameters after 24 hours of exposure.

Average inhibition halos (mm) after 24 hours for each antiseptic gel brand, considering the combined mean values against the four bacterial strains (*Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*). Results are expressed as mean $\pm$ standard error of the mean. Data were analyzed using one-way ANOVA ($F = 5.308$; $p = 0.0147$), followed by Tukey's multiple comparisons test. Statistically significant differences were found between brands B and D ($p = 0.0340$), and between brands C and D ($p = 0.0255$). Asterisks indicate statistically significant differences ($p < 0.05$).

was arranged with four disks representing the four alcohol-gel samples. Therefore, the values obtained from the halos were divided by the total number of disks in the agar, excluding the sterile disk, which served as a control. In this study, three independent experiments were conducted to verify whether the results obtained in experiment 1 would be consistent and thus reduce variability. The results are detailed in (Supplementary Table 1).

The effective sample size was 46 observations for 24 h and 41 observations for 48 h. Atypical growth of resistant colonies was observed in the area of the discs, which were marked as R* and excluded from the analysis to avoid bias. Sixteen paired disks (same plate ID) were identified between the two times, allowing points to be connected in the graphs and the random effect to be included in the mixed model (Figure 1).

After 24 hours, the average halos (mm) were calculated for each antiseptic gel brand, considering the combined mean values against *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*. The overall descriptive statistics showed very close means between times: 24h = 0.466 (95% CI 0.276–0.657) and 48h = 0.448 (95% CI 0.253–0.644). The normality test indicated a significant deviation from normality in the overall 24h set ($W = 0.866$, $p < 0.001$). However, only Brand A showed strong non-normality in the analysis by brand, while the others did not reject normality within the current sample limits. When brands were compared, one-way ANOVA indicated significant differences ($F = 5.308$; $p = 0.0147$), with Tukey's test revealing lower efficacy of Brand D compared with Brands B

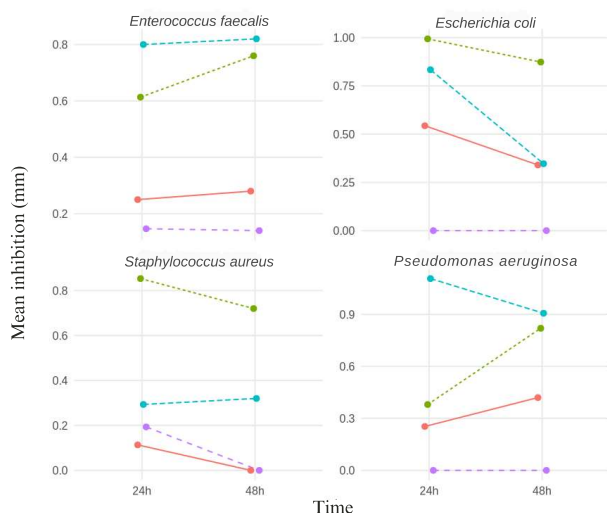


Figure 3: Mean inhibition by time for each brand within each bacterium.

An average inhibition was plotted over time for each brand in both periods. Each facet represents a bacterium; the lines/dots show brands A, B, C, and D of alcohol gel at 24 to 48 hours. Brand A in red, Brand B in green, Brand C in blue and Brand D in purple.

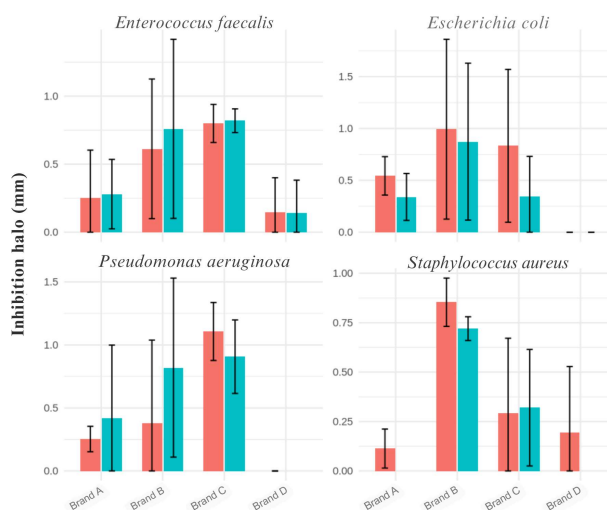


Figure 4: Mean inhibition zones by brand and incubation time against tested bacteria.

Bar plot of mean inhibition by brand and time with SD error bars, faceted by bacterium: *Enterococcus faecalis*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*. Results are shown by brand (A–D) and incubation time (24 h in red, 48 h in blue).

($p = 0.034$) and C ($p = 0.026$) (**Figure 2**). In contrast, species-specific analyses showed no significant differences among brands after adjustment: *E. faecalis* ($p = 0.127$), *E. coli* ($p = 0.240$), *S. aureus* ($p = 0.098$), and *P. aeruginosa* ($p = 0.122$) (**Supplementary Table 1**).

For *Enterococcus faecalis*, the Kruskal–Wallis test was applied, yielding a p -value of 0.127, an effect size (ϵ^2) of 0.269, and Levene's test indicated homogeneity of variances ($p = 0.630$). For *Escherichia coli*, one-way ANOVA was used, resulting in a p -value of 0.240, an effect size (η^2) of 0.392, and a non-significant Levene's test ($p = 0.477$). For *Staphylococcus aureus*, the Kruskal–Wallis

test showed a p -value of 0.098, with an effect size (ϵ^2) of 0.300, and Levene's p -value of 0.734. Finally, for *Pseudomonas aeruginosa*, the Kruskal–Wallis test produced a p -value of 0.122, an effect size (ϵ^2) of 0.279, and Levene's test was also non-significant ($p = 0.652$) (**Supplementary Table 1**).

Head of the pairwise post-hoc (**Supplementary Table 2**) results indicated that, for *Enterococcus faecalis* (Kruskal–Wallis, Dunn's test with BH adjustment), none of the pairwise comparisons reached significance after correction, with the closest contrast being C–D ($p_{\text{adj}} = 0.152$). For *Escherichia coli* (ANOVA, Tukey HSD), all pairwise comparisons were non-significant, with adjusted p -values ranging from 0.229 to 0.986. In *Staphylococcus aureus* (Kruskal–Wallis, Dunn's test with BH adjustment), the contrasts A–B ($p_{\text{adj}} = 0.129$) and B–D ($p_{\text{adj}} = 0.145$) approached significance but did not remain significant after adjustment. Finally, for *Pseudomonas aeruginosa* (Kruskal–Wallis, Dunn's test with BH adjustment), the contrast C–D yielded the lowest adjusted p -value ($p_{\text{adj}} = 0.125$), though still above the conventional significance threshold.

The REML-adjusted mixed linear model, considering time as a fixed effect and plate ID as a random effect, indicated no significant effect of time (Intercept = 0.466 ± 0.092 , $p < 0.001$; 48h effect = -0.018 ± 0.084 , $p = 0.832$). The marginal estimates were 24h = 0.466 ± 0.092 and 48h = 0.448 ± 0.095 , with a Holm-adjusted difference of 0.018 ($p = 0.833$). An ANOVA of independent measures also confirmed no significant difference ($F(1,85) = 0.009$, $p = 0.923$).

The analysis of deltas (halos 48h – halos 24h) by bacterium and brand showed small variations and no consistent trend: *Enterococcus faecalis* ranged from -0.007 to 0.147 ; *Escherichia coli* showed reductions from -0.487 to -0.120 ; *Pseudomonas aeruginosa* ranged from -0.200 to 0.440 ; and *Staphylococcus aureus* from -0.193 to 0.027 . These results suggest variability between brands and individual replicates, but do not indicate a systematic effect of time (**Figure 3**). The ranking of effectiveness by bacteria remained relatively consistent: Brand C was most effective against *E. faecalis* and *P. aeruginosa*; Brand B was most effective against *E. coli* and *S. aureus*; while Brand D consistently demonstrated minimal to no inhibitory activity in the disk-diffusion assay compared to the other tested brands (**Figure 4**).

4. Discussion

The ASTM E1174 (American Society for Testing and Materials) standard is an American method for evaluating antiseptic hand wash formulations for healthcare professionals. At the same time, EN 1500 (EuroNorm) establishes criteria for rub-in hygienic disinfectants, using the reduction of transient microbial flora under controlled experimental conditions as a reference. Both use gram-negative bacteria as marker organisms, yet they still do not align with the practical everyday use of alcoholic hand preparations [22]. Despite the high concentrations, conducting this experiment with clinically relevant pathogenic strains enabled a direct comparison of the effectiveness of alcohol gel with that of antimicrobials used clinically. Thus, the methodological choice sought not only to approximate the results of situations encountered in healthcare settings, where the easy spread of microorganisms among patients and healthcare professionals represents a constant risk of cross-contamination, but also to provide an objective comparative parameter between the activity of routinely used antiseptics and the response to pathogens of clinical interest. The evaluation of antimicrobial activity using alcohol gel has already been described

for its simplicity and cost-effectiveness. However, there have been reports of variations in halo reading due to how the drop spreads across the culture medium.

The antiseptic's action is naturally limited as its concentration decreases with distance from the disc, making chemical neutralization unnecessary. Furthermore, since the active ingredient is a volatile alcohol, its concentration significantly decreases during the 24-hour incubation period due to evaporation, further limiting the potential for artificially enlarged halos. The main objective of our study was to compare the intrinsic antimicrobial activity of the product formulations under identical conditions. Since all products were subjected to the same physical principles of diffusion and evaporation, the relative comparison of the resulting inhibition zones remains valid and straightforward for measuring their in vitro efficacy.

The detailed experimental design further supports the robustness of the methodology. Halos were measured after 24 hours, and the experiments were repeated three times with five disk replicates per gel, ensuring sufficient sampling and minimizing random variability. Atypical growth of resistant colonies occasionally observed above the discs (marked as R*) was carefully excluded from the statistical analyses to avoid bias, resulting in an effective sample size of 46 observations at 24 h and 41 at 48 h. Although infrequent, these occurrences highlight the potential emergence of resistant subpopulations under certain conditions and should be explored in future studies. Importantly, the use of paired disks allowed the incorporation of random effects into mixed models, further strengthening statistical validity. Despite minor differences across bacteria and brands, the consistent halos of activity against these bacteria between 24 and 48 hours, along with stable brand performance rankings, indicate that excluding R* data did not compromise the study's conclusions. Overall, these results reinforce confidence in the findings while ensuring transparency in experimental handling, variability, and data interpretation.

Considering the concentration of bacteria would be higher than that found on the surface of the hand, since the inoculum used was in the range of 1×10^8 CFU/mL, which is associated with a site of infection (sepsis) in the patient [20], the alcohol gel brands proved to be effective against the bacterial strains. For the analysis and discussion of the results, since there is no reference standard in the literature, alcohol gel is not considered an antibiotic but a prophylactic. Halos were only described if they demonstrated minimal activity against the bacteria.

Analyzing the results, possible aspects could be linked to the action of alcohol gel in inhibiting bacterial growth, such as: the presence of ethanol, commonly used in hand sanitizers for its effectiveness in inactivating viruses and bacteria, and propanol, another variant of alcohol that is also effective against pathogens [23]. Alcohol in gel form doesn't spread in the culture medium. Still, it can inhibit bacterial growth because its consistency increases the contact time with the contaminated surface by slowing down evaporation [24]. Cosmetic additives or emollients in alcohol improve skin tolerance and adherence to use by professionals, without interfering with the product's action [23].

Brand C was the most effective antiseptic, showing the largest halo, especially against *Pseudomonas aeruginosa*, likely due to an optimized alcohol concentration and synergistic ingredients. Brand A showed consistent, though lower, antimicrobial activity, with borderline statistical significance suggesting possible variability in formulation. Brand B achieved high effectiveness against *E. coli* and *S. aureus*, indicating that small differences in composition can

influence efficacy. In contrast, Brand D showed poor performance, failing to inhibit key pathogens, which raises concerns about its use and highlights the importance of strict quality control and efficacy testing.

The presence of ethanol and/or propanol is a key factor in the action of antiseptic gels, as both alcohols are capable of denaturing proteins and disrupting microbial membranes, particularly at concentrations above 60% [23, 25]. The gel form prolongs contact time with the skin or surface, potentially increasing antimicrobial activity [24]. Moreover, the inclusion of cosmetic agents or emollients can promote user adherence without compromising efficacy, a crucial factor for routine use by healthcare professionals.

Although all brands claim to contain 70% ethyl alcohol, significant variability in antibacterial efficacy was observed, suggesting that other components of the formulation strongly influence the results. Additives such as glycerin—present in Brands A, B, and C—act as humectants and improve skin tolerability, but several pieces of evidence indicate that they may reduce alcohol diffusion or form physical barriers, decreasing halos in vitro. For example, Suchomel et al. (2017) demonstrated that glycerin in alcohol-based formulations decreases bactericidal activity, particularly with regard to prolonged effects, compared with formulations without glycerin or with alternative humectants. Similar results were also reported by Meneguetti et al. (2019) [26, 27].

Other critical factors include the type and concentration of polymers (carbopol, carbomer, TR1 acrylic polymer), which determine the gel's viscosity and affect the evaporation rate and the diffusion of alcohol through the medium. More viscous formulations can keep alcohol on the surface longer, but they can also hinder its effective release, thereby inhibiting bacterial growth on agar plates [28]. In addition, neutralizers or pH-adjusting agents, such as triethanolamine and aminomethyl propanol, can alter the gel's acidity, influencing alcohol stability and bacterial susceptibility. However, some studies indicate that within moderate ranges, pH does not have a predominant effect [29].

In this sense, the presence of humectants, polymer formulation, viscosity, and pH are variables that modify practical effectiveness. For clinical or personal hygiene applications, it is recommended to evaluate not only the alcohol content but also the complete formulation, including additives, to ensure consistent antimicrobial performance [30].

In recent years, the rise in hospital infections from multidrug-resistant microorganisms, along with the pandemic, has underscored the vulnerability of healthcare workers, who are more susceptible to pathogen contamination through direct contact with infected patients or contaminated surfaces. In epidemiology, it is considered that "the hands of healthcare workers are the main source of outbreaks of healthcare-related infections" and are a critical factor in transmission between infected and healthy patients [31].

Currently, alcoholic formulations are widely used for hand sanitizing because they promote microbial reduction, are easy to apply, and cause less irritation than water, which can remove the skin's hydrolipidic barrier [32–34]. Alcohols denature and coagulate proteins when in concentrations of more than 60% and are more effective at reducing hand microbes than ordinary soaps or antiseptics, even in the presence of small amounts of proteinaceous material [25, 35]. However, hand washing with soap and water remains essential in cases of extreme contamination or visible dirt [36]. In addition, excessive use of alcohol-based sanitizers can

cause skin cracks [37], increasing the risk of infections and toxicity, and potentially selecting resistant strains, such as *Staphylococcus epidermidis*, which can form biofilms [38]. It is therefore essential to select products that guarantee efficacy, safety, and compatibility with users' skin.

The agar-based halo assay provides a controlled, standardized measure of minimal or no inhibitory activity in the disk-diffusion assay compared to the other tested products. Yet, it does not replicate the dynamics of skin colonization, contact time, or the variable microbial load encountered in clinical practice. The use of high inoculum concentrations—above those typically found on healthcare workers' hands—further complicates extrapolation to daily-use scenarios. Future studies should include clinical isolates, employ complementary methodologies, such as in vivo or simulated hand-contamination models, and explicitly address how experimental conditions may limit the generalizability of the findings. This would strengthen the discussion and provide clearer guidance on the applicability of results to infection prevention and control in healthcare environments.

The methodology used in this study was robust and appropriate for evaluating the efficacy of antiseptic gels. To ensure the accuracy and reliability of results, experiments were conducted in triplicate, and cross-contamination was avoided by using different strains on different days. Statistical analyses, including ANOVA, Kruskal–Wallis, and mixed-effects models, allowed rigorous comparison among brands and assessment of incubation time effects.

Nevertheless, the use of standard laboratory strains may not fully reflect the resistance patterns of clinical isolates, limiting direct extrapolation to real-world hospital settings. Despite this, the experimental design and statistical rigor provide a solid foundation for interpreting the relative antimicrobial performance of the gels, while highlighting the need for further studies with clinical strains to confirm applicability in healthcare environments.

Limitations

A primary limitation of this investigation is the use of the disk-diffusion method. We acknowledge that this is not a standard protocol for determining the clinical efficacy of hand rubs, for which methodologies like EN 1500 or ASTM E1174 are the established gold standards. Consequently, our results cannot be directly extrapolated to clinical performance, as we lack established breakpoints from standards organizations (e.g., CLSI) to classify an antiseptic as "susceptible" or "resistant" based on inhibition zone diameters. However, the objective of this study was not to validate clinical efficacy, but rather to conduct an exploratory in vitro screening to compare the antimicrobial activity of four commercially available alcohol-gel formulations under controlled laboratory conditions. The disk-diffusion method was chosen as a simple, accessible, and cost-effective screening tool to provide an objective comparative parameter between the activity of these products against pathogens of clinical interest. It is used for preliminary evaluations of antiseptics; although not a clinical standard, it has been documented in the literature for similar screening purposes.

A second, and perhaps more significant, finding emerged from the atypical results observed during the study. The data points marked as "R*" in Table 1, which denote "atypical growth of resistant colonies observed within the inhibition zone," were almost exclusively associated with Brand D. This consistent pattern strongly suggests that the observation was not due to a random breach in our sterile technique or procedural contamination. Instead, we interpret this as a critical finding related to the intrinsic characteristics of

Brand D. We hypothesize that the product either contained low-level microbial contamination from the manufacturing process or had an inadequate preservative system, which was insufficient to inhibit the growth of certain alcohol-tolerant microorganisms in the formula. Therefore, what could be initially dismissed as a methodological flaw is, in fact, a significant finding regarding the quality control of commercial antiseptic products. The possibility that a commercially sold alcohol gel may harbor viable, alcohol-resistant organisms is a serious concern with direct public health implications, underscoring the importance of rigorous quality assurance in producing these essential healthcare products.

5. Conclusions

The results of this study demonstrate that the antimicrobial efficacy of alcohol-based hand sanitizers is strongly influenced by their overall formulation, not just alcohol concentration. Brand C exhibited the most consistent and robust activity against the bacterial strains of this study, particularly against *Pseudomonas aeruginosa*, suggesting an optimized combination of alcohol and synergistic excipients. Brand B showed selective efficacy against certain strains, while Brand A displayed moderate activity with greater variability. Brand D failed to inhibit key pathogens, highlighting the need for careful quality control and formulation assessment.

These findings underscore the critical role of hand hygiene in healthcare settings. While alcohol content is essential, factors such as humectants, polymer composition, viscosity, and pH significantly modulate practical antimicrobial performance. The study confirms the bacteriostatic potential of selected alcohol gels against *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Enterococcus faecalis*. However, it also emphasizes the need for further investigations, including additional pathogens, vehicle-only controls, and clinical or simulated hand models, to fully evaluate product efficacy under real-world conditions.

Conflicts of Interest

The authors declare that there are no conflicts of interest related to this work. None of the authors has any collaboration, consultancy, or employment relationship with the companies or laboratories that manufacture the tested brands.

This research was conducted exclusively for exploratory and academic purposes, without external financial support or commercial involvement. The study design, data collection, analysis, and interpretation were carried out independently by the authors.

Funding Source

The authors declare that no specific grant or funding was received for this research from any public, commercial, or not-for-profit funding agency.

Acknowledgment

None.

Ethical Considerations

According to CNS Resolution N° 466/2012, which regulates research involving human subjects, and Law N° 11,794/2008/CON-CEA, which regulates animal research, this study does not fall under situations requiring submission to a Research Ethics Committee. The research was conducted exclusively in vitro using

reference microorganisms (ATCC strains, laboratory strains, and routine hospital isolates) and did not involve human participants, identifiable personal data, or animals. Therefore, institutional ethical approval was not necessary. The ethical oversight of research in Brazil is conducted by the CEP/CONEP system, whose jurisdiction is strictly limited to studies involving human subjects, including the use of identifiable data or human-derived biological materials. As our work employed only standardized bacterial strains in vitro, it is not classified as “research involving human beings” and is exempt from CEP/CONEP review.

Large Language Model

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) and DeepL Translate to assist with English language editing and translation. The authors reviewed and revised all tool-assisted text and take full responsibility for the content of the final manuscript.

Authors Contribution

All authors did the conceptualization. SKS and LBD performed data curation. SKS and LBD carried out formal analysis. SKS and LBD developed the methodology. LSB provided supervision. LSB completed validation. SKS and LBD wrote the original draft, and LSB conducted the review and editing.

Data Availability

The data generated and analyzed during this study are not publicly available because they consist of raw laboratory measurements obtained from in vitro disk-diffusion assays conducted exclusively for this research. All relevant data supporting the findings of this study are contained within the article and its supplementary materials. Additional details or anonymized datasets may be made available by the corresponding author upon reasonable request for academic and research purposes.

References

- Fontana RT. [Hospital infection and the historical evolution of infection]. *Rev Bras Enferm.* 2006;59(5):703-6. [PMID: 17340720, <https://doi.org/10.1590/s0034-71672006000500021>].
- L R. Dictionary of Medical and Health Technical Terms; 2003. 2nd ed. Rio de Janeiro: Guanabara Koogan. Available from: <https://www.estantevirtual.com.br/livro/dicionario-de-termos-tecnicos-de-medicina-e-saude-0M7-4908-000-BK?condition=novo&page=1>.
- Moura ME, Campelo SM, de Brito FC, Batista OM, de Araujo TM, Oliveira AD. [Nosocomial infection: study of prevalence at a public teaching hospital]. *Rev Bras Enferm.* 2007;60(4):416-21. [PMID: 18041525, <https://doi.org/10.1590/s0034-71672007000400011>].
- Portal de Boas Práticas em Saúde da Mulher dCedA. Hand hygiene as a strategy for preventing healthcare-associated infections (HAIs) and emergence of multidrug-resistant bacteria; 2018. Rio de Janeiro: Fiocruz; 2018 Apr 13 [cited 2025 Jan 19]. Available from: <https://www.arca.fiocruz.br/handle/icict/3s0268>.
- Cruz AFD, Abreu AO, Souza PA, Deveza B, Medeiros CT, Sousa VS, et al. Adaptation and validation of a method for evaluating the bactericidal activity of ethyl alcohol in gel format 70% (w/w). *J Microbiol Methods.* 2022;193:106402. [PMID: 34973997, <https://doi.org/10.1016/j.mimet.2021.106402>].
- NA G, et al.. Alcohol Sanitizer; 2023. StatPearls Publishing; 2023 Aug 9. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK513254/>.
- Berardi A, Perinelli DR, Merchant HA, Bisharat L, Basheti IA, Bonacucina G, et al. Hand sanitisers amid CoViD-19: A critical review of alcohol-based products on the market and formulation approaches to respond to increasing demand. *Int J Pharm.* 2020;584:119431. [PMID: 32461194, PMCID: PMC7229736, <https://doi.org/10.1016/j.ijpharm.2020.119431>].
- Rocha VFD, da Silva EN, Azevedo J, Ribeiro MT, Reis MG, Barros TF, et al. The impact of COVID-19 on microbiological profile and antibiotic consumption in ICU: a retrospective study in an infectious disease hospital in Brazil. *Braz J Infect Dis.* 2024;28(1):103705. [PMID: 38065221, PMCID: PMC10764251, <https://doi.org/10.1016/j.bjid.2023.103705>].
- Kalingamudali YT, Jalini P, Jayamaha A, Jayaratne DL. Knowledge, attitudes, and practices on the usage of antiseptics prior to invasive medical procedures: Evidence from Sri Lankan Healthcare Professionals. *SAGE Open Med.* 2023;11:20503121231199654. [PMID: 37750100, PMCID: PMC10518135, <https://doi.org/10.1177/20503121231199654>].
- Wichaidit W, Naknual S, Kleangkert N, Liabsuetrakul T. Installation of pedal-operated alcohol gel dispensers with behavioral nudges and changes in hand hygiene behaviors during the COVID-19 pandemic: A hospital-based quasi-experimental study. *J Public Health Res.* 2020;9(4):1863. [PMID: 33150146, PMCID: PMC7607228, <https://doi.org/10.4081/jphr.2020.1863>].
- Leslie RA, Zhou SS, Macinga DR. Inactivation of SARS-CoV-2 by commercially available alcohol-based hand sanitizers. *Am J Infect Control.* 2021;49(3):401-2. [PMID: 32818578, PMCID: PMC7431409, <https://doi.org/10.1016/j.ajic.2020.08.020>].
- Eslami H, Das S, Zhou T, Müller-Plathe F. How Alcoholic Disinfectants Affect Coronavirus Model Membranes: Membrane Fluidity, Permeability, and Disintegration. *J Phys Chem B.* 2020;124(46):10374-85. [PMID: 33172260, <https://doi.org/10.1021/acs.jpcc.0c08296>].
- GF N. Bacterial resistance in hospital settings: a review;. [Undergraduate thesis]. Campina Grande: Federal University of Paraíba; 2010. Available from: <http://dspace.bc.uepb.edu.br/jspui/handle/123456789/4136>.
- org F. Foodborne outbreaks in Brazil: data updated in December 2016; 2017. [Internet]. 2017 [cited 2025 Feb 14]. Available from: <http://foodsafetybrazil.org/surtos-alimentares-no-brasil-dados-atualizados-em-dezembro-de-2016/>.
- org F. Escherichia coli O157:H7 - understand the outbreak in the US and actions taken; 2024. [Internet]. 2024 Nov 4 [cited 2025 Feb 14]. Available from: <https://foodsafetybrazil.org/escherichia-coli-o157h7-entenda-o-surto-e-as-tratativas/>.
- (WHO) WHO. E. coli; 2016. [Internet]. 2016 [cited 2025 Feb 14]. Available from: <http://www.who.int/mediacentre/factsheets/fs125/es/>.
- for Disease Control C, (CDC) P. Pseudomonas aeruginosa in Healthcare Settings; 2014. [Internet]. 2014 [cited 2025 Feb 14]. Available from: <https://www.cdc.gov/hai/organisms/pseudomonas.html>.
- García-Domínguez C, Martín P, Santos-Hurtado I, Blanco J, Blanco MT, Gómez-García AC. Susceptibility of Pseudomonas aeruginosa to anti-pseudomonal antibiotics in a general hospital in Spain from 1989 to 1992. *J Antimicrob Chemother.* 1994;33(5):1057-9. [PMID: 7993468, <https://doi.org/10.1093/jac/33.5.1057>].
- of Health BNHSAAM. Resolution no. 485, August 21, 2008;. [Internet]. [cited 2025 Oct 2]. Available from: http://www.cff.org.br/userfiles/file/resolucoes/res485_2008.pdf.
- Laboratorial SBdPCM. Recomendações da Sociedade Brasileira de Patologia Clínica/Medicina Laboratorial: Boas práticas em Microbiologia Clínica. Barueri, SP, Brazil: Manole; 2015. [PMCID: PMCID].
- ASTA. Science ASSIST Standard Operating Procedure: Susceptibility testing of antiseptics and disinfectants; 2023. Jan. 2023 [cited 2025 Oct 2]. Available from: <https://assist.asta.edu.au/>.
- Arbogast JW, Okeke B, Brill F, Bradley T, Fraise A, Wilkinson M, et al. Advancement of the standard in vivo hand rub test methods: a critical comparison of the health care personnel handwash (ASTM E1174) and the hygiene handrub (EN1500). *BMC Proceedings.* 2011;5(S6). [https://doi.org/10.1186/1753-6561-5-s6-p269].

23. Golin AP, Choi D, Ghahary A. Hand sanitizers: A review of ingredients, mechanisms of action, modes of delivery, and efficacy against coronaviruses. *Am J Infect Control*. 2020;48(9):1062-7. [PMID: 32565272, PMCID: PMC7301780, <https://doi.org/10.1016/j.ajic.2020.06.182>].
24. Bloomfield SF, Aiello AE, Cookson B, O'Boyle C, Larson EL. The effectiveness of hand hygiene procedures in reducing the risks of infections in home and community settings including handwashing and alcohol-based hand sanitizers. *American Journal of Infection Control*. 2007;35(10):S27-64. [<https://doi.org/10.1016/j.ajic.2007.07.001>].
25. Jy K. Higiene das mãos: comparação da eficácia antimicrobiana do álcool – formulação gel e líquida – nas mãos com matéria orgânica; 2004. [PMCID: PMCID].
26. Meneguetti MG, Laus AM, Ciol MA, Auxiliadora-Martins M, Basile-Filho A, Gir E, et al. Glycerol content within the WHO ethanol-based handrub formulation: balancing tolerability with antimicrobial efficacy. *Antimicrob Resist Infect Control*. 2019;8:109. [PMID: 31285821, PMCID: PMC6591802, <https://doi.org/10.1186/s13756-019-0553-z>].
27. Suchomel M, Weinlich M, Kundi M. Influence of glycerol and an alternative humectant on the immediate and 3-hours bactericidal efficacies of two isopropanol-based antiseptics in laboratory experiments in vivo according to EN 12791. *Antimicrob Resist Infect Control*. 2017;6:72. [PMID: 28670452, PMCID: PMC5488408, <https://doi.org/10.1186/s13756-017-0229-5>].
28. d'Angelo I, Provenzano R, Florio E, Pagliuca C, Mantova G, Scaglione E, et al. Alcohol-Based Hand Sanitizers: Does Gelling Agent Really Matter? *Gels*. 2022;8(2). [PMID: 35200468, PMCID: PMC8871833, <https://doi.org/10.3390/gels8020087>].
29. Zhou Y, Yan H, Zhou Q, Feng R, Zhai B. Impact of COVID-19 control measures on *Legionella pneumophila* infections in children in Henan, China. *J Infect*. 2023;87(1):85-7. [PMID: 37116618, PMCID: PMC10131741, <https://doi.org/10.1016/j.jinf.2023.04.020>].
30. Thaddeus N, Francis E, Jane O, Obumneme A, Okechukwu E. Effects of some common additives on the antimicrobial activities of alcohol-based hand sanitizers. *Asian Pacific Journal of Tropical Medicine*. 2018;11(3). [<https://doi.org/10.4103/1995-7645.228437>].
31. (ANVISA) BNHSA. Module 10: Clinical Microbiology Manual; [Internet]. [cited 2025 Feb 14]. Available from: https://www.gov.br/anvisa/pt-br/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes/modulo-10_manual-de-microbiologia.pdf.
32. Chopin-Doroteo M, Krotzsch E. Soap or alcohol-based products? The effect of hand hygiene on skin characteristics during the COVID-19 pandemic. *J Cosmet Dermatol*. 2023;22(2):347-53. [PMID: 36409429, <https://doi.org/10.1111/jocd.15523>].
33. Loveday HP, Wilson JA, Pratt RJ, Golsorkhi M, Tingle A, Bak A, et al. epic3: national evidence-based guidelines for preventing healthcare-associated infections in NHS hospitals in England. *J Hosp Infect*. 2014;86 Suppl 1:S1-70. [PMID: 24330862, PMCID: PMC7114876, [https://doi.org/10.1016/S0195-6701\(13\)60012-2](https://doi.org/10.1016/S0195-6701(13)60012-2)].
34. Pratt RJ, Pellowe CM, Wilson JA, Loveday HP, Harper PJ, Jones SR, et al. epic2: National evidence-based guidelines for preventing healthcare-associated infections in NHS hospitals in England. *J Hosp Infect*. 2007;65 Suppl 1:S1-64. [PMID: 17307562, PMCID: PMC7134414, [https://doi.org/10.1016/S0195-6701\(07\)60002-4](https://doi.org/10.1016/S0195-6701(07)60002-4)].
35. Health BNHSAMo. Resolution no. 485, August 21, 2008; [PMCID: PMCID].
36. Andrade D, Beraldo CC, Watanabe E, Oliveira BA, Ito IY. Atividade Antimicrobiana in Vitro Do Álcool Gel a 70% Frente Às Bactérias Hospitalares E Da Comunidade. *Medicina (Ribeirão Preto)*. 2007;40(2):250-4. [<https://doi.org/10.11606/issn.2176-7262.v40i2p250-254>].
37. Alhalwani A, Husain A, Saemaldahar A, Makhdom F, Alhakami M, Ashi R, et al. The impact of alcohol hand sanitizer use on skin health between healthcare worker: Cross-sectional study. *Skin Res Technol*. 2024;30(1):e13527. [PMID: 38186055, PMCID: PMC10772474, <https://doi.org/10.1111/srt.13527>].
38. Lopez-Gigosos RM, Mariscal-Lopez E, Gutierrez-Bedmar M, Mariscal A. Effect of Long-Term Use of Alcohol-Containing Handwashing Gels on the Biofilm-Forming Capacity of *Staphylococcus epidermidis*. *Int J Environ Res Public Health*. 2023;20(6). [PMID: 36981945, PMCID: PMC10049709, <https://doi.org/10.3390/ijerph20065037>].



Review Article

Search Filters to Identify Automation in HEOR: An Umbrella Review of Performance and Overlap

Maria Rizzo¹, Omar Irfan^{2,*}

1-Independent Consultant, London, United Kingdom

2-Independent Consultant, Milton, Ontario, Canada

ARTICLE INFO

Article history:

Received 3 Aug. 2025

Received in revised form 28 Nov. 2025

Accepted 7 Dec. 2025

Published 21 Dec. 2025

Keywords:

Automation

Artificial intelligence

Health economics and outcomes research

HEOR

Search filters

ABSTRACT

Background: Search strategies used to identify evidence on automation in Health Economics and Outcomes Research (HEOR) often lack sensitivity and specificity, resulting in information overload or missed studies. This umbrella review evaluated and compared the performance and overlap of search filters commonly used to retrieve automation-related HEOR evidence.

Methods: Systematic literature reviews (SLRs) and search filters focusing on any form of automation in HEOR, were included. Searches (January 01, 2023–July 03, 2024) were conducted in EMBASE, ISSG Resource, and Google Scholar. Subject headings, search terms, and performance metrics were extracted. Reference lists were cross-checked. Screening was performed by one reviewer, with 20% verified by a second reviewer. The PRESS checklist was used to assess search strategy quality. The protocol was registered with Open Science Framework (OSF).

Results: Seven SLRs and one standalone filter, reporting 11 search strategies, met inclusion criteria. HEOR relevance was defined by studies applying search filters in contexts of SLRs, indirect treatment comparisons, and economic modelling. Included SLRs retrieved between 5–273 studies. PubMed was the most frequently searched database. Commonly exploded subject-headings included “artificial intelligence,” “deep learning,” “machine learning,” and “natural language processing,” with “artificial intelligence” the most frequent free-text term. Inclusion rates varied: title/abstract (1%–8%), full-text (27%–86%), final inclusion (0.13%–2.31%). Time to include one study ranged from 0.6–8 hours.

Conclusions: Considerable variability in search filter performance was observed, causing lower specificity and inefficient evidence retrieval. Standardized, high-performing search strategies are needed to enhance efficiency and reliability in identifying automation-related HEOR evidence.

1. Introduction

Health Economics and Outcomes Research (HEOR) plays a pivotal role in ensuring patients have access to new technologies that have been evaluated for efficacy, safety, and cost-effectiveness, thereby optimizing resource allocation within the healthcare sector [1]. With the escalating costs of new and innovative medicines and the constraints on public funding, the efficient allocation of healthcare resources is imperative to ensure that cost-effective treatments are prioritized and accessible to those in need. However, the health technology appraisal (HTA) landscape is also becoming more complex, with frequent changes to reimbursement legislations, and with more advanced and complex HEOR analytics (e.g., multi-level network meta regression [ML-NMR], synthetic control arms, quantitative bias analysis), it is not surprising that the consideration on the use of artificial intelligence (AI) in HEOR is being considered, to expedite timelines, to get new technology to patients on time [2].

HTA bodies have begun using AI in reimbursement submissions. Most recently, the National Institute for Health and Care Excellence (NICE) accepts the use of AI, however, its use must be declared, the methods transparently reported, should only be applied if there is a rationale, and its use adds value, and most critically, that the AI only augments human involvement, and does not replace it [3]. Additionally, CADTH outlines the use of AI in systematic literature reviews (SLRs) to automate data extraction and searching [4]. Additionally, the Institut National d'Excellence en Santé et Services Sociaux (INESSS) created a proof-of-concept demonstrating the application of a large language model (LLM) based on GPT-4 to aid title and abstract screening [5].

As AI continues to be integrated into HTA submissions, researchers, decision-makers, and HEOR experts must remain up to date with the latest developments in AI methodologies, applications, and software. The rapid expansion of research on AI and automation in HEOR has created a vast, continually growing body of literature, posing significant challenges for efficiently identifying the most relevant studies. Current search strategies often struggle to balance sensitivity (capturing all relevant studies) and specificity (excluding irrelevant studies), leading to either information overload or missed key evidence [6]. There is therefore an urgent need for robust, validated search filters that can accurately navigate the complex, abundant research landscape in this domain. This umbrella review aims to systematically identify, evaluate, and compare the performance and overlap of search filters used

*Corresponding author: Omar Irfan, Independent Consultant, Milton, Ontario, Canada, Email: omarirfan1@hotmail.com

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Rizzo M, Irfan O. Search Filters to Identify Automation in HEOR: An Umbrella Review of Performance and Overlap. ASIDE Health Sci. 2025;2(2):9-17, doi:10.71079/ASIDE.HS.122125202

Title and Abstract Inclusion Rate:

$$\text{Inclusion Rate}_{\text{TA}} = \frac{\text{Full-texts Retrieved}}{\text{Title and Abstract Screened}}$$

Full-text Inclusion Rate:

$$\text{Inclusion Rate}_{\text{FT}} = \frac{\text{Final Included Studies}}{\text{Full-texts Retrieved}}$$

Overall Inclusion Rate:

$$\text{Inclusion Rate}_{\text{Overall}} = \frac{\text{Final Included Studies}}{\text{Title and Abstract Screened}}$$

Figure 1: Search performance metrics

FT, Full-text; TA, Title and Abstract

to identify evidence on automation in HEOR. To our knowledge, few studies have compared the overlap and performance of filters used to retrieve automation studies in HEOR-relevant contexts, highlighting a critical gap in the literature.

2. Methods

The umbrella review followed the reporting requirements of the Preferred Reporting Items for Systematic reviews and Meta-Analyses 2020 (PRISMA) [7] and the methodology outlined in the Cochrane Handbook for Overviews of Reviews [8]. The protocol for this systematic review is registered in the Open Science Framework (OSF) platform (<https://osf.io/dzumv>).

Studies were included if they were (a) SLRs, with reproducible search filters, or publications on search filters; (b) SLR or search filter was on any automation; (c) SLRs on automation was tested in HEOR, defined as SLRs, indirect treatment comparisons (ITCs) or health economic modeling; (d) full-text publication; (e) written in English; and (f) published in 2023 or later, to capture studies since the availability of ChatGPT.

2.1. Search Strategy

The search was run from January 01, 2023, to July 03, 2024. The review was limited to studies published from January 2023 onwards to capture the post-ChatGPT era, during which large language models (LLMs) became widely accessible and integrated into academic and clinical discourse. This period also reflects a shift in indexing terminology, focus areas, and applications relevant to current LLM capabilities [9].

Terms on automation and artificial intelligence (AI) were combined with terms for SLR, ITCs, or economic models. The full search strategy is available in (**Supplement Table 1**). Studies were searched in EMBASE via OVID. EMBASE was selected as the primary database because it offers comprehensive coverage of biomedical, health technology, and pharmacoeconomic literature, and provides more detailed indexing of artificial intelligence and automation-related subject headings than other databases, such as MEDLINE. Its broad inclusion of conference abstracts and global publications also enhances the capture of emerging AI-HEOR evidence, which is often presented in early or technical formats before journal publication.

The reference lists of included studies were checked, and forward citation chasing was conducted in Google Scholar to identify any missing SLRs. The InterTASC Information Specialists' Sub-Group

(ISSG) Search Filter Resource was also manually searched to identify additional search filters.

2.2. Selection Process

Studies were screened by one reviewer, with 20% of records checked by a second reviewer to ensure accuracy in study selection. Any discrepancies were resolved through discussion and consensus, with a third reviewer available for arbitration if needed, although this was not required. This process was designed to ensure consistency and mitigate potential selection bias inherent in single-reviewer screening approaches.

2.3. Data Extraction

All studies were extracted in Microsoft Excel® using a piloted data extraction template. A second reviewer validated all data. Details on SLR characteristics (objectives, search sources, inclusion criteria, exclusion criteria); search terms; and search results (number of records screened, number of reports retrieved, and number of studies included) were extracted. The time to include studies during the SLR phases of title/abstract screening and full-text screening was directly extracted from the included reviews, not calculated by the authors.

2.4. Quality Assessment

The Peer Review of Electronic Search Strategies (PRESS) was used to evaluate the electronic search strategies [18].

2.5. Evidence Synthesis

Subject headings and search terms used across search filters were qualitatively summarized. In addition, search performance was assessed at the title and abstract, full-text, and overall levels (**Figure 1**).

3. Results

A total of 11,319 records were identified from EMBASE via OVID. No duplicates were identified after the records were imported into EndNote X9 software for deduplication. Out of these, 25 reports were retrieved for full-text review. Among these 25 reports, five SLRs were included. A flow of literature and reasons for exclusion are outlined in (**Figure 2**).

An additional three reports were included: a search filter from the ISSG Search Filter Resource [17] and two SLRs identified through forward citation chasing in Google Scholar [10, 15]. Among the eight reports [10–17], seven were search filters used in an SLR [10–16] and one was a standalone search filter on the ISSG website [17]. One SLR, combined searches from four previously identified reviews on SLR automation [15]. As a result, this umbrella review identified 11 individual search filters.

Among the seven SLRs [10–16], two had objectives related to the automation and processing of biomedical literature, as well as to the preparation of SLRs [11, 14]. Two evaluated the automation of any SLR task [10, 15]; one restricted to the automation of data extraction in preparation of an SLR [13]; another to the automation of data extraction from grey literature and soft data [12]. Lastly, one SLR was restricted to automation used in the screening of cancer topics [16]. The number of title and abstract records screened across the SLRs ranged from 3,947 [16] to 27,542 [13]; between 7 [16] and 721 [14] full-texts assessed for eligibility; and between 5 [16] and 273 [11] reports included. A summary of the seven SLRs is provided in (**Table 1**).

Table 1: Summary of Included SLRs

Primary author	Year	Objectives	Inclusion criteria	Exclusion criteria	Title/abstract records	Full-text reports	Included reports
de la Torre-Lopez [10]	2023	To determine which phases of the SLR process have been automated using AI, identify supporting AI techniques, and assess human involvement in AI-based SLR automation.	Automation of one or more SLR phases using an AI-based approach.	Non-English language reports.	9,027	50	34
Santos [11]	2023	To examine AI use in automated or semi-automated analysis of biomedical literature and identify state-of-the-art methods and knowledge gaps.	AI methods and tools applied to medical literature; AI tools used in preparation of SLRs.	AI in diagnosis or treatment; AI applied to non-literature medical documents (e.g., electronic medical records).	12,145	316	273
Schmidt [12]	2023a	To review methods and tools for (semi-)automated data extraction in medical SLRs using a living SLR approach.	Original NLP-based full-text studies describing complete implementation and evaluation cycles; published from 2005; data from RCTs or comparative observational studies.	Image-only processing; protocol-only or synthesis-only studies; electronic health records or genetic data mining.	27,542	278	76
Schmidt [13]	2023b	To provide an overview of automated data extraction methods for health-related research using grey and soft data.	Original extraction tools or methods; non-peer-reviewed healthcare datasets; English-language full texts.	Patient-level electronic health records; genomic or biological data extraction.	8,927	131	84
Doneva [14]	2024	To identify BioNLP tasks addressed using large language models, map LLM architectures, and assess methodological transparency.	LLMs applied to biomedical text; SLRs or meta-analyses using LLMs for automation.	Non-English language reports; clinical questionnaires or surveys; reviews.	13,825	721	197
Toth [15]	2024	To determine which SLR stages were automated, the applied tools and data sources, and the research impact of SLR automation studies.	SLRs using automation tools to assist or replace human judgment tasks; full-text reports.	Non-English language reports.	5,321	411	123
Yao [16]	2024	To evaluate accuracy and workload savings of AI-based screening tools versus human reviewers in cancer-related SLRs.	AI tools for automated screening; cancer topics; reported sensitivity/specificity or workload savings.	Non-English language reports; letters, editorials, commentaries; non-public tools.	3,947	7	5

AI, Artificial Intelligence; LLM, Large Language Models; MA, Meta-analysis; NLP, Natural Language Processing; PDFs, Portable Document Formats; RCT; Randomized Controlled Trial; SLR, Systematic Literature Review

Across the seven SLRs [10–16], 16 bibliographic databases were searched, with PubMed used most frequently (6 out of 7 SLRs) [11–16]; followed by EMBASE [11, 14, 16], Web of Science [10, 11, 13] and Institute of Electrical and Electronics Engineers (IEEE) [10, 11, 13], each searched by three SLRs. In addition to IEEE, three more bibliographic databases were identified and searched by two SLRs each, specializing in computing and technology; these databases were: the Association for Computing Machinery (ACM) Digital Library [10, 11], ACL Anthology [12, 13], and dblp [12, 13]. For a full list of bibliographic databases searched by the SLRs, see (Table 2). One of the reports identified was a search filter published by the University of Alberta, and available on the ISSG Search Filter Resource website [17]. The search filter, which is not validated, was translated for EMBASE, MEDLINE,

PsycINFO, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Scopus, and Cochrane Library [17].

Of the eight reports summarizing 11 search filters on automation, five used subject headings [11–14, 17]. Four search filters exploded the subject headings ‘artificial intelligence’ [11–13, 17], ‘deep learning’ [11–13, 17], ‘machine learning’ [11–13, 17], and ‘natural language processing’ [11–14]. Followed by two search filters, each exploding subject headings on ‘data mining’ [12, 13], ‘supervised machine learning’ [13, 17], ‘support vector machine’ [13, 14], and ‘unsupervised machine learning’ [13, 17]. Subject headings such as ‘artificial intelligence’, ‘automation’, ‘data mining’, ‘machine learning’ (using both the exact term and the wildcard search ‘machine learn*’), and ‘natural language processing’ were also

Table 2: Electronic Databases Searched for Literature on AI

Database	de la Torre-Lopez	Santos	Doneva	Schmidt(a)	Schmidt(b)	Toth	Yao
PubMed		x	x	x	x	x	x
Embase		x	x				x
IEEE Xplore	x	x	x				
Web of Science	x	x		x			
ACL- Anthology				x	x		
ACM Library	x	x					
Cochrane Library		x					x
dblp				x	x		
Scopus	x				x		
arXiv					x		
CINAHL		x					
Google Scholar		x					
MedRxiv					x		
Prospero						x	
SciELO		x					

ACM, Association for Computing Machinery; ACL, Association for Computational Linguistics; ACM, Association for Computing Machinery; AI, Artificial Intelligence; CINAHL, Cumulative Index to Nursing and Allied Health Literature; dblp, database systems and logic programming; IEEE, Institute of Electrical and Electronics Engineers; SciELO, Scientific Electronic Library Online.

searched as major headings in PubMed and as descriptors in EMBASE. For a full list of subject headings searched, see (Table 3).

Across the 11 search filters, 196 variations of free text terms were used to capture studies on automation. Less than a third of these free-text terms were used by more than one search, with only twelve used in three or more search filters. “Artificial intelligence” was the most frequently utilized in all search filters, followed by “machine learning” and “natural language processing”, each present in six search filters (see (Figure 3)). For the full list of search terms used, please see (Supplement Table 2).

When grouping subject headings and free text terms into common concepts, for example, the following search terms were grouped under machine learning (general terms): “machine learning”, “ml”, [machine NEXT/1 (intelligence OR learning)], exp machine learning/, machine learning/de); “machine learning (general terms)” and “automation” were the most common constructs searched across the search filters (8 search filters each). This was followed by “mining” and “natural language processing” (7 search filters each), and “text grouping”, “neural network”, “identification, interpretation and pattern recognition”, “support vector machine”, and “artificial intelligence” (6 search filters each). “Deep learning” and “machine learning approaches” were searched in 5 search filters, while “semi automation” and “decision support systems” were searched in 4. The remaining concepts were searched in 3 (“AI models and application”, “extraction”, and “text analysis”), and 2 search filters (“knowledge systems”). The cluster with the most variation in terms used was “AI models and applications”, followed by “machine learning approaches”. All concepts identified are shown in (Figure 4), with a list of terms for each concept in (Supplement Table 3). Similar search filter characteristics were identified by the PRESS checklist (Supplement Table 4): only 5 of the included reports used subject headings, and the majority of the

search filters reported a narrow list of keywords. Only five of the 11 search filters included truncation. No incorrect use of system syntax and spelling errors were identified. A mapping framework (Supplement Table 5) was developed to link each included automation review or search filter to HEOR-relevant evidence-generation tasks. Reviews were categorized by their core automation focus (e.g., AI/NLP, ML, decision systems) and cross-referenced with HEOR functions, including SLR automation for economic model inputs, ITCs, and HTA submissions. Each review/filter was coded for “HEOR applicability (Y/N)” with justification provided under the column “Basis.”

Among the seven SLRs [10–16], one of which searched using 4 previously identified search filters, and combined using “or” [15]; none had a title and abstract inclusion rate (defined as the proportion of studies selected for full-text retrieval out of the total number of title and abstract records screened) above 8%, with two searches below 1% [10, 16] (see (Table 4)). The full-text inclusion rate (defined as the proportion of studies included after full-text review compared to the total number of full-text articles retrieved for evaluation) varied from 27% [13, 14] to 86% [11]. The overall inclusion rate (defined as the proportion of studies included after the full review compared to the total number of title and abstract records initially screened) ranged from 0.13% [16] to 2.31% [15]. When evaluating the number of hours needed to include one study the SLR by Santos [11], followed by Toth [15] had the best performance (0.6 hours and 0.8 hours, respectively), followed by Doneva (1.1 hours) [14], Schmidt(b) (1.2 hours) [12], De la Torre-Lopes (2.8 hours), Schmidt(a) (4.0 hours) [13], with Yao, which was only interested in AI used in HEOR among oncology studies, requiring 8 hours of screening to identify one included study [16] (see (Table 4)). Of note: performance metrics could not be calculated for the standalone search filter identified from the ISSG Search Filter resource, as it did not present a flow of literature

Table 3: Subject Headings Used to Identify Literature on AI

Search Filter	Corresponding Citations
artificial intelligence/	Campbell [17], Schmidt(a) [12], Schmidt(b) [13]
artificial intelligence/de	Doneva [14]
artificial intelligence[MH]	Schmidt(a) [12]
artificial intelligence[MeSH Terms]	Santos [11]
artificial neural network/exp	Doneva [14]
automated pattern recognition/	Doneva [14]
automation/	Schmidt(b) [13]
automation[MH]	Schmidt(a) [12]
automated[MH]	Schmidt(a) [12]
automating[MH]	Schmidt(a) [12]
data mining/	Schmidt(a) [12], Schmidt(b) [13]
data mining[MH]	Schmidt(a) [12]
decision trees/	Schmidt(b) [13]
deep learning/	Campbell [17], Schmidt(a) [12], Schmidt(b) [13]
deep learning [MeSH Terms]	Santos [11]
knowledge bases/	Schmidt(b) [13]
learning algorithm*[MH]	Schmidt(a) [12]
literature mining[MH]	Schmidt(a) [12]
machine learn*[MH]	Schmidt(a) [12]
machine learning/	Campbell [17], Schmidt(a) [12], Schmidt(b) [13]
machine learning/de	Doneva [14]
machine learning [MeSH Terms]	Santos [11]
natural language processing/	Schmidt(a) [12], Schmidt(b) [13], Doneva [14]
natural Language processing[MH]	Schmidt(a) [12]
natural Language processing [MeSH Terms]	Santos [11]
neural networks (computer)/	Schmidt(b) [13]
neural networks/	Schmidt(a) [12]
neural network [MeSH Terms]	Santos [11]
predictive modelling[MH]	Schmidt(a) [12]
semi-automated[MH]	Schmidt(a) [12]
semi-automating[MH]	Schmidt(a) [12]
semi-automation[MH]	Schmidt(a) [12]
supervised machine learning/	Schmidt(a) [12], Schmidt(b) [13]
support vector machine/	Schmidt(b) [13], Doneva [14]
text mining[MH]	Schmidt(a) [12]
unsupervised machine learning/	Schmidt(a) [12], Schmidt(b) [13]

MH, Medical Subject Headings; MeSH, Medical Subject Headings.

as part of an SLR. These time metrics were directly extracted from the included reviews, not calculated by the authors.

For this umbrella review, a comprehensive search was conducted in EMBASE (via Ovid), the InterTASC Information Specialists' Sub-Group (ISSG) Search Filter Resource, and Google Scholar. The search covered studies published between January 1, 2023, and July 3, 2024, to capture the post-ChatGPT era when large language

models (LLMs) became widely adopted. Search terms combined concepts for automation and artificial intelligence (AI) with terms representing systematic evidence synthesis (systematic review, indirect treatment comparison, economic model). In contrast, the included SLRs and search filters used a variety of databases and approaches. Across the eight included publications, 16 databases were searched, with PubMed being the most frequently used (6/7 SLRs), followed by EMBASE (3/7). This distinction between our

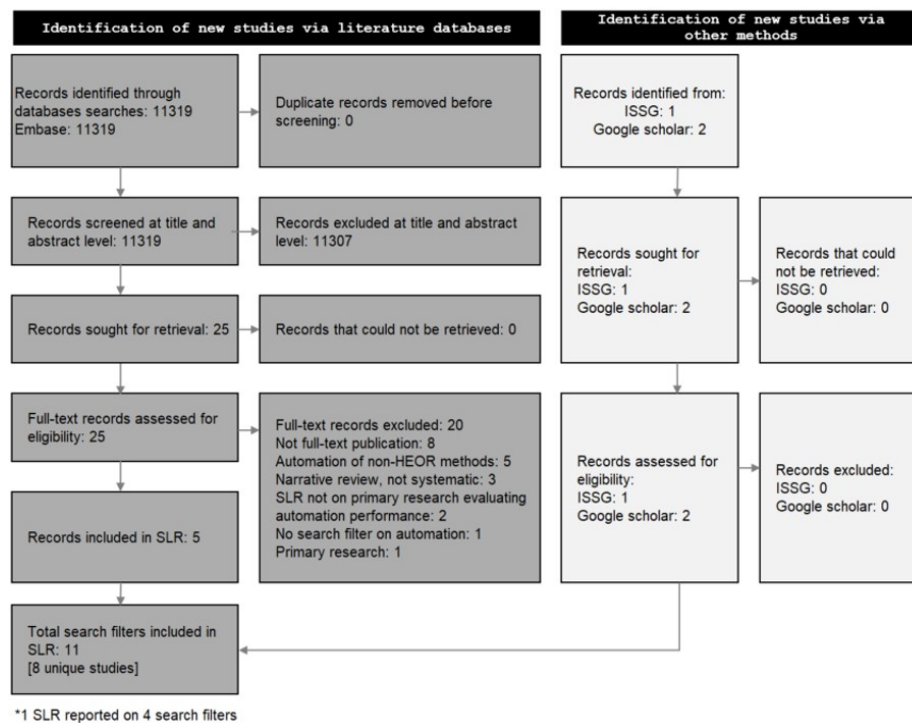


Figure 2: PRISMA flow diagram of the Umbrella review study selection

HEOR, Health Economics Outcomes Research; ISSG, InterTASC Information Specialists' Sub-Group; SLR, Systematic Literature Review

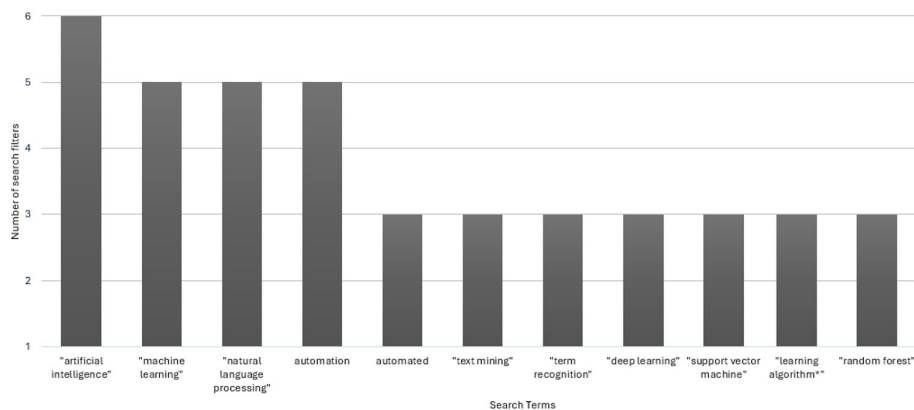


Figure 3: Free text terms used in three or more search filters.

umbrella review's search process and that of the included reviews is summarized in Table X ("Databases Searched: This Review vs. Included SLRs"). Among the eight included publications (seven SLRs and one standalone filter), three reviews (38%) explicitly evaluated automation methods within HEOR-relevant contexts, such as SLRs for cost-effectiveness models or indirect treatment comparisons.

4. Discussion

There is very little overlap or consensus among existing search filters for identifying papers on AI in HEOR. Of the 11 search filters identified, only five used subject headings, and among these, no subject heading was searched across all five. For example,

only four out of 31 different subject headings ('artificial intelligence', 'deep learning', 'machine learning', and 'natural language processing') were searched across four of the five filters. There were also differences in how subject headings were searched: as major headings in PubMed and as descriptors in EMBASE. The PRESS checklist further confirmed the deficits in the search filters, identifying fewer reports using subject headings and the majority of the searches reporting a narrow list of keywords. Quantitative overlap analysis (e.g., Jaccard similarity) was not conducted because the terminology across filters was highly heterogeneous and inconsistently formatted, making reliable computation of similarity indices infeasible without extensive normalization.

A lack of consistency in search filters was identified among free-text terms, with less than a third used by more than one search, and only 12 by more than three searches. With the abundance of

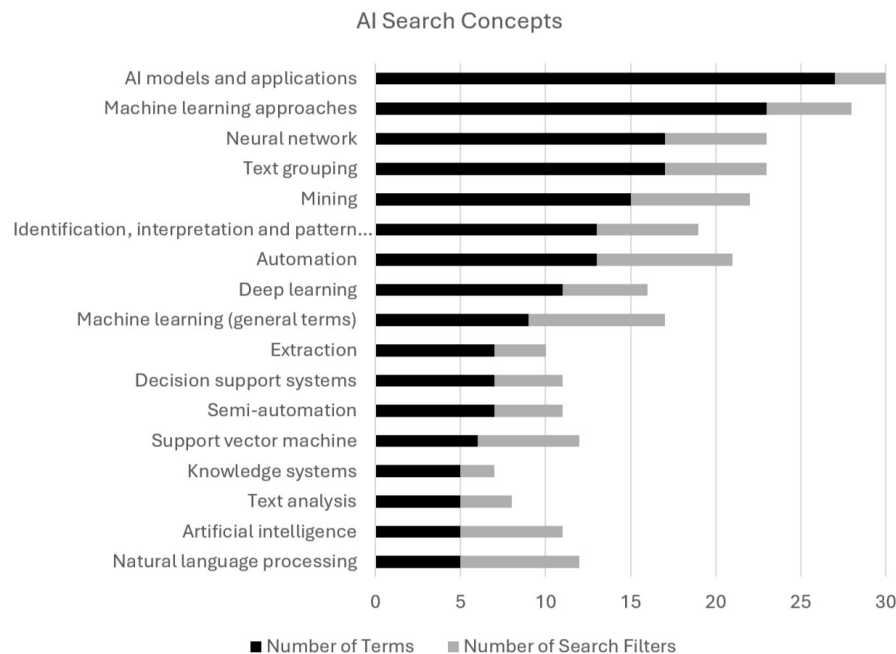


Figure 4: AI Constructs Searched in the search filters.

Table 4: Inclusion rates at different stages of SLR in the included studies

Primary Author	Title and Abstract (%)	Full Text (%)	Overall (%)
de la Torre-Lopez [10]	0.55%	68.00%	0.38%
Doneva [14]	5.22%	27.32%	1.42%
Santos [11]	2.60%	86.39%	2.25%
Schmidt(a) [12]	1.01%	27.34%	0.28%
Schmidt(b) [13]	1.47%	64.12%	0.94%
Toth [15]	7.72%	29.93%	2.31%
Yao [16]	0.18%	71.43%	0.13%

SLR, Systematic Literature Review

AI research in the HEOR field and no uniform, validated search filter available, it is not surprising that the SLRs ranged from just under 4,000 to over 27,000 records. This equated to a wide range of title and abstract screening time, from 0.6 hours to 8 hours, to identify one relevant study. This wide variation underscores the potential importance of a validated, optimized filter that can supplement rapid access to high-quality evidence supporting timely health technology assessment and reimbursement decisions.

To the best of the author's knowledge, this umbrella review appears to be the first systematic review to explore the identification and performance of existing AI search filters in HEOR. One of the identified SLRs combined the search strategies of previous reviews to provide a comprehensive overview of automation studies in the SLR process, but did not evaluate the search filters themselves.

Table 5: Number of hours required at different stages of SLR in the included studies

Primary Author	Hours for Title Abstract Screening	Hours for Full Text Screening	Overall Screening Hours	Ratio of Hours to Each 1 Included Study
de la Torre-Lopez [10]	90	5	95	2.8
Doneva [14]	138	72	210	1.1
Santos [11]	121	32	153	0.6
Schmidt(a) [12]	275	28	303	4.0
Schmidt(b) [13]	89	13	102	1.2
Toth [15]	53	41	94	0.8
Yao [16]	39	1	40	8.0

SLR, Systematic Literature Review

Our research identified seven additional search filters that were not included in the previous publication. Our research provides a focused synthesis of the overlap among existing search filters for retrieving AI studies. It explores performance using metrics such as the title-abstract inclusion rate, full-text inclusion rate, and overall inclusion rate.

With key decision-making bodies such as CADTH, INESSS, and NICE exploring the use of AI in HTA submissions, as a community, we must stay ahead of new research to adopt the most reliable and proven automation methods in HEOR. To do this, we must develop

a validated search filter that balances the sensitivity and specificity of AI records, specifically for HEOR, to make this manageable. We observed from our search that we had to screen 11,319 records, to full-text review 25 reports, and to include 5 reports from the electronic database search. Most research was excluded at the title and abstract stages because it focused on the direct application of AI in medical and healthcare settings, not HEOR. By creating a tested and validated search strategy, it is more viable that a living SLR on this topic can be maintained, with insights shared with HEOR experts and decision-makers.

Although this was a systematic, umbrella review of search filters used in AI research for HEOR, a validated search filter was not used, as none had yet been developed. In addition, only a single screening was conducted with quality checks, which may indicate some human error in missing studies; however, this was mitigated by reference checking and forward citation chasing. Only EMBASE was searched, while the field of AI may require automation-specific databases such as the ACM Digital Library and IEEE; this needs to be thoroughly investigated. A systematic search restricted to the EMBASE database represents a methodological limitation, especially in the context of AI and LLMs, where relevant work often appears in interdisciplinary and technical venues. Rapidly evolving AI terminology can shorten the longevity of search filters, as keywords and concepts may quickly become outdated or replaced. This can reduce search precision and recall over time, requiring regular updates or adaptive filtering to maintain relevance and effectiveness. As included SLRs were not quality-assessed using a validated tool, such as the AMSTAR2 tool, this may affect the strength of the conclusions. Lastly, restricting studies published in 2023 or later might introduce bias and limit the generalizability of the review's findings. Broader multilingual and multi-database searches could yield a more accurate picture of true variability across automation-related reviews.

To ensure technical accuracy and reproducibility, all search filters extracted from the included reviews were verified against the native database syntax for both PubMed (MEDLINE) and EMBASE (via Ovid). Verification focused on identifying and correcting invalid or outdated field tags, truncation errors, Boolean logic inconsistencies, and incorrect use of subject headings. For PubMed, MeSH terms were verified against the current 2024 MeSH database. Only exact headings (e.g., "Artificial Intelligence"[Mesh]) were retained; wildcards and free-text modifiers were removed from MeSH field tags. Free-text terms were confirmed to use the [tiab] or [tw] fields as appropriate. For EMBASE, Emtree terms were verified using the 2024 Emtree browser. Exploded subject headings (e.g., 'artificial intelligence'/exp) were checked for availability and correct hierarchy placement. Incorrect truncation (e.g., "machine learn"/de*) and misplaced slashes were corrected according to Ovid syntax.

Future studies must focus on using these initial findings to create a validated search filter on AI in HEOR. Filters will need to explore the right AI search terms, but the most appropriate combination of terms to identify research in HEOR specifically. Suppose a search can be highly specific while still sensitive enough to ensure relevant studies are not missed. In that case, we can start standardizing how the industry searches for research in a rapidly evolving field. However, any search filter for AI must acknowledge that periodic review will be required as new AI methods emerge that may require additional search concepts, and as older methods are retired.

This umbrella review identified significant variability in the search filters used for identifying literature on automation in HEOR, highlighting the lack of a standardized approach. The low inclusion rates across the SLRs indicate challenges in optimizing search

filter performance. The variation in search strategies and the high yield of records on AI in HEOR call for more research to develop, validate, and standardize search filters in this field to optimize efficiency and keep up to date with the ongoing stream of new evidence.

5. Conclusion

This umbrella review identified significant variability in the search filters used for identifying literature on automation in HEOR, highlighting the lack of a standardized approach. The low inclusion rates across the SLRs indicate challenges in optimizing search filter performance. The variation in search strategies and the high yield of records on AI in HEOR call for the need for more research to create, validate, and standardize search filters in this field to optimize efficiency, and keep up to date, with ongoing stream of new evidence.

Conflicts of Interest

The authors declare no competing interests that could have influenced the objectivity or outcome of this research.

Funding Source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgment

None

Institutional Review Board (IRB)

None

Large Language Model

None

Authors Contribution

MR conceptualised the study. MR and OI conducted the literature search, study screening, selection, and data extraction, and drafted the manuscript. MR drafted the initial manuscript, and OI reviewed and revised the manuscript. All authors critically reviewed the manuscript for important intellectual content and approved the final manuscript as submitted. MR is the guarantor.

Data Availability

No new datasets were generated or analyzed in this umbrella review. All data supporting the findings are derived from previously published sources and are presented within the article and its supplementary materials.

References

1. van Nooten F, Holmstrom S, Green J, Wiklund I, Odeyemi IA, Wilcox TK. Health economics and outcomes research within drug development: challenges and opportunities for reimbursement and market access within biopharma research. *Drug Discov Today*. 2012;17(11-12):615-22. [PMID: 22366662, <https://doi.org/10.1016/j.drudis.2012.01.021>].

2. Reason T, Rawlinson W, Langham J, Gimblett A, Malcolm B, Klijn S. Artificial Intelligence to Automate Health Economic Modelling: A Case Study to Evaluate the Potential Application of Large Language Models. *Pharmacoecon Open*. 2024;8(2):191-203. [PMID: 38340276, PMCID: PMC10884386, <https://doi.org/10.1007/s41669-024-00477-8>].
3. NICE. Use of AI in evidence generation: NICE position statement; 2024. Available from: <https://www.nice.org.uk/about/what-we-do/our-research-work/use-of-ai-in-evidence-generation-nice-position-statement>. Accessed on: 09 Nov 2025.
4. Mason J, Morrison A, Visintini S. An overview of clinical applications of artificial intelligence. Ottawa: CADTH. 2018.
5. Health NioEi, Services S. Development and evaluation of a software tool based on GPT-4 to assist in sorting documents in the context of literature reviews: Literature, a proof of concept; 2024. Available from: https://www.inesss.qc.ca/fileadmin/doc/INESSS/Demarche/INESSS_Outil_tri_GPT_4_VF.pdf. Accessed on: 09 Nov 2025.
6. Escobar Liquitay CM, Garegnani L, Garrote V, Sola I, Franco JV. Search strategies (filters) to identify systematic reviews in MEDLINE and Embase. *Cochrane Database Syst Rev*. 2023;9(9):MR000054. [PMID: 37681507, PMCID: PMC10485899, <https://doi.org/10.1002/14651858.MR000054.pub2>].
7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. [PMID: 33782057, PMCID: PMC8005924, <https://doi.org/10.1136/bmj.n71>].
8. Pollock M, Fernandes RM, Becker LA, Pieper D, Hartling L. Chapter V: overviews of reviews. *Cochrane handbook for systematic reviews of interventions version*. 2020;6.
9. OpenAI. GPT-4 Technical Report 2023; 2023. Available online: <https://openai.com/research/gpt-4>. Accessed on: 09 Nov 2025.
10. de la Torre-López J, Ramírez A, Romero JR. Artificial intelligence to automate the systematic review of scientific literature. *Computing*. 2023;105(10):2171-94. [<https://doi.org/10.1007/s00607-023-01181-x>].
11. Santos AOD, da Silva ES, Couto LM, Reis GVL, Belo VS. The use of artificial intelligence for automating or semi-automating biomedical literature analyses: A scoping review. *J Biomed Inform*. 2023;142:104389. [PMID: 37187321, <https://doi.org/10.1016/j.jbi.2023.104389>].
12. Schmidt L, Mohamed S, Meader N, Bacardit J, Craig D. Automated data analysis of unstructured grey literature in health research: A mapping review. *Res Synth Methods*. 2024;15(2):178-97. [PMID: 38115736, <https://doi.org/10.1002/jrsm.1692>].
13. Schmidt L, Finnerty Mutlu AN, Elmore R, Olorisade BK, Thomas J, Higgins JPT. Data extraction methods for systematic review (semi)automation: Update of a living systematic review. *F1000Res*. 2021;10:401. [PMID: 34408850, PMCID: PMC8361807, <https://doi.org/10.12688/f1000research.51117.3>].
14. Doneva SE, Qin S, Sick B, Ellendorff T, Goldman JP, Schneider G, et al. Large language models to process, analyze, and synthesize biomedical texts: a scoping review. *Discover Artificial Intelligence*. 2024;4(1). [<https://doi.org/10.1007/s44163-024-00197-2>].
15. Toth B, Berek L, Gulacsi L, Pentek M, Zrubka Z. Automation of systematic reviews of biomedical literature: a scoping review of studies indexed in PubMed. *Syst Rev*. 2024;13(1):174. [PMID: 38978132, PMCID: PMC11229257, <https://doi.org/10.1186/s13643-024-02592-3>].
16. Yao X, Kumar MV, Su E, Flores Miranda A, Saha A, Sussman J. Evaluating the efficacy of artificial intelligence tools for the automation of systematic reviews in cancer research: A systematic review. *Cancer Epidemiol*. 2024;88:102511. [PMID: 38071872, <https://doi.org/10.1016/j.canep.2023.102511>].
17. J G, C L, P M, S R, I B, Woods L e. ISSG Search Filter Resource; 2006. York (UK): The InterTASC Information Specialists' Sub-Group; updated 17 September 2025. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home>. Accessed on: 09 Nov 2025.
18. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-6. [PMID: 27005575, <https://doi.org/10.1016/j.jclinepi.2016.01.021>].



Original Article

War Injuries of External Male Genitalia: A Prospective Single-Center Study from the General Military Hospital, Sana'a, Yemen (2022–2024)

Wael Alsaqaf¹, Khaled Alkohlany^{1,2}, Majdi Alshami¹, Mahmood Shaif¹, Abdullah Al-Sakkaf^{1,3}, Ahmed Farei¹, Aziz Shamsan¹, Nasser Albaddai^{1,*}

1-General Military Hospital, Sana'a, Yemen

2-Sana'a University, Sana'a, Yemen

3-21 September University of Medical and Applied Sciences, Sana'a, Yemen

ARTICLE INFO

Article history:

Received 23 Dec. 2025

Received in revised form 5 Jan. 2026

Accepted 19 Jan. 2026

Published 21 Jan. 2026

Keywords:

War injuries

External genital trauma

Penile injury

Testicular injury

Surgical management

ABSTRACT

Background: War-related external male genital injuries are uncommon but clinically significant, often coexisting with severe pelvic, vascular, and musculoskeletal trauma and causing potential long-term functional, cosmetic, and psychological effects.

Methods: This prospective, single-center study at the General Military Hospital, Sana'a, Yemen, included all consecutive male patients presenting with war-related external genital injuries from April 2022 to March 2024. Minor isolated injuries were managed in the emergency department, while severe or associated injuries required admission. Data collected included injury mechanism, management, complications, and short-term outcomes. Primary outcomes were injury patterns, management strategies, and one-month clinical outcomes. Secondary outcomes included complications, erectile function, cosmetic acceptability, and semen analysis at one month.

Results: Fifty-two patients (mean age 23.3 ± 4.9 years) were included. Penetrating trauma occurred in 44 (84.6%) and blunt trauma in 8 (15.4%). Penile injuries were seen in 16 (30.8%) and testicular injuries in 46 (88.4%). Isolated genital injuries occurred in 20 (38.4%), with 32 (61.5%) having associated injuries. All patients underwent surgery for penetrating wounds, large hematomas, or hematoceles. At one-month follow-up, early complications occurred in 12/52 patients, including wound infections (15.4%) and urinary tract infections (7.7%). Of 42 patients completing investigations, semen analysis was normal in 71.4%, and 28.6% had abnormal semen parameters. Late complications included testicular atrophy (33.3%), erectile dysfunction (4.8%), and post-operative penile curvature (4.8%).

Conclusion: Prompt assessment and timely surgical management are essential. While most patients had satisfactory short-term outcomes, a notable proportion experienced complications, highlighting the need for structured protocols and longer follow-up.

1. Introduction

Genital injuries are clinically significant due to their frequent association with major pelvic and vascular trauma from both blunt and penetrating mechanisms, as well as the potential for chronic impairment following penile and scrotal injuries [1]. While injuries to the external genitalia are rarely life-threatening on their own, they are often accompanied by other potentially fatal injuries to the abdomen, pelvis, or groin [2]. Traumatic injuries to the genitourinary (GU) tract occur in 2.2–10.3% of hospitalized trauma patients, with external genital injuries accounting for one-third to two-thirds of these cases [3]. Approximately 70% of GU injuries in combat settings involve the external genitalia, representing up to 5% of all military-related trauma [4, 5, 6]. GU trauma occurs across all age

groups but is most common in males aged 15–40 years, with 5% of cases occurring in children under ten years of age [7].

The testes, scrotum, and penis are vulnerable to both blunt and penetrating trauma. Testicular injuries may range from contusions to ruptures, while scrotal trauma often involves deep lacerations, burns, or avulsions. Penile injuries include fractures (rupture of the corpus cavernosum), amputations, strangulation, or bite-related injuries, as well as missile injuries [8].

Because trauma predominantly affects young individuals, injuries to the external genitalia can profoundly impact health-related quality of life, compromising urinary function, sexual performance, fertility potential, cosmetic appearance, and psychological well-being [9].

In Yemen, the ongoing conflict over the past decade has led to a notable increase in war-related male genital injuries, primarily from gunshots, explosions, and blast injuries. However, there is a scarcity of data on the epidemiology, management, and short-term outcomes of these injuries in low-resource, conflict settings. Aim: This study aimed to evaluate the mechanisms, types, and anatomical distribution of war-related external genital injuries, associated injuries, management strategies, and short-term clinical outcomes,

*Corresponding author: Nasser Albaddai, General Military Hospital, Sana'a, Yemen. Email: wmaaa2300@gmail.com

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Alsaqaf W, Alkohlany K, Alshami M, et al. War Injuries of External Male Genitalia: A Prospective Single-Center Study from the General Military Hospital, Sana'a, Yemen (2022–2024). ASIDE Health Sci. 2026;2(2):18-23. doi:10.71079/ASIDE.HS.012126461

including complications, sexual and reproductive function, and cosmetic results.

2. Patients and Methods

This prospective, single-center study was conducted at the General Military Hospital, Sana'a, Yemen, from April 2022 to March 2024. The study included all consecutive male patients presenting to the emergency department with war-related external genital injuries.

Inclusion criteria were males of any age with war-related external genital trauma. Exclusion criteria included non-war-related injuries, previous genital surgery, or incomplete clinical data.

Data collected included patient demographics, injury mechanism (penetrating vs blunt), anatomical site, associated injuries, time from trauma to hospital presentation and surgery, management approach, intraoperative findings, hospital stay, and follow-up outcomes. Patients were examined systematically, including the penis, scrotum, and perineum, documenting penetrating injuries, scrotal swelling, or hematomas.

2.1. Management

Mild isolated injuries were treated conservatively in the emergency department with local wound care. Surgical intervention was performed for penetrating injuries, large hematomas, or urethral involvement. Urethral injuries were managed with urethral catheterization or suprapubic cystostomy following preoperative retrograde urethrogram. Scrotal ultrasound and Doppler were selectively performed to evaluate testicular integrity and perfusion. Standard laboratory investigations included CBC, kidney and liver function tests, urinalysis, and urine culture.

2.2. Follow-up

One month after injury, patients underwent semen analysis to assess post-traumatic reproductive function. Evaluation followed standard laboratory protocols and included assessment of semen volume, sperm concentration, motility, and morphology. This timing allowed stabilization of acute post-injury changes and provided an initial assessment of fertility-related outcomes following management of external genital trauma. Outcomes were classified as "excellent" (good wound healing, satisfactory cosmetic appearance, preserved sexual function) or "unfavorable" (significant complications such as testicular atrophy, abnormal semen parameters, erectile dysfunction, or penile deviation). All patient information, including injury details, course of treatment, and follow-up findings, was recorded on a specifically designed data collection sheet.

2.3. Statistical analysis

Only descriptive statistics were used, with percentages and means calculated for all variables. This study represents a consecutive series of eligible patients; no formal sample size calculation was performed.

3. Results

3.1. Epidemiology

During the study period, 52 patients with external genital injuries were admitted to the urology department (**Table 1**). All patients were male, with a mean age of 23.3 ± 4.9 years (range: 10–32 years). Most patients were aged 20–29 years (38/52, 73.1%). Thirty patients (57.7%) were single, and 22 (42.3%) were married.

Table 1: Demographic characteristics and Mechanism of Injury

Category	Subcategory	Frequency	Percentage%
1. Age category	<20 years	10	19.2
	20-29 years	38	73.1
	30-39 years	4	7.7
2. Marital status	Single	30	57.7
	Married	22	42.3
Mechanism of Injury	Blunt: RTA, Fall astride	8	15.4
	Penetrating:	44	84.6
	Gunshot, Stabbed,	20	38.4
	Shells (explosive devices)	24	46.2

RTA, Road Traffic Accident; n, Number of participants.

Penetrating trauma was the predominant mechanism, occurring in 44 patients (84.6%). Injuries caused by explosive devices accounted for 24 cases (46.2%), while gunshot or stabbing injuries occurred in 20 cases (38.4%). Blunt trauma was observed in 8 patients (15.4%) (**Table 1**).

3.2. Types and Distribution of Genital Injuries

Scrotal injuries were identified in 44 patients (84.6%), penile injuries in 16 patients (30.8%), and testicular injuries in 46 patients (88.5%). Testicular injuries involved the right testis in 12 patients (23.1%), the left testis in 20 (38.5%), and were bilateral in 14 patients (26.9%) (**Table 2**).

3.3. AAST Injury Grading

According to the AAST classification, most scrotal injuries were Grade II (32/44, 72.7%). Testicular injuries were most commonly Grade I (16/46, 34.8%) and Grade V (14/46, 30.4%). Penile injuries were predominantly Grade II (12/16, 75.0%) (**Table 2**).

3.4. Associated Injuries

Concomitant injuries were present in 32 patients (61.5%). Femoral–pelvic fractures were the most frequent associated injury (24/52, 46.2%), followed by vascular injuries (12/52, 23.1%), abdominal–pelvic organ injuries (4/52, 7.7%), and urethral injuries (2/52, 3.8%) (**Figure 1**).

3.5. Surgical Repair Indications

Six patients (11.5%) had extensive hematomas, six patients (11.5%) had hematocele, and the majority of patients (n=40, 76.9%) had indications for intervention because of penetrating injuries.

3.6. Surgical Management and Discharge Outcomes

All patients underwent surgical management. Fourteen patients (26.9%) required scrotal exploration and/or orchiectomy due to testicular destruction or avulsion. Blood transfusion was required in 14 patients (26.9%) because of associated vascular or orthopedic injuries. Thirty patients (57.7%) were discharged home, while 22 (42.3%) were transferred to other departments for continued care (**Table 3**).

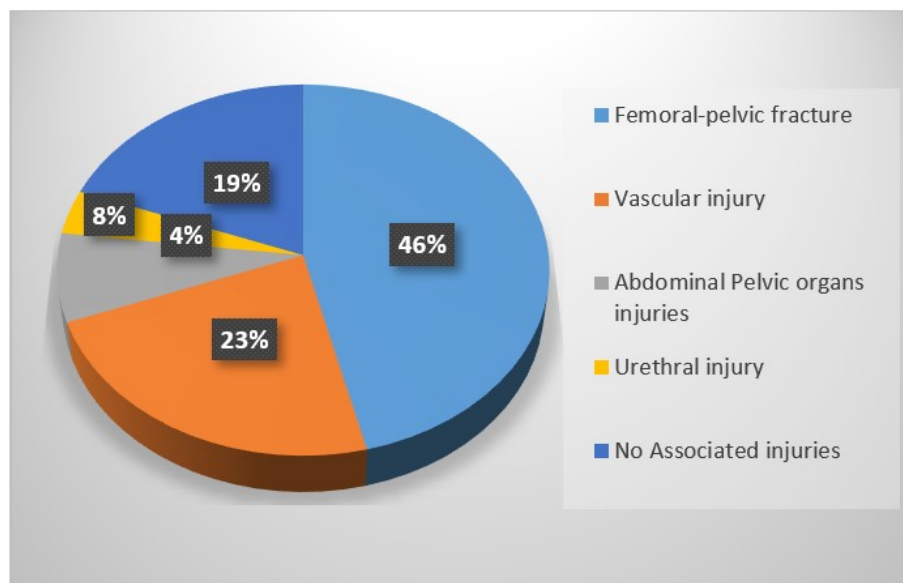


Figure 1: Distribution of associated injuries distribution.

Table 2: Types, Distribution, and Genital Lesions according to the AAST Injury Grading System

External Genitalia Examination	Type of Injury	Frequency	Percentage%
Penis	Laceration	16	30.8
Scrotum	Laceration	18	34.6
	Swelling	2	3.8
	Skin loss	16	30.8
Testes	Testicular dislocation	6	11.5
	Testicular rupture	12	23.1
	Hematocoele	10	19.2
Side of Testes injury	Right testis	12	23.1
	Left testis	20	38.5
	Bilateral	14	26.9
AAST Grade			
	Penis (n = 16)	Scrotum (n = 44)	Testes (n = 46)
Grade I	2	6	16
Grade II	12	32	8
Grade III	2	0	4
Grade IV	0	4	4
Grade V	0	2	14

AAST, American Association for the Surgery of Trauma; n, Number of participants.

3.7. Complications and Follow-up

During the study, early complications were observed in 12 of the 52 patients, including wound infections in 8 patients (15.4%) and urinary tract infections in 4 patients (7.7%). At the one-month

follow-up, 42 patients completed post-injury investigations. Urine analysis was normal in all 42 patients (100%). Semen analysis was normal in 30 patients (71.4%), while 12 patients (28.6%) demonstrated abnormal semen parameters at one month, including oligospermia and/or motility disorders in 8 patients (19.0%) and azoospermia in 4 patients (9.5%). Semen culture was negative in 40 patients (95.2%) and positive in 2 patients (4.8%). Late complications included testicular atrophy in 14 patients (33.3%), erectile dysfunction in 2 patients (4.8%), and post-operative penile curvature in 2 patients (4.8%). Abnormal semen parameters at one month were observed in 12 patients (28.6%). At one-month follow-up, 28 patients (66.7%) achieved a satisfactory outcome, characterized by good wound healing, preserved sexual function, and acceptable cosmetic appearance, while 14 patients (33.3%) experienced an unfavorable outcome, including complications such as testicular atrophy, abnormal semen parameters, erectile dysfunction, or post-operative penile deformity (Table 3).

4. Discussion

Young men are frequently affected by injuries to the male external genitalia. Several studies have documented similar findings: Lee et al., in a 10-year review, reported a mean age of 27.8 years [10], while Ahmed et al. observed a mean age of 28 ± 3.2 years [11]. In our study, the mean age was 23.3 ± 4.9 years, which aligns with these reports. Diagnosis was primarily based on clinical assessment, followed by emergency surgery. Scrotal ultrasound was performed in 14 patients (53.8%) and was not systematically applied. This approach aligns with Djè and Coulibaly [12, 13], who suggest emergency surgical exploration except in cases of severe trauma. This therapy would decrease the likelihood of problems and aftereffects. The treatment attitude to scrotal injuries is diverse and depends above all on the existence or absence of a hematocoele on clinical evaluation [14].

Mechanisms of injury were predominantly penetrating (84.6%), with blunt trauma representing 15.4%. RTA and gunshot injuries accounted for 68% and 16% of injuries, respectively, in the Ahmed et al. study [11], but penetrating injuries accounted for 79% of injuries in other reports [15]. This is comparable to the situation

Table 3: Type of Surgical Intervention, Management, Complications, and Follow-up Outcomes

Surgical category	Intervention details	Frequency	Percentage (%)
Penile injuries	Glans laceration, Buck's fascia laceration treated by Penile exploration and suturing	14	26.9
	Cutaneous avulsion/laceration involving the glans treated by Penile stump refashioning with primary suturing	2	3.8
Scrotal injuries	Superficial or deep laceration/contusion treated by debridement and primary/secondary suturing	40	76.9
	Scrotal hematoma evacuation	4	7.7
Testicular injuries	Scrotal exploration and/or orchiectomy for testicular destruction or avulsion	14	26.9
	Hematocele evacuation for contusion/hematoma/subclinical laceration	24	46.2
	Scrotal exploration with testicular repair	8	15.4
Supportive management	Blood transfusion for associated vascular or orthopedic bleeding	14	26.9
Discharge and Associated Injuries	Discharge to another department	22	42.3
	Discharge home	30	57.7
Early Complications (n = 52)			
	Wound infection	8	15.4
	Urinary tract infection	4	7.7
Follow-up investigations at 1 Month (n = 42)			
Urine analysis	Normal	42	100
Semen analysis	Normal	30	71.4
	Abnormal semen parameters at one month	12	28.6
	Oligospermia/motility disorder	8	19.0
	Azoospermia	4	9.5
Semen culture	Negative	40	95.2
	Positive	2	4.8
Late complications			
	Erectile dysfunction	2	4.8
	Testicular atrophy	14	33.3
	Penile abnormality: post-operative curvature formation	2	4.8
	Abnormal semen parameters at one month	12	28.6
Outcome			
	Satisfactory outcome	28	66.7
	Unfavorable outcome	14	33.3

n, Number of patients.

in other nations where penetrating wounds are more common [16]. Gunshot wounds accounted for 38.4%, while explosive devices caused 46.2% of injuries. Penile injuries were noted in 30.8% of patients, lower than the 60% reported in large cohort studies by Bertini and Corriere [17], likely due to differences in sample size and study period.

All patients underwent prompt assessment and management, including hemodynamic stabilization and urethral evaluation when indicated. Penetrating penile injuries were treated with debridement and exploration, preserving tissue for potential reconstruction, with corpora repaired using absorbable sutures [18]. Scrotal

and testicular injuries were addressed immediately: hematomas were evacuated, necrotic tissue excised, and tunica albuginea closed, using tunica vaginalis flaps in selected cases (15.4% of testicular repairs; 46.2% of hematocele evacuations). Orchiectomy was required in 26.9% of patients with non-viable testes, mostly due to high-velocity gunshot injuries, consistent with previous reports [19, 20].

Penetrating scrotal wounds were generally managed with exploration; partial scrotal skin loss was repaired primarily. Large hematomas were drained promptly to prevent infection and ischemia,

whereas stable hematomas were managed conservatively with rest, ice, and elevation [21, 22].

Associated injuries occurred in 32 patients (61.5%), primarily vascular and musculoskeletal, requiring multidisciplinary management. In Ahmed et al.'s study, related injuries were observed in 41% of cases [11]. Mean hospital stay was 5.2 ± 5.1 days, comparable to Coulibaly [12] but shorter than reports from Burkina Faso [22].

At one-month follow-up, 28 patients (66.7%) achieved satisfactory cosmetic and functional outcomes, including good wound healing, acceptable genital appearance, and preserved sexual function. Exceptions included two patients requiring bilateral orchiectomy with subsequent testosterone therapy. Semen analysis, completed in 42 patients, showed normal results in 71.4%, oligoasthenozoospermia in 19.0%, and azoospermia in 9.5%. Early infection was uncommon and managed with local care. Cosmetic and functional outcomes of therapy of penetrating genital damage are usually favorable if adequate, on-time treatment and repair are undertaken [23]. In testicular injuries, sperm production and sustained hormonal function have been observed in men with adequately bilateral repaired testes after penetrating injuries [24].

4.1. Limitations

This study is limited by its single-center military setting, small sample size, and potential selection bias (only patients reaching this hospital were included). Follow-up was incomplete, and validated instruments for sexual function or quality of life (e.g., IIEF) were not used.

5. Conclusion

External male genital injuries during war are caused by a variety of mechanisms, ranging from minor penile lacerations to complete avulsion of the testes and scrotum. Timely evaluation, meticulous wound irrigation with saline and mild antiseptics, and prompt surgical intervention remain critical for minimizing infection and preserving function. In our cohort, while a majority of patients achieved satisfactory cosmetic and functional outcomes at one month, a substantial proportion experienced complications, including testicular atrophy, abnormal semen parameters, penile deformity, and erectile dysfunction. These findings highlight the need for standardized management protocols for war-related genital trauma and for future multi-center studies with longer follow-up to better assess fertility, sexual function, and quality of life outcomes.

Conflicts of Interest

The authors declare no conflict of interest.

Funding Source

This research received no external funding.

Acknowledgment

The authors would like to thank the medical and nursing staff of the General Military Hospital for their support during data collection.

Informed consent

This study was approved by the Ethics Committee of the General Military Hospital, Sana'a, Yemen (Approval No. GMH-URO-2024-7). Written informed consent was obtained from all participants.

Large Language Model

Artificial intelligence or large language model tools were used for language editing and formatting only. No AI tools were used for data analysis, interpretation, or scientific content generation.

Authors Contribution

WAKA, MANA, MSA, A were responsible for conceptualization. WA, AAS, AF, AS, NA conducted data collection. MS, AAS, AF, AS, NA performed data analysis. WAKA, MAMA, MS, AAS, AF, AS, NA carried out manuscript writing as well as manuscript review and editing.

Data Availability

None.

References

- Shewakramani S, Reed KC. Genitourinary trauma. *Emerg Med Clin North Am.* 2011;29(3):501-18. [PMID: 21782071, <https://doi.org/10.1016/j.emc.2011.04.009>].
- Waxman S, Beckley A, Morey A, Soderdahl D. Penetrating trauma to the external genitalia in Operation Iraqi Freedom. *Int J Impot Res.* 2009;21(2):145-8. [PMID: 19020522, <https://doi.org/10.1038/ijir.2008.59>].
- Al-Mamari SA. Epidemiology of Scrotal and Testicular Trauma. *Urogenital Trauma: A Practical Guide.* 2023:371-3. [https://doi.org/10.1007/978-981-99-6171-9_43].
- Hudak SJ, Hakim S. Operative management of wartime genitourinary injuries at Balad Air Force Theater Hospital, 2005 to 2008. *J Urol.* 2009;182(1):180-3. [PMID: 19450817, <https://doi.org/10.1016/j.juro.2009.02.150>].
- McCormick CS, Dumais MG, Johnsen NV, Voelzke BB, Hagedorn JC. Male genital trauma at a level 1 trauma center. *World J Urol.* 2020;38(12):3283-9. [PMID: 32077992, <https://doi.org/10.1007/s00345-020-03115-0>].
- Williams M, Jezior J. Management of combat-related urological trauma in the modern era. *Nat Rev Urol.* 2013;10(9):504-12. [PMID: 23877722, <https://doi.org/10.1038/nrurol.2013.148>].
- Gioffre-Florio M, Murabito LM, Visalli C, Pergolizzi FP, Fama F. Trauma in elderly patients: a study of prevalence, comorbidities and gender differences. *G Chir.* 2018;39(1):35-40. [PMID: 29549679, PMCID: PMC5902142, <https://doi.org/10.11138/gchir/2018.39.1.035>].
- Hadjipavlou M, Sharma D. Traumatic Scrotal and Testicular Injuries. *A Clinical Guide to Urologic Emergencies.* 2021:202-15. [<https://doi.org/10.1002/9781119021506.ch15>].
- Harvey-Kelly KF, Kanakaris NK, Obakponovwe O, West RM, Giannoudis PV. Quality of life and sexual function after traumatic pelvic fracture. *J Orthop Trauma.* 2014;28(1):28-35. [PMID: 23481925, <https://doi.org/10.1097/BOT.0b013e31828fc063>].
- Lee SH, Bak CW, Choi MH, Lee HS, Lee MS, Yoon SJ. Trauma to male genital organs: a 10-year review of 156 patients, including 118 treated by surgery. *BJU Int.* 2008;101(2):211-5. [PMID: 17922859, <https://doi.org/10.1111/j.1464-410X.2007.07265.x>].
- Ahmed A, Mbibu NH. Aetiology and management of injuries to male external genitalia in Nigeria. *Injury.* 2008;39(1):128-33. [PMID: 17572420, <https://doi.org/10.1016/j.injury.2007.02.039>].
- Coulibaly M, Issa A, Kossogué A, Ouattara Z. Traumatisme des bourses: aspects cliniques et thérapeutiques au service d'urologie du CHU Gabriel Touré. *Mali Médical.* 2017;32:13-6.
- Dje K, Yao B, Sanou B, Kokoua A, Gnananzan KG. Les complications uro-génitales des traumatismes du bassin a propos de 41 cas. *SOPEMI;* 2007. [PMID: 18679344, PMCID: PMCID].
- Fouedjio JH, Fouelifack FY, Nnang NA, Mbu RE. [Epidemiological, clinical and therapeutic aspects of supposed benign ovarian tumors at the Central Hospital of Yaounde]. *Pan Afr*

- Med J. 2016;25:207. [PMID: 28270904, PMCID: PMC5326262, <https://doi.org/10.11604/pamj.2016.25.207.9939>].
15. Mohr AM, Pham AM, Lavery RF, Sifri Z, Bargman V, Livingston DH. Management of trauma to the male external genitalia: the usefulness of American Association for the Surgery of Trauma organ injury scales. *J Urol.* 2003;170(6 Pt 1):2311-5. [PMID: 14634403, <https://doi.org/10.1097/01.ju.0000089241.71369.faj>].
 16. Cline KJ, Mata JA, Venable DD, Eastham JA. Penetrating trauma to the male external genitalia. *J Trauma.* 1998;44(3):492-4. [PMID: 9529176, <https://doi.org/10.1097/00005373-199803000-00011>].
 17. Bertini J J E, Corriere J J N. The etiology and management of genital injuries. *J Trauma.* 1988;28(8):1278-81. [PMID: 3411650, <https://doi.org/10.1097/00005373-198808000-00025>].
 18. Kunkle DA, Lebed BD, Mydlo JH, Pontari MA. Evaluation and management of gunshot wounds of the penis: 20-year experience at an urban trauma center. *J Trauma.* 2008;64(4):1038-42. [PMID: 18404072, <https://doi.org/10.1097/TA.0b013e3180342036>].
 19. Gomez RG, Castanheira AC, McAninch JW. Gunshot wounds to the male external genitalia. *J Urol.* 1993;150(4):1147-9. [PMID: 8371375, [https://doi.org/10.1016/s0022-5347\(17\)35710-5](https://doi.org/10.1016/s0022-5347(17)35710-5)].
 20. Morey AF, Metro MJ, Carney KJ, Miller KS, McAninch JW. Consensus on genitourinary trauma: external genitalia. *BJU Int.* 2004;94(4):507-15. [PMID: 15329102, <https://doi.org/10.1111/j.1464-410X.2004.04993.x>].
 21. Cass AS. Testicular trauma. *J Urol.* 1983;129(2):299-300. [PMID: 6834494, [https://doi.org/10.1016/s0022-5347\(17\)52062-5](https://doi.org/10.1016/s0022-5347(17)52062-5)].
 22. Kambou T ZC Ouattara A. TRAUMATISMES UROGENITAUX : PROFIL EPIDEMIOLOGIQUE ET ASPECTS LESIONNELS AU CENTRE HOSPITALIER UNIVERSITAIRE SOURO SANON DE BOBO DIOULASSO (BURKINA FASO); 2014. [PMCID: PMCID].
 23. Cavalcanti AG, Krambeck R, Araujo A, Manes CH, Favorito LA. Penile lesion from gunshot wound: a 43-case experience. *Int Braz J Urol.* 2006;32(1):56-60; discussion 60-3. [PMID: 16519829, <https://doi.org/10.1590/s1677-55382006000100009>].
 24. Brandes SB, Buckman RF, Chelsky MJ, Hanno PM. External genitalia gunshot wounds: a ten-year experience with fifty-six cases. *J Trauma.* 1995;39(2):266-71; discussion 271-2. [PMID: 7674395, <https://doi.org/10.1097/00005373-199508000-00013>].

Review Article

Cross-Cultural Application of the Beck Depression Inventory (BDI-I/II) for Assessing Adolescent Depression in Low- and Middle-Income Countries: A Critical Review of Evidence from the Last Quarter-Century

Anas Al-Qatefy^{1,*}, Abdelrahman Tolba², Zeyad Eldawy², Mahmoud Ghallab³

1-Faculty of Medicine, Mansoura National University, Mansoura, Egypt

2-Neurology Department, Faculty of Medicine, Alexandria University, Alexandria, Egypt

3-Psychiatry Department, Faculty of Medicine, Mansoura National University, Mansoura, Egypt

ARTICLE INFO

Article history:

Received 9 Dec. 2025

Received in revised form 28 Dec. 2025

Accepted 19 Jan. 2026

Published 22 Jan. 2026

Keywords:

Beck Depression Inventory (BDI)

Adolescent Depression

Cross-Cultural Validation

Low- and Middle-Income Countries (LMICs)

Measurement Invariance

Psychiatry

Psychology

ABSTRACT

Background: The Beck Depression Inventory (BDI-I/II) versions are widely used for identifying depression among adolescents in low- and middle-income countries (LMICs). Nevertheless, their global usage depends on the questionable assumption that symptoms are expressed the same way in different cultures.

Methods: This comprehensive review used narrative synthesis to explore the psychometric, conceptual, and ethical aspects of the BDI studies done on adolescents in LMICs. We purposively selected 8 case exemplars from the literature based on a pre-specified framework to illustrate key cross-cultural and methodological issues. A custom rubric was used to weight studies by methodological quality. Invariance evidence was defined as explicit reporting of multi-group confirmatory factor analysis (MG-CFA) for configural, metric, or scalar invariance.

Results: Although the BDI showed high internal consistency across the 8 exemplar studies (Cronbach's : 0.79–0.96), it revealed significant psychometric variation across cultures. Somatic items showed substantial overlap with medical comorbidities (e.g., HIV, malnutrition) in vulnerable populations, confounding diagnostic specificity. Factor structures were inconsistent, and rigorous measurement invariance testing was largely absent.

Discussion: The widespread absence of scalar invariance suggests that current cross-cultural score comparisons may reflect measurement artifacts rather than true differences in depression severity. Beyond psychometrics, BDI presents ethical challenges regarding developmentally inappropriate items, particularly concerning sexuality and suicidality, that limit its validity in conservative or resource-limited contexts.

Conclusion: Investigations into this issue should focus more on the use of invariance-testing frameworks, the implementation of a mixed-methods strategy, and the recognition of the need for sensitive translation and culturally appropriate assessment techniques.

1. Introduction

Depression among adolescents is a major worldwide public health problem. Unfortunately, the majority of cases occur in low- and middle-income countries (LMICs) that have very limited mental health resources [1, 2]. The Beck Depression Inventory (BDI-I and BDI-II), grounded in Beck's cognitive theory of depression, has become commonly used for depression symptom measurement worldwide [3]. Nonetheless, the extensive use of BDI in LMICs often follows what has been termed "methodological imperialism"—the practice of translating and validating Western-developed

instruments without adequate consideration of their cultural and epistemological foundations [4, 5].

The main premise of cross-cultural psychological assessment, i.e. one tool can measure the same constructs in different populations, requires empirical proof of measurement invariance [6]. This challenge intensifies when assessing adolescents, whose emotional awareness and capacity for cognitive abstraction follow developmental trajectories that may interact with cultural context [7]. Furthermore, cultural syndromes and idioms of distress, particularly the somatization of psychological distress, which is strongly present in numerous non-Western cultural groups, challenge the BDI's cognitive-affective-somatic framework and point to the importance of cultural-contextual factors in adolescent mental health assessment [8].

Recent epidemiological data highlight the need for such an investigation. A systematic review of mental health problems among adolescents in sub-Saharan Africa revealed that 26.9% (IQR 20.1–31.1) of general population adolescents are showing depressive symptoms, and the percentage goes even higher in at-risk groups [9]. Correspondingly, approximately 24.4% (95% CI 19.2–30.5)

*Corresponding author: Anas Al-Qatefy, Faculty of Medicine, Mansoura National University, Mansoura, Egypt. Email: aanwar@std.mansnu.edu.eg

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Al-Qatefy A, Tolba A, Eldawy Z, Ghallab M. Cross-Cultural Application of the Beck Depression Inventory (BDI-I/II) for Assessing Adolescent Depression in Low- and Middle-Income Countries: A Critical Review of Evidence from the Last Quarter-Century. ASIDE Health Sci. 2026;2(2):24-32, doi:10.71079/ASIDE.HS.012226419

of university students in LMICs were identified with depressive symptoms [10]. Almost 90% of adolescents globally reside in LMICs, meaning that even moderate prevalence translates into a massive public health burden [11].

Multiple attempts to validate the BDI have been made. Still, no comprehensive synthesis has yet psychometrically and conceptually examined the BDI (both versions, I and II) among adolescents in LMICs over the last 25 years. This critical narrative review builds on prior validation studies of the BDI with adolescent populations in LMICs over the last 25 years. We go beyond simply listing psychometric properties to re-evaluate the evidence for the BDI's cross-cultural construct validity, examine methodological rigor, interpret challenges in understanding conceptual equivalence, and reflect on the ethical and scientific aspects of global adolescent mental health assessment solutions that stem from this literature.

2. Methods

2.1. Review Design

The study employs a critical narrative review method, selected because it is appropriate for synthesizing heterogeneous literature and for offering a conceptual critique of intricate phenomena, according to Greenhalgh et al. [12]. Our synthesis includes studies using both the original BDI (BDI-I) and the revised BDI-II. While both versions assess depressive symptomatology, they differ substantively: BDI-I reflects symptoms over the past week and includes items on weight loss, somatic preoccupation, and work difficulty, whereas BDI-II assesses symptoms over the past two weeks and replaces several items (e.g., adding "changes in sleeping pattern" and "concentration difficulty") while removing others. These structural differences may independently influence factor structure, reliability, and optimal cut-offs, regardless of cultural context.

2.2. Search Strategy and Selection Criteria

A structured search was conducted in PubMed, PsycINFO, and Scopus covering January 2000 to January 2025. The search strategy, as elaborated in Supplementary File 1, employed: ("Beck Depression Inventory" OR "BDI") AND (adolescent* OR youth OR teenage*) AND (valid* OR reliab* OR psychometric*) AND (LMIC OR "low and middle income" OR "developing countr*" OR "global health"). Reference lists of assessed papers were manually screened to identify additional relevant studies. Inclusion criteria covered empirical studies that demonstrated the validation of the BDI-I or BDI-II in adolescent populations (ages 10–19) living in low- and middle-income countries (LMICs) and that reported at least one psychometric property (e.g., reliability, validity, factor structure). The status of lower- and middle-income countries was defined according to the World Bank Group's 2023 country income classification. Exclusion criteria were studies that were completely adult-only, non-English, or non-translated papers without an English abstract, and theoretical discussions.

2.3. Synthesis

Critical narrative reviews are well-suited to interrogating conceptual assumptions and synthesizing heterogeneous evidence through purposive selection of illustrative cases. A narrative synthesis was performed through thematic analysis organized around cultural, methodological, and developmental challenges. From the 54 full-text studies assessed for relevance, we purposively selected 8 case exemplars based on a pre-specified framework emphasizing: (1) geographic diversity across LMIC regions, (2) variation in cultural adaptation procedures, and (3) depth of psychometric reporting

(particularly factor structure or measurement invariance). These exemplars were chosen to maximize insight into cross-cultural validity challenges rather than statistical generalizability. The remaining 46 studies informed contextual understanding but were excluded from detailed synthesis due to insufficient psychometric reporting, methodological limitations, or sample ineligibility, as elaborated in Supplementary File 2. Additionally, we applied a custom quality rubric assessing four criteria: (1) sample representativeness, (2) use of cognitive interviewing or local input in adaptation (beyond back-translation), (3) reliability reporting (e.g., Cronbach's α), and (4) critical discussion of validity (e.g., item equivalence or cut-off appropriateness). Each criterion scored 1 (met) or 0 (not met); studies scoring 2 were selected as exemplars. Two reviewers appraised independently; disagreements were resolved by consensus. Invariance evidence was defined as explicit reporting of multi-group confirmatory factor analysis (MG-CFA) for configural, metric, or scalar invariance; claims without statistical testing were not counted. Following Cohen's general effect size guidelines (often applied in psychological assessment), convergent validity coefficients 0.50 are interpreted as strong, 0.30–0.49 as moderate, and < 0.30 as weak [13].

3. Review

3.1. Overview of Included Studies

This narrative review summarized the results of eight representative validation studies of the Beck Depression Inventory (BDI-I and BDI-II) conducted in adolescent populations in low- and middle-income countries (LMICs), as shown in (Table 1). Our exemplar studies span three distinct population contexts: (1) general adolescent populations (school-based samples in Nigeria, India, Brazil, Chile, Iran, and Turkey), (2) clinical or medically vulnerable groups (HIV-affected youth in Malawi [22]), and (3) institutionally at-risk cohorts (orphaned adolescents in Nepal [18]). These settings differ markedly in baseline distress levels, symptom expression, and comorbidity profiles. The studies included 181–1,412 adolescent participants, and in most cases, the adolescents ranged in age from 13 to 18 years. Various forms of BDI were also applied, including the original BDI, BDI-II in several translated versions, and culturally adapted versions. Translation-back-translation was a common protocol across all studies; however, the extent and quality of the cultural adaptation process varied significantly from one study to another.

3.1.1. Misleading Reliability Paradox

Empirical evidence from the included exemplar studies indicates that the Cronbach's α coefficients were high across all eight studies, ranging from 0.79 to 0.96. In India, the back-translated Tamil version of the BDI showed excellent internal consistency ($\alpha = 0.96$) and good 4-week test-retest reliability ($r = 0.82$). It also had high convergent validity with the Children's Depression Rating Scale-Revised ($r = 0.72$) [15]. In orphanage environments, the Nepali BDI-II translation had $\alpha = 0.79$ [18], while the Persian BDI-II in Iranian high school students showed $\alpha = 0.83$ and a test-retest reliability of $r = 0.55$ [19]. Several studies on Brazilian adolescents reported good internal consistency across samples, and Ferreira-Maia and colleagues, in particular, reported elevated alpha values in their large São Paulo sample of 1,184 adolescents aged 13–18 years ($\alpha = 0.88$) [21].

These findings raise critical methodological implications: the consistent pattern of very high internal consistency across the board is quite impressive but may also be somewhat deceptive. The high alpha coefficients indicate that the items correlate strongly within each culture, i.e., respondents in each culture respond to the items in a statistically coherent way. Nevertheless, this single psychometric

Table 1: Cross-cultural validation studies of the Beck Depression Inventory (BDI) in adolescent populations

Author (Year)	Country / Region	BDI Version and Cultural Adaptation	Sample (n, age)	Factor Structure / Validity Evidence
(Adewuya et al., 2007) [14]	Nigeria (South-West)	BDI-I (Translation and back-translation)	n = 1,095; school adolescents aged 13–18 years	Criterion validation against K-SADS/E; ROC analyses confirmed strong diagnostic accuracy.
(Basker et al., 2007) [15]	India (Vellore, Tamil Nadu)	BDI-I (Tamil back-translation)	n = 181; secondary school adolescents	EFA and CFA replicated a 2-factor solution (30.5% variance explained); distinct screening and diagnostic cut-offs were proposed (5 vs 22).
(Araya et al., 2013) [16]	Chile (Santiago)	BDI-II (Spanish adaptation)	n = 592; mean age 15.5 years	EFA/CFA supported a unidimensional structure; criterion validity was established against MINI-KID; sex-specific cut-offs were identified (higher for girls).
(Kim et al., 2014) [17]	Malawi	BDI-II (HIV-affected cohort)	n = 562; adolescents aged 12–18 years in pediatric HIV program	Criterion validity confirmed; authors noted potential inflation of somatic symptoms due to HIV comorbidity. AUC = 0.82 (95% CI: 0.78–0.89) against CDRS-R as criterion.
(Bhatt et al., 2020) [18]	Nepal (Childcare homes)	BDI-II (Nepali version with pilot testing)	n = 602; adolescents aged 13–17 years	Internal consistency and convergent validity are reported; the authors caution against interpreting the results in the context of the study.
(Toosi et al., 2017) [19]	Iran (Shiraz)	BDI-II (Persian translation with expert review)	n = 1,150 (main sample); reliability subsample n = 128; high-school students	EFA (varimax) yielded a 4-factor solution, explaining 43.7% of the variance; concurrent validity with DASS-21 and BAI was confirmed.
(Jha et al., 2017) [20]	India (Bihar and multi-site school surveys)	BDI-II (Hindi translations)	In 1,412, school adolescents aged 14–18 years	Reported high prevalence estimates; methodological warnings against adult cut-off application to adolescents.
(Ferreira-Maia et al., 2018) [21]	Brazil (São Paulo)	BDI-II (Portuguese adaptation)	In 1,184, adolescents aged 13–18 years	EFA/CFA supported a dominant general factor; gender invariance analyses indicated minor differential item functioning.

BDI, Beck Depression Inventory; BDI-II, Beck Depression Inventory—Second Edition; , Cronbach's alpha; r, correlation coefficient; ROC, Receiver Operating Characteristic; Sn, Sensitivity; Sp, Specificity; K-SADS/E, Schedule for Affective Disorders and Schizophrenia for School-Age Children—Epidemiologic Version; CDRS-R, Children's Depression Rating Scale—Revised; MINI-KID, Mini International Neuropsychiatric Interview for Children and Adolescents; DASS-21, Depression Anxiety Stress Scales—21; BAI, Beck Anxiety Inventory; EFA, Exploratory Factor Analysis; CFA, Confirmatory Factor Analysis.

characteristic cannot establish the construct's equivalence across cultures, which is the most important issue [23]. Sijtsma (2009) argued that high internal consistency is a prerequisite but not sufficient for construct validity, especially in cross-cultural contexts, where high item correlations may be due to cultural factors rather than to shared cultural response biases, acquiescence, or socially desirable responding patterns [24, 25]. In other words, internal consistency supports within-group reliability but does not guarantee between-group comparability—a distinction often overlooked in cross-cultural psychometric research.

The Nigerian validation study conducted by Adewuya et al. offers valuable insights. The study demonstrated excellent screening performance, with a sensitivity of 0.91 and specificity of 0.97 at a cut-off of 18 on the Kiddie Schedule for Affective Disorders and Schizophrenia/Epidemiologic version (K-SADS/E) diagnostic interview. However, the authors strongly advised against the direct use of Western cut-offs, stressing the importance of local threshold validation and noting significant gender differences, with a higher prevalence in female adolescents. The finding that female adolescents are most affected underscores the critical methodological point:

acceptable psychometric properties within a single cultural context do not automatically justify cross-cultural score comparisons or the application of universal cut-off values [14].

These reliability measures were almost entirely based on school-based convenience sampling of children recruited. Thus, the alpha coefficients may have been inflated due to restriction of range and spectrum bias [26]. Large samples of schools (n = 1,412 in one Bihar sample) using Hindi translations of the BDI-II, as reported in the Indian school surveys by Jha and colleagues [27], showed high internal consistency but very high point prevalence estimates. Critics suggested that these prevalence figures were inflated because adult-derived Western cut-offs for adolescents were uncritically applied to populations without validation, thereby illustrating how psychometric properties that appear to be reliable can be used to conceal cross-cultural issues in instrument application.

Additionally, cross-cultural response styles—including extreme responding, midpoint avoidance, and socially desirable answering—may further inflate internal consistency and distort item means independently of depressive symptomatology. Yet, these patterns were rarely assessed in the included validations. The Chimera of

Conceptual Equivalence A major finding from the studies' review is that the research identified significant differences in the factor structures across different cultural settings, thereby endangering the conceptual equivalence assumption. To investigate this, Basker et al. conducted an Indian Tamil validation and used both exploratory and confirmatory factor analyses; the final two-factor solution they identified explained about 30.5% of the variance [15]. The study particularly distinguished between screening cut-offs (5) and diagnostic cut-offs (22), suggesting that the BDI could be effective in a clinical context where situations and severity levels change.

While in the Indian Tamil context the personality of BDI may differ, the Chilean study by Araya et al. found that exploratory and confirmatory analyses both supported a largely unidimensional solution among 592 adolescents (mean age 15.5 years) in Santiago [16]. This result could have been taken as evidence for the cross-cultural stability of the model if not for the study's disclosure of a critical complication: cut-offs for males and females differed substantially, with girls requiring higher thresholds than boys. To further support their point, the authors noted that using a single cut-off value derived from a Western sample can lead to errors when applied to other genders, and that a need for sex- and context-specific validation should be recognized instead. This sex-based differential functioning implies that the association between the observed and the latent variables of depression constructs may change for demographic subgroups even within a single cultural context.

The Persian BDI-II validation among Iranian high school students by Toosi et al. revealed yet another factor structure. The exploratory factor analysis with varimax rotation gave a four-factor solution accounting for approximately 43.7% of the variance. The authors concluded that the BDI-II could still be a viable instrument for measuring depression in Iranian teen populations; however, due to disparities at the item-level, changes were required in the cultural context, and conducting cognitive interviews was necessary, especially with younger adolescents, as they might understand the items in a different way than older ones [19].

The Brazilian study by Ferreira-Maia and colleagues indeed represents the most methodologically sophisticated examination of factor structure and measurement invariance among the reviewed studies. Their abstract and subsequent Brazilian adolescent research reported that exploratory and confirmatory factor analyses revealed a dominant general depression factor alongside reliable somatic and cognitive group factors. Critically, they tested measurement invariance across gender and documented gender differences in item functioning. This finding supports the presence of a strong general depression construct. Still, it demonstrates that the relationship between this latent construct and specific item responses varies by gender—a psychometric property that undermines simple score comparisons between male and female adolescents.

Interpretively, this heterogeneity challenges the assumption that the BDI measures a single, equivalent depression construct across cultures. These divergent factor structures—two-factor, unidimensional, four-factor, and hierarchical models—across cultural contexts are not merely technical statistical details. They reflect fundamental differences in how depressive symptoms cluster and co-occur in different populations. As Chen argued, “when factor structures differ across groups, the construct being measured may itself differ, rendering cross-group comparisons conceptually problematic” [28]. The BDI may be assessing somewhat different constellations of distress in Nigeria, India, Chile, Iran, and Brazil, even when items are competently translated.

Medical Comorbidity and the Somatic Symptom Confound Kim et al.'s paper on the Malawian HIV-affected adolescent population brings to light how medical comorbidity can weaken the discriminant validity of the BDI's somatic items. While the study reported acceptable internal consistency and concurrent validity against structured diagnostic scales (CDRS-R), the authors issued a critical caution: somatic items assessing fatigue, sleep disturbance, and appetite loss showed substantial overlap with HIV-related symptoms [22]. In contexts where adolescents suffer from chronic diseases, these items become powerless in terms of diagnostic specificity for depression.

The Nepali study done by Bhatt et al. concentrated on the performance of the BDI-II among 371 orphaned and vulnerable adolescents (13-17 years old) in childcare institutions. The authors emphasized the critical importance of accounting for trauma, loss, and institutional context when interpreting somatic and affective items. They recommended mixed-methods approaches for adaptation, recognizing that symptoms arising from grief, institutional deprivation, and trauma-related stress may be phenomenologically and etiologically distinct from symptoms of major depressive disorder, despite producing similar BDI score elevations [18].

This necessitates a reconceptualization of somatic items through the lens of “local biologies”—the acknowledgement that although human physiology is universal structurally, the manifestation and the meaning given to bodily experiences are very much dependent on the ecological, epidemiological, and social contexts [29]. In settings where malnutrition, parasitic infections, and infectious diseases are endemic, the phenomenological experience of “fatigue” or “appetite loss” fundamentally differs from experiences in resource-rich environments where these symptoms more reliably signal mood pathology [30]. The BDI is not capable of distinguishing fatigue caused by anemia, malaria, or intestinal parasites from the anhedonic fatigue that is typical of melancholic depression.

3.2. Criterion Validation and Cut-off Score Variability

Empirical evidence from exemplar studies demonstrates substantial variability in optimal cut-offs; several criterion-validation studies that underwent structured diagnostic interviews showed a wide range of optimal cut-offs. Adewuya et al., in their study in Nigeria, identified a cut-off of 18 as optimal for screening, yielding excellent sensitivity (0.91) and specificity (0.97) against K-SADS/E diagnoses. [14]. In contrast, the Indian Tamil validation by Basker et al. distinguished between a screening cut-off of 5 and a diagnostic cut-off of 22, suggesting that different thresholds may be appropriate depending on assessment purpose and clinical context [15].

This variability in cut-off results seriously challenges clinical practice and epidemiological research. Firstly, when researchers from various countries decide on different cut-offs to apply, or in case they uncritically apply a universal Western-derived threshold, cross-national prevalence estimates are not comparable. Secondly, the Chilean study's finding of sex-specific optimal cut-offs [16] Makes it even more difficult to deal with. If optimal thresholds vary not only across countries but also within each country by gender, this would suggest that population-level norms may be insufficient and that gender-specific reference data are needed.

3.3. The Neglected Adolescent Cognitive Context

Several studies noted developmental considerations that warrant attention. Toosi et al. even went so far as to suggest cognitive interviewing for younger adolescents, recognizing that early adolescents might interpret the abstract or metaphorical meanings of the items differently than older adolescents. The Basker et al. study, conducted

in primary care pediatric settings, emphasized the importance of pediatrician-administered support, suggesting that adolescents may need more structured support than is typically provided in self-administration.

These observations align with developmental neuroscience research documenting that adolescence involves ongoing maturation of prefrontal cortex regions responsible for abstract reasoning, emotional regulation, and metacognitive awareness [31]. Younger adolescents who are still mainly dependent on concrete operational thinking may interpret an item such as "I feel I am being punished" literally rather than metaphorically [32]. Similarly, items assessing self-evaluation and future orientation presuppose cognitive capacities for abstract self-reflection and temporal projection that continue to develop through adolescence [33]. Administration mode varied across exemplar studies and may contribute to response heterogeneity. Adewuya et al. [14] and most Brazilian/Chilean studies used standard self-administration. In contrast, Basker et al. [15] administered the BDI with pediatrician support in a clinical setting, while Kim et al. [22] and Bhatt et al. [18] used trained interviewers for orphaned or HIV-affected youth with potential literacy or trauma-related barriers. Clinician- or interviewer-assisted administration may reduce missing data and increase disclosure of sensitive items (e.g., suicidality, guilt), whereas self-report in school settings may heighten social desirability or literal interpretation—particularly in low-literacy contexts.

Considering that literacy and educational heterogeneity may further threaten construct validity in LMIC settings. The BDI assumes a minimum 5th–6th grade reading level in its original English form; translated versions rarely undergo readability validation. In contexts where adolescent literacy is variable—due to school dropout, limited educational access, or language mismatch—item comprehension may be compromised, biasing responses toward literal interpretation, acquiescence, or non-response. This risk is especially acute for abstract or metaphorical items (e.g., "I feel I am being punished"), which demand higher-order cognitive and linguistic processing [3].

Measurement Invariance and the Fallacy of Universal Cut-offs
Perhaps the most critical methodological gap across the reviewed studies is the near-complete absence of measurement invariance testing, without which no valid cross-cultural score comparisons can be made. Among the eight exemplar studies, only the Chilean study by Araya et al. and the Brazilian study by Ferreira-Maia et al. explicitly referenced invariance testing—and even these established it only partially or noted gender-specific differences, which made interpretation difficult [16, 21]. The finding of Araya et al. that optimal cut-offs differ by sex is a direct indication that scalar invariance—the state where item intercepts are equal across groups—does not hold, because equal cut-offs would only be valid under scalar invariance.

This gap has profound psychometric consequences, as measurement invariance testing proceeds hierarchically through configural, metric, and scalar levels [34]. Configural invariance requires equivalent factor structure; metric invariance, equivalent factor loadings; and scalar invariance, equivalent item intercepts. Only after establishing scalar invariance can latent mean differences be interpreted as reflecting true differences in the underlying construct rather than measurement artifacts [28].

The lack of invariance testing represents a fundamental psychometric gap. A score of 18 on the BDI in a Nigerian, Chilean, Indian, or Iranian adolescent sample cannot be assumed to reflect equivalent depression severity if item intercepts differ due to cultural response styles, differential stigma, linguistic nuance, or divergent symptom

conceptualization [35]. The use of Western-derived cut-offs in these circumstances conflates true disorder prevalence with measurement bias—a problem that is raised in Indian school prevalence studies [27].

Empirically, the failure to establish scalar invariance is mirrored in the divergent factor structures reported across validation studies [15, 19, 21], sex-specific optimal thresholds [16], and context-dependent somatic item interpretations [18, 22]. The findings, in fact, challenge the whole idea that BDI cut-off scores have the same diagnostic value across different cultural-linguistic settings. When scalar invariance is not achieved, the symptoms of two adolescents having the same level of depression—for example, one in Chile and the other in Nigeria—may reflect differently on their respective test scores because of culture-influenced ways of responding, stigma, or the translation meaning [36].

Epidemiological reports based on Western cut-offs in LMIC populations will likely yield an incorrect estimate of prevalence [24], while clinical screening or monitoring based on unvalidated thresholds risks systematic misclassification. Future validation research must therefore prioritize multi-group confirmatory factor analysis to sequentially test configural, metric, and scalar invariance across gender, age, and cultural subgroups [34, 37]. When invariance fails, researchers are required to identify items that function differentially, modify or discard them, or create context-specific norms indicative of culturally grounded symptom expression rather than the application of external psychometric universals. **Cultural Psychology and the Limits of Beck's Cognitive Model**
The poor and inconsistent performance of cognitive-affective items across multiple validation studies [14, 16, 19] reveals fundamental tension between the BDI's theoretical foundation and non-Western constructions of selfhood. This reflects deeper ontological differences in selfhood; as Beck's cognitive model presupposes an internalized, individualized self, in which negative automatic thoughts about personal worth, competence, and the future constitute core depressive pathology [38]. Yet, in collectivist cultures existing in numerous LMICs, the concept of self is more interdependent and is mainly determined by one's social roles, family relationships, and community status rather than personal characteristics [39].

This ontological difference has profound implications for symptom expression. Chentsova-Dutton and Tsai demonstrate that cultural values fundamentally shape not merely how depression is described but how it is phenomenologically experienced [40]. In individualist cultures, depressed individuals emphasize personal failure, worthlessness, and internal negative self-evaluation—precisely the cognitive content the BDI was designed to assess. On the other hand, in collectivist cultures, individuals may feel and communicate their distress primarily through physical symptoms, disruptions in interpersonal relations, or spiritual aspects rather than self-criticism [41, 42].

The sex differences in emotional expressivity items reported by Adewuya et al., and the sex-specific cut-offs identified by Araya et al., likely reflect culturally prescribed emotional display rules that interact with gender socialization [14, 16]. In many cultures, masculine ideals emphasize emotional stoicism, making male adolescents less likely to acknowledge symptoms related to crying, sadness display, or emotional vulnerability even when experiencing significant depressive distress [43].

These patterns necessitate a fundamental shift from "imposed etics"—treating Western psychological constructs as universal standards—toward combined "emic-etic" approaches that integrate universal dimensions of negative affect with locally salient idioms

of distress [44, 45]. This might involve retaining certain BDI items that assess transculturally relevant symptoms, while adding or changing those culturally problematic cognitive items with the local expressions of distress that make sense in the particular culture [46].

Ethical Imperatives and Guidance Beyond Psychometric Adequacy
Beyond psychometric limitations, ethical concerns require explicit addressing. Item 21 ("Loss of interest in sex") has been repeatedly identified as developmentally inappropriate and culturally problematic for adolescents, particularly in contexts where discussing sexuality is taboo, where many adolescents lack sexual experience, or where sexual activity before marriage is religiously proscribed [47–49]. This item routinely produces missing data, systematic under-reporting, and discomfort among respondents [50–52].

Similarly, Item 9 evaluating suicidal ideation raises ethical challenges in contexts where suicidality is stigmatized from a moral point of view or where talking about such thoughts may result in negative consequences. This demands ethically grounded instrument design: ethical principles for research require that instruments be not only psychometrically robust but also culturally sensitive and that they take into account participants' developmental stage and the social/environmental context [53]. To address these concerns, we recommend: (1) omit or reword Item 21 in developmentally or culturally inappropriate contexts; (2) administer Item 9 only with a safety protocol (trained staff, debriefing, and referral pathways); (3) pilot-test all items with local adolescents to assess acceptability and comprehension prior to use; and (4) tell participants they may skip any item—no coercion, no full-score pressure.

3.4. Methodological Recommendations

Advancing cross-cultural adolescent mental health research requires methodological rigor balanced with contextual feasibility and ethical sensitivity. We propose a three-component framework:

First, adopt an invariance-first approach as standard practice. Before any cross-group comparison, a multi-group confirmatory factor analysis should test configural, metric, and scalar invariance across gender, age, and cultural groups [28, 34]. If scalar invariance fails, the researchers are advised to highlight the items that display differential functioning and then decide whether to modify, eliminate, or create group-specific scoring algorithms.

Second, integrate mixed-methods approaches. Quantitative validation must be complemented by qualitative inquiry, particularly cognitive interviews and focus groups with adolescents from target populations [46]. Cognitive interviewing is an approach that can reveal semantic misunderstandings, identify culturally inappropriate content, and even detect developmentally challenging phrasing, all of which can be done before large-scale validation.

Third, incorporate criterion validity assessment. Even when gold-standard diagnostic interviews (K-SADS, MINI) are resource-intensive, conducting them with purposefully selected subsamples can establish locally appropriate cut-off scores. The Nigerian study's approach offers a feasible example: carrying out diagnostic interviews with extreme scorers to figure out the best sensitivity-specificity trade-offs [14].

Fourth, embrace decolonizing co-creative frameworks. Rather than merely adapting Western instruments, future research should involve local researchers and adolescents in the generation of items and pretesting [54]. Co-creative methods can initially involve questions that are directed towards adolescents and local mental health stakeholders, asking them how they identify, narrate, and comprehend depression in their context.

Fifth, as mobile health (mHealth) screening expands in LMICs [55], future validations must assess whether BDI's psychometric properties—established predominantly via paper-pencil—are preserved in digital formats, where interface design, literacy demands, and privacy perceptions may alter response behavior.

3.5. Limitations

This narrative review has several important limitations. First, our selective synthesis of eight exemplar studies, while appropriate for narrative review methodology, as Green et al. discussed [56], necessarily excludes potentially relevant validation research and cannot provide precise quantitative effect-size estimates that meta-analysis would afford. Second, the restriction to English-language publications constitutes a major limitation, likely biasing regional representation and underrepresenting culturally grounded validation work from non-Anglophone LMICs. Third, publication bias likely affects the literature we reviewed, as exemplar-based synthesis limits generalizability. Fourth, our synthesis did not stratify findings by urban/rural setting, socioeconomic status, or time of data collection—factors that may substantially influence symptom expression and instrument validity in rapidly changing LMIC contexts, and did not evaluate alternative instruments (e.g., PHQ-A, CES-D, or locally developed tools) that may offer superior cross-cultural validity or feasibility for adolescents in LMICs. Additionally, although two reviewers independently appraised study quality using our rubric and resolved disagreements by consensus, the initial exclusion of 46 full-text studies from in-depth synthesis was not independently verified. To enhance auditability, we now provide a supplementary table listing all excluded studies with their rubric scores and exclusion reasons. These version-specific differences (between BDI-I and BDI-II) may also independently influence psychometric properties, potentially confounding cultural and instrument-level variance. Future systematic reviews with meta-analytic synthesis offer more detailed quantitative summaries by utilizing tools such as COSMIN [53].

4. Conclusion

The Beck Depression Inventory, while representing a historically important milestone in depression assessment, embodies assumptions about symptom expression, selfhood, and psychological distress that demonstrably do not generalize uniformly across all cultural contexts. Based on the exemplar studies and broader literature, our critical review suggests that the BDI is a tool that consistently ranks (individuals within specific cultural-linguistic groups), as shown by high internal consistency; however, current evidence is insufficient to confirm whether the BDI measures equivalent constructs across diverse cultural contexts, highlighting the urgent need for formal measurement invariance testing.

The consistent pattern of high internal reliability creates what we term a "reliability paradox"—a dangerous illusion of cross-cultural utility that masks profound threats to construct validity. Factor structure heterogeneity, somatic symptom confounding in medically compromised populations, the almost complete lack of formal testing for scalar invariance, sex-specific cut-off requirements, and the context-dependent interpretation of cognitive-affective items together point out the fundamental limits to the cross-cultural applicability of the BDI.

These limitations reflect deeper epistemological tensions between universalist assumptions inherent in Western psychiatric nosology and the cultural psychology recognition that cultural meaning systems, social structures, and local biologies fundamentally shape psychological experiences.

Ethically and accurately assessing adolescent depression in LMICs necessitates the abandonment of the administrative convenience of mere exported instruments in favor of contextually grounded, developmentally sensitive, and psychometrically sound tools that emerge from collaborative, co-creative processes. The continued use of Western-derived instruments without rigorous cross-cultural validation—particularly without measurement invariance testing and locally derived cut-offs—raises serious concerns about the validity of epidemiological estimates, the appropriateness of clinical thresholds, and the comparability of research findings across cultural contexts.

Conflicts of Interest

The authors declare that they have no competing interests.

Funding Source

No funding was received for this work.

Acknowledgment

None.

Institutional Review Board (IRB)

The institutional review board (IRB) approval was not required for this study.

Large Language Model

A locally deployed large language model (Qwen2-VL-72B-Instruct API) for personal research use was used exclusively to enhance language quality, coherence, and logical consistency in limited parts of the Review section. No confidential, proprietary, or participant-related information was entered into the model. All AI-assisted content was carefully reviewed and approved by the authors, who assume full responsibility for the accuracy, originality, and integrity of the manuscript.

Authors Contribution

AA conceptualized the topic, conducted the literature search, ZE interpreted the data, and AT and MG supervised the manuscript. All authors read and approved the final version of the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and its reference list.

References

- Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. *Lancet*. 2011;378(9801):1515-25. [PMID: 22008427, [https://doi.org/10.1016/S0140-6736\(11\)60827-1](https://doi.org/10.1016/S0140-6736(11)60827-1)].
- WHO. Mental health of adolescents [Internet]; 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health>.
- Beck AT, Steer RA, Brown G. Beck Depression Inventory-II. *PsychTESTS Dataset*. 2011. [<https://doi.org/10.1037/t00742-000>].
- Fernando S. Race and culture issues in mental health and some thoughts on ethnic identity. *Counselling Psychology Quarterly*. 2012;25(2):113-23. [<https://doi.org/10.1080/09515070.2012.674299>].
- van dVf, Tanzer NK. Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*. 2004. [<https://doi.org/10.1016/j.erap.2003.12.004>].
- Byrne BM, van de Vijver FJR. The maximum likelihood alignment approach to testing for approximate measurement invariance: A paradigmatic cross-cultural application. *Psicothema*. 2017;29(4):539-51. [PMID: 29048316, <https://doi.org/10.7334/psicothema2017.178>].
- Crone EA, Dahl RE. Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nat Rev Neurosci*. 2012;13(9):636-50. [PMID: 22903221, <https://doi.org/10.1038/nrn3313>].
- Kirmayer LJ. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *J Clin Psychiatry*. 2001;62 Suppl 13:22-8; discussion 29-30. [PMID: 11434415].
- Jorns-Presentati A, Napp AK, Dessauvage AS, Stein DJ, Jonker D, Breet E, et al. The prevalence of mental health problems in sub-Saharan adolescents: A systematic review. *PLoS One*. 2021;16(5):e0251689. [PMID: 33989357, PMCID: PMC8121357, <https://doi.org/10.1371/journal.pone.0251689>].
- Akhtar P, Ma L, Waqas A, Naveed S, Li Y, Rahman A, et al. Prevalence of depression among university students in low and middle income countries (LMICs): a systematic review and meta-analysis. *J Affect Disord*. 2020;274:911-9. [PMID: 32664032, <https://doi.org/10.1016/j.jad.2020.03.183>].
- Wubshet I, Engdawork K, Gebremariam S, Kanazayire C, Abbott P. The Social Determinants of Depression among Adolescents in Low- and Middle-Income Countries: A Scoping Review. <https://doi.org/10.1101/2025040725325376>. 2025. [<https://doi.org/10.1101/2025.04.07.25325376>].
- Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? *Eur J Clin Invest*. 2018;48(6):e12931. [PMID: 29578574, PMCID: PMC6001568, <https://doi.org/10.1111/eci.12931>].
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge. 2013. [<https://doi.org/10.4324/9780203771587>].
- Adewuya AO, Ola BA, Aloba OO. Prevalence of major depressive disorders and a validation of the Beck Depression Inventory among Nigerian adolescents. *Eur Child Adolesc Psychiatry*. 2007;16(5):287-92. [PMID: 17473949, <https://doi.org/10.1007/s00787-006-0557-0>].
- Basker M, Moses PD, Russell S, Russell PS. The psychometric properties of Beck Depression Inventory for adolescent depression in a primary-care paediatric setting in India. *Child Adolesc Psychiatry Ment Health*. 2007;1(1):8. [PMID: 17688697, PMCID: PMC1976414, <https://doi.org/10.1186/1753-2000-1-8>].
- Araya R, Montero-Marin J, Barrioilhet S, Fritsch R, Gaete J, Montgomery A. Detecting depression among adolescents in Santiago, Chile: sex differences. *BMC Psychiatry*. 2013;13(1):122. [PMID: 23617306, PMCID: PMC3637822, <https://doi.org/10.1186/1471-244X-13-122>].
- Kim MH, Mazenga AC, Devandra A, Ahmed S, Kazembe PN, Yu X, et al. Prevalence of depression and validation of the Beck Depression Inventory-II and the Children's Depression Inventory-Short amongst HIV-positive adolescents in Malawi. *Journal of the International AIDS Society*. 2014;17(1). [<https://doi.org/10.7448/ias.17.1.18965>].
- Bhatt KB, Apidechkul T, Srichan P, Bhatt N. Depressive symptoms among orphans and vulnerable adolescents in childcare homes in Nepal: a cross-sectional study. *BMC Psychiatry*. 2020;20(1):466. [PMID: 32977787, PMCID: PMC7517808, <https://doi.org/10.1186/s12888-020-02863-y>].
- Toosi F, Rahimi C, Sajjadi S. Psychometric Properties of Beck Depression Inventory-II for High School Children in Shiraz City, Iran. *International Journal of School Health*. 2017;In press(In press). [<https://doi.org/10.5812/intjsh.41069>].
- Jha KK, Singh SK, Nirala SK, Kumar C, Kumar P, Aggrawal N. Prevalence of Depression among School-going Adolescents in an Urban Area of Bihar, India. *Indian Journal of Psychological Medicine*. 2017;39(3):287-92. [<https://doi.org/10.4103/0253-7176.207326>].

21. Ferreira-Maia AP, Boronat AC, Gorenstein C, Wang YP. 5.39 Reliability and Dimensional Validity of the Beck Depression Inventory-II (BDI-II) Among Brazilian Adolescents. *Journal of the American Academy of Child Adolescent Psychiatry*. 2018;57(10):S239. [https://doi.org/10.1016/j.jaac.2018.09.332].
22. Kim MH, Mazenga AC, Devandra A, Ahmed S, Kazembe PN, Yu X, et al. Prevalence of depression and validation of the Beck Depression Inventory-II and the Children's Depression Inventory-Short amongst HIV-positive adolescents in Malawi. *J Int AIDS Soc*. 2014;17(1):18965. [PMID: 25085002, PMCID: PMC4119288, https://doi.org/10.7448/IAS.17.1.18965.18965].
23. Flake JK, Pek J, Hehman E. Construct Validation in Social and Personality Research. *Social Psychological and Personality Science*. 2017;8(4):370-8. [https://doi.org/10.1177/1948550617693063].
24. Bolton P, Neugebauer R, Ndogoni L. Prevalence of depression in rural Rwanda based on symptom and functional criteria. *J Nerv Ment Dis*. 2002;190(9):631-7. [PMID: 12357098, https://doi.org/10.1097/00005053-200209000-00009].
25. Sijsma K. On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*. 2009;74(1):107-20. [PMID: 20037639, PMCID: PMC2792363, https://doi.org/10.1007/s11336-008-9101-0].
26. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care Pain*. 2008;8(6):221-3. [https://doi.org/10.1093/bjaceaccp/mkn041].
27. Jha KK, Singh SK, Nirala SK, Kumar C, Kumar P, Aggrawal N. Prevalence of Depression among School-going Adolescents in an Urban Area of Bihar, India. *Indian J Psychol Med*. 2017;39(3):287-92. [PMID: 28615762, PMCID: PMC5461838, https://doi.org/10.4103/0253-7176.207326].
28. Chen FF. What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *J Pers Soc Psychol*. 2008;95(5):1005-18. [PMID: 18954190, https://doi.org/10.1037/a0013193].
29. Lock M. Cultivating the Body: Anthropology and Epistemologies of Bodily Practice and Knowledge. *Annual Review of Anthropology*. 1993;22(1):133-55. [https://doi.org/10.1146/annurev.an.22.100193.001025].
30. Kohrt BA, Hruschka DJ. Nepali concepts of psychological trauma: the role of idioms of distress, ethnopsychology and ethnophysiology in alleviating suffering and preventing stigma. *Cult Med Psychiatry*. 2010;34(2):322-52. [PMID: 20309724, PMCID: PMC3819627, https://doi.org/10.1007/s11013-010-9170-2].
31. Blakemore SJ, Mills KL. Is adolescence a sensitive period for socio-cultural processing? *Annu Rev Psychol*. 2014;65(1):187-207. [PMID: 24016274, https://doi.org/10.1146/annurev-psych-010213-115202].
32. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. *J Abnorm Psychol*. 1998;107(1):128-40. [PMID: 9505045, https://doi.org/10.1037/0021-843x.107.1.128].
33. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. *Lancet*. 2012;379(9820):1056-67. [PMID: 22305766, PMCID: PMC3488279, https://doi.org/10.1016/S0140-6736(11)60871-4].
34. Putnick DL, Bornstein MH. Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Dev Rev*. 2016;41:71-90. [PMID: 27942093, PMCID: PMC5145197, https://doi.org/10.1016/j.dr.2016.06.004].
35. Gregorich SE. Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care*. 2006;44(11 Suppl 3):S78-94. [PMID: 17060839, PMCID: PMC1808350, https://doi.org/10.1097/01.mlr.0000245454.12228.8f].
36. Weiss B, Dang M, Trung L, Nguyen MC, Thuy NT, Pollack A. A Nationally-Representative Epidemiological and Risk Factor Assessment of Child Mental Health in Vietnam. *Int Perspect Psychol*. 2014;3(3):139-53. [PMID: 25328817, PMCID: PMC4199236, https://doi.org/10.1037/ipp0000016].
37. Davidov E, Meuleman B, Cieciuch J, Schmidt P, Billiet J. Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*. 2014;40(1):55-75. [https://doi.org/10.1146/annurev-soc-071913-043137].
38. Beck AT. The evolution of the cognitive model of depression and its neurobiological correlates. *Am J Psychiatry*. 2008;165(8):969-77. [PMID: 18628348, https://doi.org/10.1176/appi.ajp.2008.08050721].
39. Markus HR, Kitayama S. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*. 1991;98(2):224-53. [https://doi.org/10.1037/0033-295x.98.2.224].
40. Chentsova-Dutton YE, Tsai JL. Self-focused attention and emotional reactivity: the role of culture. *J Pers Soc Psychol*. 2010;98(3):507-19. [PMID: 20175627, https://doi.org/10.1037/a0018534].
41. Kirmayer LJ, Young A. Culture and somatization: clinical, epidemiological, and ethnographic perspectives. *Psychosom Med*. 1998;60(4):420-30. [PMID: 9710287, https://doi.org/10.1097/00006842-199807000-00006].
42. Ryder AG, Yang J, Zhu X, Yao S, Yi J, Heine SJ, et al. The cultural shaping of depression: somatic symptoms in China, psychological symptoms in North America? *J Abnorm Psychol*. 2008;117(2):300-13. [PMID: 18489206, https://doi.org/10.1037/0021-843X.117.2.300].
43. Chaplin TM, Aldao A. Gender differences in emotion expression in children: a meta-analytic review. *Psychol Bull*. 2013;139(4):735-65. [PMID: 23231534, PMCID: PMC3597769, https://doi.org/10.1037/a0030737].
44. Canino G, Alegria M. Psychiatric diagnosis - is it universal or relative to culture? *J Child Psychol Psychiatry*. 2008;49(3):237-50. [PMID: 18333929, PMCID: PMC3104469, https://doi.org/10.1111/j.1469-7610.2007.01854.x].
45. Kleinman AM. Depression, somatization and the "new cross-cultural psychiatry". *Soc Sci Med* (1967). 1977;11(1):3-10. [PMID: 887955, https://doi.org/10.1016/0037-7856(77)90138-x].
46. Kaiser BN, Ticao C, Anojé C, Minto J, Boglosa J, Kohrt BA. Adapting culturally appropriate mental health screening tools for use among conflict-affected and other vulnerable adolescents in Nigeria. *Glob Ment Health (Camb)*. 2019;6:e10. [PMID: 31258924, PMCID: PMC6582460, https://doi.org/10.1017/gmh.2019.8].
47. Askari F, Mirzaiinajmabadi K, Saeedy Rezvani M, Asgharinekah SM. Sexual health education issues (challenges) for adolescent boys in Iran: A qualitative study. *J Educ Health Promot*. 2020;9(1):33. [PMID: 32318601, PMCID: PMC7161661, https://doi.org/10.4103/jehp.jehp_462_19].
48. Joodaki K, Nedjat S, Vahid Dastjerdi M, Larijani B. Ethical considerations and challenges of sex education for adolescents in Iran: a qualitative study. *J Med Ethics Hist Med*. 2020;13:2. [PMID: 33088429, PMCID: PMC7569533, https://doi.org/10.18502/jmehm.v13i2.2664].
49. Osman A, Barrios FX, Gutierrez PM, Williams JE, Bailey J. Psychometric properties of the Beck Depression Inventory-II in nonclinical adolescent samples. *J Clin Psychol*. 2008;64(1):83-102. [PMID: 18161034, https://doi.org/10.1002/jclp.20433].
50. Lipton MF, Augenstein TM, Weeks JW, De LRA. A Multi-informant Approach to Assessing Fear of Positive Evaluation in Socially Anxious Adolescents. *Journal of Child and Family Studies*. 2013;23(7):1247-57. [https://doi.org/10.1007/s10826-013-9785-3].
51. Los Reyes A, Makol BA, Racz SJ, Youngstrom EA, Lerner MD, Keeley LM. The Work and Social Adjustment Scale for Youth: A Measure for Assessing Youth Psychosocial Impairment Regardless of Mental Health Status. *J Child Fam Stud*. 2019;28(1):1-16. [PMID: 33311964, PMCID: PMC7731438, https://doi.org/10.1007/s10826-018-1238-6].
52. Rausch E, Racz SJ, Augenstein TM, Keeley L, Lipton MF, Szollos S, et al. A Multi-Informant Approach to Measuring Depressive Symptoms in Clinical Assessments of Adolescent Social Anxiety Using the Beck Depression Inventory-II: Convergent, Incremental, and Criterion-Related Validity. *Child Youth Care Forum*. 2017;46(5):661-83. [https://doi.org/10.1007/s10566-017-9403-4].
53. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-57. [PMID: 29435801, PMCID: PMC5891568, https://doi.org/10.1007/s11136-018-1798-3].

54. Lewis-Fernández R, Aggarwal Neil K, Hinton L, Hinton Devon E, Kirmayer Laurence J. DSM-5® Handbook on the Cultural Formulation Interview. DSM-5® Handbook on the Cultural Formulation Interview. 2015. [<https://doi.org/10.1176/appi.books.9781615373567>].
55. Bhatia RK, Bhardwaj P, Chaudhry A, Ghosh A, Parmar A, Ahluwalia Y, et al. A review of systematic reviews for evidence on use of mobile applications for mental disorders, including substance use disorders, in low and middle-income countries. Digit Health. 2025;11:20552076251387056. [PMID: 41143094, PMCID: PMC12547136, <https://doi.org/10.1177/20552076251387056>].
56. Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. J Chiropr Med. 2006;5(3):101-17. [PMID: 19674681, PMCID: PMC2647067, [https://doi.org/10.1016/S0899-3467\(07\)60142-6](https://doi.org/10.1016/S0899-3467(07)60142-6)].



Original Article

Selective Attenuation of Acetaminophen-induced Hepatobiliary Dysfunction by *Pseudoelephantopus spicatus* in Sprague-Dawley Rats

Uzzal Chondra^{1,*}, Palash C. Roy², Khaled H. Riad², Masud A. Sumon², Md. Zahirul Islam², S.M. Nazmul Islam², Md. Jahid Hasan², Pabitra K. Shill², Mithun Sikder³, Shawon S. Joy⁴, Annesha Majumder²

1-Chikitsa Foundation Bangladesh (CF-BD), Dhaka, Bangladesh

2-Department of Pharmacy, Gono Bishwabidyalay, Dhaka, Bangladesh

3-Bangladesh Rural Advancement Committee (BRAC), Dhaka, Bangladesh

4-Department of Physiotherapy, Gono Bishwabidyalay, Dhaka, Bangladesh

ARTICLE INFO

Article history:

Received 31 Jan. 2026

Received in revised form 27 Mar. 2026

Accepted 4 Apr. 2026

Published 8 Apr. 2026

Keywords:

Acetaminophen

Hepatobiliary biomarkers

Liver enzymes

Phytotherapy

Pseudoelephantopus spicatus

Rat model

ABSTRACT

Background: *Pseudoelephantopus spicatus* (PS) is used in ethnomedicine for hepatic disorders; however, its effects on biochemical markers of drug-induced liver injury remain incompletely characterized. This study evaluated the effects of the ethanolic root extract of PS on serum biochemical indices in an acetaminophen-induced acute hepatotoxicity model.

Methods: Sprague-Dawley rats (150-200 g) were allocated into five groups (n = 6/group). PS extract (500 and 1000 mg/kg) and silymarin (100 mg/kg) were administered orally for 10 days. On day 10, hepatotoxicity was induced by a single oral dose of acetaminophen (700 mg/kg) after a 12 h fast. Serum ALT, AST, ALP, total bilirubin, total protein, albumin, and globulin were quantified 24 h post-administration.

Results: Acetaminophen produced marked elevations in ALT, AST, ALP, and total bilirubin, with reduced serum protein fractions. PS treatment did not significantly reduce ALT or AST levels. However, PS significantly decreased ALP activity at both 500 and 1000 mg/kg (p < 0.001) and reduced total bilirubin dose-dependently (p < 0.01-p < 0.001 vs. acetaminophen control). At 1000 mg/kg, PS significantly restored total protein (p < 0.001) and increased albumin and globulin levels (p < 0.05), indicating partial recovery of hepatic synthetic function.

Conclusion: PS ethanolic root extract produced selective biochemical attenuation of acetaminophen-induced hepatobiliary dysfunction, significantly modulating ALP, total bilirubin, and circulating proteins without significant transaminase normalization. These findings indicate partial functional improvement rather than comprehensive hepatocellular protection. Further studies incorporating histopathology, oxidative stress, and mechanistic analyses are required to establish structural and molecular correlates.

1. Introduction

The liver is the primary site for the metabolism and elimination of xenobiotics and endogenous compounds, rendering it particularly vulnerable to chemically induced injury [1]. Drug-induced liver injury represents a major cause of acute hepatic failure, with acetaminophen overdose being one of the most extensively studied and clinically relevant models of hepatotoxicity [2]. Under therapeutic conditions, acetaminophen is predominantly metabolized through glucuronidation and sulfation pathways; however, excessive dosing leads to cytochrome P450-mediated bioactivation to the highly reactive metabolite N-acetyl-p-benzoquinone imine (NAPQI) [3]. Accumulation of NAPQI results in depletion of intracellular glutathione reserves, covalent binding to hepatic proteins, mitochondrial dysfunction, enhanced generation of reactive oxygen species, and subsequent hepatocellular necrosis [4]. Acetaminophen-induced

liver damage is commonly associated with altered ALT, AST, ALP, bilirubin, and protein levels, which are used as biochemical indicators of liver function [5].

Currently available hepatoprotective therapies are limited in efficacy, necessitating the exploration of alternative agents capable of modulating oxidative, inflammatory, and metabolic pathways involved in liver damage [6]. While oxidative stress and inflammation are recognized contributors to acetaminophen-induced hepatotoxicity, this study focused specifically on routine serum biochemical markers-ALT, AST, ALP, bilirubin, and circulating protein fractions-as primary endpoints. These biomarkers provide reliable, accessible measures of hepatocellular and hepatobiliary function, reflecting functional impairment and recovery without requiring direct assessment of oxidative or inflammatory pathways [7, 8]. In this context, medicinal plants constitute a valuable source of structurally diverse bioactive compounds with potential hepatoprotective properties [9].

P. spicatus (Juss.) Rohr. (Asteraceae) is a perennial herb traditionally used to manage inflammatory disorders, infectious diseases, and hepatic dysfunction [10]. Phytochemical investigations have revealed the presence of sesquiterpene lactones, triterpenoids, sterols, and related secondary metabolites, several of which exhibit antioxidant, anti-inflammatory, and cytoprotective activities [11]. Experimental studies have reported moderate hepatoprotective effects of *P. spicatus*

*Corresponding author: Uzzal Chondra, Chikitsa Foundation Bangladesh (CF-BD), Dhaka, Bangladesh. Email: uzzal.chondra@gmail.com

Published by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. ©2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosted by ASIDE Journals.

Citation: Chondra U, C. Roy P, H. Riad K, et al. Selective Attenuation of Acetaminophen-induced Hepatobiliary Dysfunction by *Pseudoelephantopus spicatus* in Sprague-Dawley Rats. ASIDE Health Sci. 2026;2(2):33-39, doi:10.71079/ASIDE.HS.040826549

extracts against chemically induced liver injury, including toxin and drug-mediated models [12]; however, systematic evaluation of organ-specific extracts, particularly roots, in acetaminophen-induced hepatotoxicity remains limited.

Accordingly, the present study was designed to evaluate the effects of the ethanolic root extract of *P. spicatus* on serum biochemical markers of hepatocellular and hepatobiliary injury in a rat model of acetaminophen-induced acute hepatotoxicity. By focusing on these routine endpoints, the study aims to establish the functional hepatoprotective potential of *P. spicatus* before mechanistic or histopathological investigations.

2. Methods

2.1. Plant Collection and Extraction

The plant of *P. spicatus* was collected from Dhamrai, Dhaka, Bangladesh, and identified by the Bangladesh National Herbarium (BNH) and accession number (DACB87366). Thoroughly washed, shade-dried at 35–40 °C for 7 days, pulverized to a fine powder, and subjected to Soxhlet extraction using 96% ethanol (900 g powder; 30 h extraction). The combined ethanolic extract was concentrated under reduced pressure using a rotary evaporator (BUCHI Rota Vapor R-114, Switzerland), further dried on a water bath at room temperature to yield a solid extract (100 g), which was stored at 4 °C until use. Although the plant material was taxonomically authenticated and a voucher specimen deposited, the ethanolic extract was not standardized using a quantified chemical marker or chromatographic fingerprint, and variation in phytochemical composition may therefore influence reproducibility.

2.2. Chemicals and Reagents

Acetaminophen (APAP) was procured from Gonoshasthaya Pharmaceuticals Ltd. aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), and total bilirubin (TB) were obtained from HUMAN GmbH (Germany), while total protein and albumin were quantified using colorimetric diagnostic kits supplied by Erba Mannheim (Transasia Bio-Medicals Ltd., India); all other reagents used in the study were of analytical grade.

2.3. Experimental Animals

Sprague-Dawley rats of either sex (4–8 weeks old) were housed three per polypropylene cage with standard bedding and environmental enrichment materials under standard hygienic conditions at the animal house of the Department of Pharmacy, Jahangirnagar University. Environmental conditions were maintained at 27 ± 2 °C with a 12 h light/12 h dark cycle, and animals had free access to a standard pellet diet and water. All animals allocated to the experimental groups (n = 6 per group) were included in the final biochemical evaluation, and no animals were excluded from analysis. All procedures were conducted in accordance with institutional and international ethical guidelines. Animals of both sexes were included; however, the allocation of males and females within individual experimental groups was not recorded during the study. Consequently, sex-balanced group assignment and sex-specific analyses could not be performed.

2.4. Acute Oral Toxicity Study

A preliminary acute oral toxicity study was conducted in Swiss albino female mice (25–30 g, n = 4 animals per dose step) as an exploratory dose-tolerability assessment using a stepwise design broadly based on the acute toxic class approach described in OECD Test Guideline 423, in which animals received a single oral dose of ethanolic *P. spicatus* root extract and were continuously monitored for 4 h, periodically over 24 h, and daily up to 14 days for clinical

signs, including locomotor activity, grooming, respiration, food and water intake, tremors, and mortality, with body weights recorded on days 0, 7, and 14. No mortality or treatment-related clinical signs were observed at doses up to 1000 mg/kg. Based on these observations, 500 and 1000 mg/kg were selected for subsequent biochemical evaluation in Sprague-Dawley rats [10, 13].

2.5. Experimental Design and Assessment of Liver Function

The in vivo study was conducted at the Pharmacology Laboratory, Department of Pharmacy, Jahangirnagar University (Savar, Dhaka), using adult Sprague-Dawley rats of either sex (150–200 g) housed in polypropylene cages under controlled environmental conditions (27 ± 2 °C, standard pellet diet, ad libitum access to water) following a minimum 2-week acclimatization period. All protocols for the animal experiment were approved by the Research Ethical Committee under the Center for Multidisciplinary Research of Chikitsa Foundation Bangladesh (Approval No. CF-BD/EC/023). Animals were monitored daily for clinical signs of toxicity or distress, including changes in behavior, locomotion, grooming, and food and water intake, and humane endpoints were applied in accordance with approved ethical guidelines. Before acetaminophen administration, animals received their respective treatments once daily for 10 consecutive days. Rats were randomly assigned to five groups (n = 6 per group) after body weight measurement to ensure comparable mean body weights across groups. Group allocation was performed using a simple randomization method based on a random number sequence.

Group I received distilled water (10 mL/kg, p.o.) for 10 days and served as the normal control.

Group II received distilled water (10 mL/kg, p.o.) for 10 days, followed by a single oral dose of acetaminophen (700 mg/kg).

Group III received silymarin (100 mg/kg, p.o.) for 10 days, followed by a single oral dose of acetaminophen (700 mg/kg).

Groups IV and V received *P. spicatus* root extract (500 and 1000 mg/kg, p.o., respectively) for 10 days, followed by a single oral dose of acetaminophen (700 mg/kg).

Acute hepatotoxicity was induced by a single oral administration of acetaminophen (700 mg/kg) on day 10, administered 1 h after the final dose of the respective treatments, following 12 h of overnight fasting [14, 15]. Food was withheld for 12 h before acetaminophen administration, while water was provided ad libitum. The animals were observed for 24 h after dosing. The dried extract yield was 11.1% (w/w) relative to the starting plant material. The extract was suspended in distilled water and administered orally at a dosing volume of 10 mL/kg body weight. The formulation was freshly prepared, stored at 25 °C protected from light, and used within 2 h to ensure stability. Following treatment, animals were anesthetized with ketamine (50 mg/kg, intraperitoneal) [16, 17] to achieve deep anesthesia before blood collection. Animals were sacrificed under anesthesia in accordance with institutional ethical approval. Blood samples were collected and centrifuged (3000 rpm, 10 min) to obtain serum, and liver function biomarkers, including ALT, AST, ALP, and total bilirubin [18], total protein, and albumin were determined using standard enzymatic colorimetric assays (Figure 1) [19]. Serum biochemical assays and statistical analyses were conducted using coded samples so that investigators performing the analyses were unaware of the treatment group allocation.

2.6. Statistical Analysis

Data were analyzed using one-way ANOVA followed by Bonferroni's post hoc test in SPSS v16.0. Results are presented as mean ±

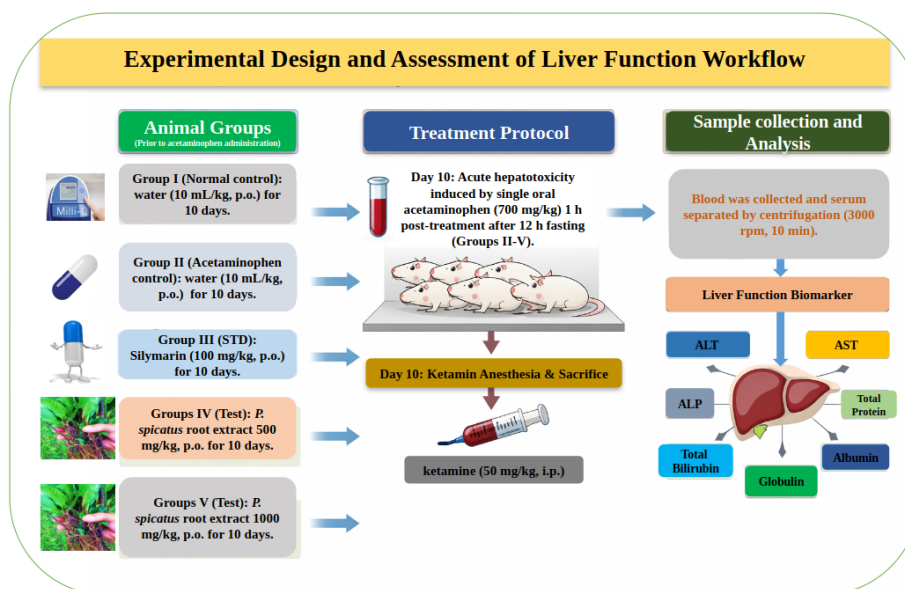


Figure 1: Schematic overview of the experimental design and liver function assessment workflow illustrating animal grouping, treatment regimen, sample collection, and biochemical evaluation of hepatic biomarkers in acetaminophen-induced hepatotoxicity in rats.

SEM, with statistical significance was defined as * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$, where the normal control group (Group I) was compared with the acetaminophen control (Group II) to confirm hepatotoxicity, and each treatment group-PS 500 mg/kg (Group IV), PS 1000 mg/kg (Group V), and silymarin 100 mg/kg (Group III)-was statistically compared exclusively against the acetaminophen control (Group II) without intergroup comparisons or reference to the normal control.

3. Results

3.1. Acetaminophen-induced hepatic injury and effects on serum transaminases and ALP by *P. spicatus*

Acetaminophen administration resulted in elevated serum ALT, AST, and ALP levels compared with the normal control group, consistent with the expected biochemical pattern of hepatic injury in this experimental model. Relative to the acetaminophen control group, treatment with *P. spicatus* (PS) at 500 mg/kg did not produce statistically significant reductions in ALT or AST levels; however, ALP activity was significantly decreased ($p < 0.001$). Similarly, PS at 1000 mg/kg did not significantly alter ALT or AST levels compared with the acetaminophen control group, while ALP levels were significantly reduced ($p < 0.001$), suggesting selective modulation of hepatobiliary enzyme activity. In contrast, silymarin (100 mg/kg) significantly decreased ALT, AST, and ALP levels ($p < 0.001$) relative to the acetaminophen control group. Collectively, PS treatment was associated with a significant reduction in ALP but not in transaminase levels, whereas silymarin significantly reduced all measured hepatic enzyme parameters (Table 1) and (Figure 2).

3.2. Acetaminophen-induced alterations in serum total bilirubin and effects of *P. spicatus*

Acetaminophen administration resulted in elevated serum total bilirubin levels compared with the normal control group. Relative to the acetaminophen control group, treatment with *P. spicatus* (PS) at 500 mg/kg significantly reduced serum total bilirubin levels ($p < 0.01$). Administration of PS at 1000 mg/kg produced a further significant reduction in bilirubin levels ($p < 0.001$) compared

with the acetaminophen control group. Similarly, silymarin (100 mg/kg) significantly decreased serum total bilirubin levels ($p < 0.001$) relative to the acetaminophen control group. Collectively, PS treatment was associated with a dose-dependent reduction in serum bilirubin levels compared with acetaminophen-induced elevation (Table 1) and (Figure 3).

3.3. Acetaminophen-induced alterations in serum total protein, albumin, and globulin, and effects of *P. spicatus*

Acetaminophen administration resulted in marked reductions in serum total protein, albumin, and globulin levels compared with those in the normal control group. Treatment with *P. spicatus* (PS) at 500 mg/kg did not produce statistically significant changes in any of these parameters relative to the acetaminophen control group. In contrast, PS at 1000 mg/kg significantly increased serum total protein levels ($p < 0.001$) and elevated albumin and globulin concentrations ($p < 0.05$) compared with the acetaminophen control group. Silymarin (100 mg/kg) significantly increased serum total protein levels ($p < 0.01$) relative to the acetaminophen control group, whereas albumin and globulin showed no significant changes (Table 2) and (Figure 4). Mean $\pm$ SEM values for serum ALT, AST, ALP, total bilirubin, total protein, albumin, and globulin across all experimental groups, together with corresponding statistical significance levels, are provided in (Tables 1 and 2) to ensure full transparency of inter-animal variability and the dispersion of biochemical responses.

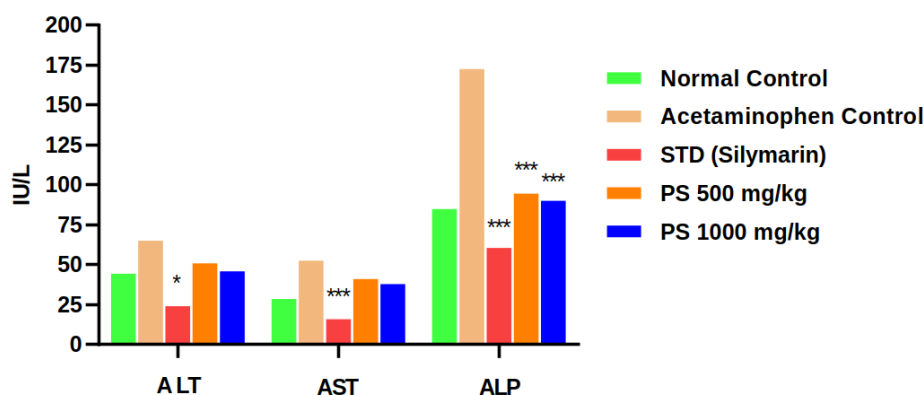
4. Discussion

Acetaminophen (APAP) is widely used as an analgesic and antipyretic, yet overdose results in severe hepatotoxicity due to metabolic activation and oxidative stress [20]. Under therapeutic conditions, acetaminophen is primarily detoxified via glucuronidation and sulfation pathways; however, a minor fraction is bioactivated by cytochrome P450 enzymes to the reactive intermediate N-acetyl-p-benzoquinone imine (NAPQI), which is normally neutralized through conjugation with reduced glutathione (GSH) [21]. Following overdose, saturation of conjugation pathways leads to excessive

Table 1: Effects of *P. spicatus* on hepatic injury biomarkers in acetaminophen-induced rat hepatotoxicity

Group	ALT (IU/L)	AST (IU/L)	ALP (IU/L)	Total Bilirubin (mg/dl)
Normal Control	44.0 ± 2.25	28.33 ± 3.72	84.50 ± 9.62	0.78 ± 0.08
Acetaminophen Control	64.67 ± 10.80	52.17 ± 5.78	172.33 ± 13.36	1.92 ± 0.15
STD (Silymarin 100 mg/kg)	23.83 ± 8.85*	15.50 ± 1.67***	60.33 ± 4.09***	0.55 ± 0.12***
PS 500 mg/kg	50.50 ± 7.31	40.67 ± 2.09	94.33 ± 9.83***	1.07 ± 0.18**
PS 1000 mg/kg	45.50 ± 12.71	37.50 ± 4.91	89.83 ± 3.69***	0.65 ± 0.06***

Data are expressed as mean ± SEM (n = 6) and were analyzed by one-way ANOVA followed by Bonferroni's post hoc test; p* < 0.05, p** < 0.01, and p*** < 0.001 indicate statistically significant, highly significant, and very highly significant differences, respectively, where the normal control group (Group I) was compared with the acetaminophen control (Group II) to confirm hepatotoxicity, and each treatment group—PS 500 mg/kg (Group IV), PS 1000 mg/kg (Group V), and silymarin 100 mg/kg (Group III)—was statistically compared exclusively against the acetaminophen control (Group II) without intergroup comparisons or reference to the normal control.

**Figure 2:** Modulatory effects of *P. spicatus* root extract on serum alanine aminotransferase (ALT), aspartate aminotransferase (AST), and alkaline phosphatase (ALP) activities in acetaminophen-induced hepatotoxicity in rats.

NAPQI formation, hepatic GSH depletion, and covalent binding of NAPQI to cellular macromolecules, triggering lipid peroxidation, calcium dysregulation, mitochondrial dysfunction, and hepatocellular necrosis [22, 23]. Hepatocellular damage disrupts membrane integrity, resulting in leakage of cytosolic enzymes into the circulation, making serum transaminases and alkaline phosphatase reliable biomarkers of liver injury [24, 25]. These biomarkers reflect both hepatocellular membrane integrity and hepatobiliary functional status, and differential modulation of these parameters may indicate selective biochemical effects rather than uniform cytoprotection.

In the present study, acetaminophen administration was associated with elevated serum ALT, AST, ALP, and bilirubin levels, confirming hepatocellular and hepatobiliary dysfunction. Treatment with *P. spicatus* (PS) root extract attenuated these alterations, with significant reductions observed in ALP at both doses and in bilirubin in a dose-dependent manner. However, ALT and AST levels were not significantly reduced at either dose, indicating that the extract did not demonstrate statistically significant protection against hepatocellular enzyme leakage under the present experimental conditions. The higher dose produced marked reductions in ALP and bilirubin levels, suggesting a possible modulatory effect on hepatobiliary parameters rather than comprehensive hepatocellular protection [26]. Because

Table 2: Modulation of serum total protein, albumin, and globulin by *P. spicatus* in acetaminophen-induced hepatotoxicity

Group	Total Protein (g/dl)	Albumin (g/dl)	Globulin (g/dl)
Normal Control	4.95 ± 0.36	3.06 ± 0.11	2.72 ± 0.35
Acetaminophen Control	3.08 ± 0.33	2.12 ± 0.27	1.75 ± 0.16
STD (Silymarin 100 mg/kg)	4.90 ± 0.20**	2.87 ± 0.31	2.60 ± 0.36
PS 500 mg/kg	3.50 ± 0.20	2.22 ± 0.12	2.23 ± 0.20
PS 1000 mg/kg	5.28 ± 0.16***	3.25 ± 0.23*	2.98 ± 0.14*

Data are expressed as mean ± SEM (n = 6) and were analyzed by one-way ANOVA followed by Bonferroni's post hoc test; p* < 0.05, p** < 0.01, and p*** < 0.001 indicate statistically significant, highly significant, and very highly significant differences, respectively, where the normal control group (Group I) was compared with the acetaminophen control (Group II) to confirm hepatotoxicity, and each treatment group—PS 500 mg/kg (Group IV), PS 1000 mg/kg (Group V), and silymarin 100 mg/kg (Group III)—was statistically compared exclusively against the acetaminophen control (Group II) without intergroup comparisons or reference to the normal control.

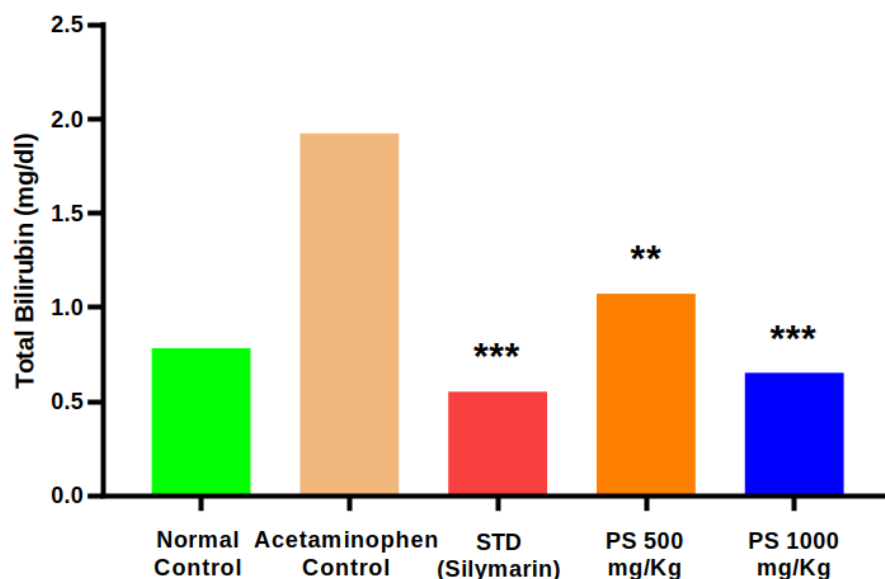


Figure 3: Attenuation of acetaminophen-induced elevations in serum total bilirubin by *P. spicatus* in rats.

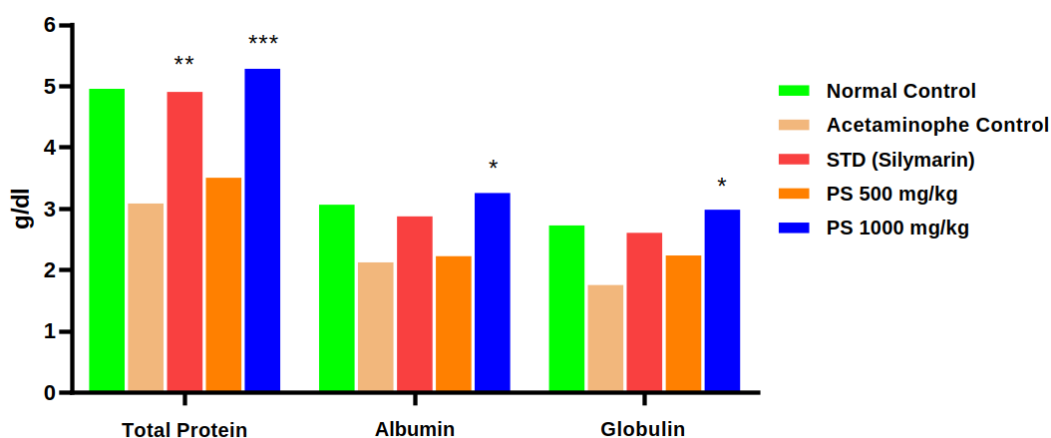


Figure 4: Restoration of acetaminophen-induced impairment of serum total protein, albumin, and globulin levels by *P. spicatus* in rats.

histopathological assessment and additional cholestatic markers (e.g., GGT or bile acids) were not evaluated, any inference regarding selective canalicular or biliary membrane protection should be interpreted cautiously and considered hypothesis-generating rather than definitive. Accordingly, the observed reductions in ALP and bilirubin, in the absence of significant normalization of ALT and AST, suggest selective modulation of hepatobiliary biochemical parameters rather than comprehensive hepatocellular protection under the present experimental conditions [26]. Similar patterns of partial biochemical modulation, characterized by improvement in cholestatic markers with limited transaminase normalization, have been reported in certain plant-derived hepatoprotective studies, suggesting that phytochemical extracts may differentially influence hepatobiliary and hepatocellular components depending on dose, composition, and injury severity [27, 28].

In addition to enzymatic changes, PS treatment at 1000 mg/kg significantly restored serum total protein, albumin, and globulin levels, indicating improved circulating protein levels. Notably, serum protein fractions in the 1000 mg/kg group exceeded values observed

in the normal control group. Such supranormal elevations may reflect biological variability in circulating proteins after acute injury and recovery, adaptive compensatory responses of the liver, acute phase reactant changes, or transient shifts in hydration and plasma volume rather than definitive supraphysiological protein synthesis. Without histopathological correlation and comprehensive biomarker profiling, these elevated protein levels should be interpreted cautiously as exploratory observations rather than conclusive evidence of enhanced synthetic capacity [29, 30]. Since the liver is the primary site of plasma protein synthesis [31], reductions in these parameters are indicative of impaired hepatocellular function in acute hepatic injury [32]. Restoration of protein levels following PS administration suggests enhancement of hepatocyte functional integrity and biosynthetic activity [33].

Nevertheless, because transaminase levels were not significantly normalized and structural assessment was not performed, this improvement should be interpreted as functional modulation rather than confirmed cytoprotective regeneration. Thus, the improvements

in circulating protein fractions likely reflect partial restoration of hepatic synthetic function rather than confirmed structural regeneration of hepatocytes.

Overall, the findings demonstrate that *P. spicatus* root extract ameliorated selected biochemical alterations induced by acetaminophen, particularly ALP and bilirubin elevation, and improved hepatic synthetic parameters at the higher dose. These effects indicate selective biochemical improvement and partial functional restoration, but do not establish comprehensive hepatocellular protection. In addition to statistical significance, the magnitude of change in ALP and bilirubin relative to the acetaminophen control group indicates a biologically meaningful attenuation of hepatobiliary dysfunction. However, effect size estimation would further strengthen interpretation in future studies [34, 35]. Although oxidative stress is central to acetaminophen-induced hepatotoxicity, oxidative stress markers, antioxidant enzyme activity, and histopathological validation were not assessed in this study; therefore, attribution of antioxidant or membrane-stabilizing mechanisms would be premature [36]. Therefore, any mechanistic interpretation regarding antioxidant, anti-inflammatory, or membrane-stabilizing activity should be considered speculative and requires direct biochemical and structural validation [37]. Future studies should incorporate oxidative stress profiling, molecular pathway analyses, and liver histology to clarify the mechanistic basis of the observed effects. Additionally, the relatively high doses employed (500-1000 mg/kg crude extract) necessitate careful consideration of dose optimization, extract standardization, pharmacokinetic characterization, and safety evaluation before translational extrapolation can be justified [38]. Inclusion of extract-only control groups in future investigations would further help delineate intrinsic hepatic effects independent of acetaminophen-induced injury. Importantly, because an extract-only control group was not included in the present study, the extract's independent biochemical effects in the absence of an acetaminophen challenge cannot be excluded. Another limitation of the present study is that the crude ethanolic extract was not chemically standardized using chromatographic fingerprinting or quantified marker compounds; therefore, potential variability in phytochemical composition may influence reproducibility. In addition, because the sex distribution within individual groups was not recorded, it was not possible to assess whether males and females responded differently to treatment. This remains a potential source of biological variability that could not be addressed in the present study. This represents a design limitation that should be addressed in future investigations.

5. Conclusion

In this acetaminophen-induced acute liver injury model, the ethanolic root extract of *P. spicatus* produced selective biochemical attenuation of hepatobiliary dysfunction, demonstrated by significant reductions in ALP and total bilirubin, as well as restoration of circulating protein fractions at 1000 mg/kg. The absence of significant normalization of ALT and AST indicates that comprehensive hepatocellular protection was not achieved under the present experimental conditions. These findings indicate partial biochemical improvement in hepatobiliary functional markers. Definitive validation will require histological examination, assessment of oxidative stress and inflammatory markers, inclusion of extract-only controls, and standardized pharmacological characterization before translational extrapolation can be made.

Conflicts of Interest

The authors declare no competing interests that could have influenced the objectivity or outcome of this research.

Funding Source

The authors declare that no specific grant or funding was received for this research from any public, commercial, or not-for-profit funding agency.

Acknowledgments

We gratefully acknowledge Professor Masum Shahriar (Department of Pharmacy, Jahangirnagar University, Savar, Dhaka, Bangladesh) and Professor Dr. Md. Iqbal Hossain, Dean, Faculty of Health Sciences, Gono Bishwabidyalay, Savar, Dhaka, Bangladesh, for their academic guidance and institutional support of this work.

Ethical approval

All animal experimental procedures were approved by the Research Ethical Committee under the Center for Multidisciplinary Research, Chikitsa Foundation Bangladesh (Approval No. CF-BD/EC/023), and conducted in accordance with institutional and international guidelines for laboratory animal care.

Large Language Model

None.

Authors' Contributions

KHR, MAS, MZI, SMNI, and MJH conducted the experiments and acquired the data. PKS contributed to the study design and experimental strategy. MS, SSJ, PCR, and AM contributed to manuscript editing. UC conceived the study, oversaw the overall workflow, and finalized the manuscript.

Data Availability

All analyzed data are presented in the manuscript as group-level summary statistics (mean $\pm$ SEM) for each experimental endpoint.

References

1. Wang L, Shao Z, Wang X, Lu W, Sun H. Xenobiotic-induced liver injury: Molecular mechanisms and disease progression. *Ecotoxicol Environ Saf*. 2025;303:118854. [PMID: 40845633, <https://doi.org/10.1016/j.ecoenv.2025.118854>].
2. Jaeschke H, Xie Y, McGill MR. Acetaminophen-induced Liver Injury: from Animal Models to Humans. *J Clin Transl Hepatol*. 2014;2(3):153-61. [PMID: 26355817, PMCID: PMC4521247, <https://doi.org/10.14218/JCTH.2014.00014>].
3. Mazaleuskaya LL, Sangkuhl K, Thorn CF, FitzGerald GA, Altman RB, Klein TE. PharmGKB summary: pathways of acetaminophen metabolism at the therapeutic versus toxic doses. *Pharmacogenet Genomics*. 2015;25(8):416-26. [PMID: 26049587, PMCID: PMC4498995, <https://doi.org/10.1097/FPC.0000000000000150>].
4. Ramachandran A, Jaeschke H. Acetaminophen hepatotoxicity: A mitochondrial perspective. *Adv Pharmacol*. 2019;85:195-219. [PMID: 31307587, PMCID: PMC7350903, <https://doi.org/10.1016/bs.apha.2019.01.007>].
5. Maharaja P, Sengottuvel T, Aarthi A, Gopalsatheeskumar K. Review on Antioxidant and Hepatoprotective activity of Medicinal plants against Paracetamol Induced animal model. *Res J Pharmacognosy and Phytochem*. 2020;12(2):114-9. [<https://doi.org/https://doi.org/10.5958/0975-4385.2020.00020.55>].
6. Wang C, Feng X, Li W, Chen L, Wang X, Lan Y, et al. Apigenin as an emerging hepatoprotective agent: current status and future perspectives. *Front Pharmacol*. 2024;15:1508060. [PMID: 39749193, PMCID: PMC11693974, <https://doi.org/10.3389/fphar.2024.1508060>].

7. Kaplowitz N. Drug-induced liver injury. *Clin Infect Dis*. 2004;38 Suppl 2:S44-8. [PMID: 14986274, <https://doi.org/10.1086/381446>].
8. Maddrey WC. Hepatotoxicity: The adverse effects of drugs and other chemicals on the liver. *Gastroenterology*. 2000;118(5):984-5. [[https://doi.org/10.1016/s0016-5085\(00\)70192-2](https://doi.org/10.1016/s0016-5085(00)70192-2)].
9. El-Saadony MT, Saad AM, Mohammed DM, Korma SA, Alshahrani MY, Ahmed AE, et al. Medicinal plants: bioactive compounds, biological activities, combating multidrug-resistant microorganisms, and human health benefits - a comprehensive review. *Front Immunol*. 2025;16:1491777. [PMID: 40375989, PMCID: PMC12079674, <https://doi.org/10.3389/fimmu.2025.1491777>].
10. Md HA, Shariful I, Mohammad MH, Md HR, Asheka R, Mohammad AT, et al. Studies of phytochemical analysis, in vitro and in vivo evaluation of the local *Pseudoelephantopus spicatus* plant. *African Journal of Pharmacy and Pharmacology*. 2024;18(7):114-26. [<https://doi.org/10.5897/ajpp2023.5354>].
11. Jimenez Usuga NdS, Malafronte N, Durango EJO, Braca A, De Tommasi N. Phytochemical investigation of *Pseudelephantopus spiralis* (Less.) Cronquist. *Phytochemistry Letters*. 2016;15:256-9. [<https://doi.org/10.1016/j.phytol.2016.02.013>].
12. Lin CC, Yen MH, Chiu HF. The pharmacological and pathological studies on Taiwan folk medicine (VI): The effects of *Elephantopus scaber* subsp. *oblanceolata*, *E. mollis* and *Pseudoelephantopus spicatus*. *Am J Chin Med*. 1991;19(1):41-50. [PMID: 1897489, <https://doi.org/10.1142/S0192415X91000065>].
13. Jonsson M, Jestoi M, Nathanail AV, Kokkonen UM, Anttila M, Koivisto P, et al. Application of OECD Guideline 423 in assessing the acute oral toxicity of moniliformin. *Food Chem Toxicol*. 2013;53:27-32. [PMID: 23201451, <https://doi.org/10.1016/j.fct.2012.11.023>].
14. Mitchell JR, Jollow DJ, Potter WZ, Davis DC, Gillette JR, Brodie BB. Acetaminophen-induced hepatic necrosis. I. Role of drug metabolism. *J Pharmacol Exp Ther*;187(1):185-94. [PMID: 4746326].
15. Mossanen JC, Tacke F. Acetaminophen-induced acute liver injury in mice. *Lab Anim*. 2015;49(1 Suppl):30-6. [PMID: 25835736, <https://doi.org/10.1177/0023677215570992>].
16. Li C, Peng J, Hu R, Yan J, Sun Y, Zhang L, et al. Safety and Efficacy of Ketamine Versus Ketamine-Fentanyl-Dexmedetomidine Combination for Anesthesia and Analgesia in Rats. *Dose Response*. 2019;17(1):1559325819825902. [PMID: 30792614, PMCID: PMC6376518, <https://doi.org/10.1177/1559325819825902>].
17. Yang J, Li W, Duan M, Zhou Z, Lin N, Wang Z, et al. Large dose ketamine inhibits lipopolysaccharide-induced acute lung injury in rats. *Inflamm Res*. 2005;54(3):133-7. [PMID: 15883747, <https://doi.org/10.1007/s00011-004-1334-5>].
18. Idoh K, Agbonon A, Potchoo Y, Gbeassor M. Toxicological assessment of the hydroethanolic leaf extract of *clerodendrum capitatum* in Wistar rats. *Pan Afr Med J*. 2016;24:66. [PMID: 27642406, PMCID: PMC5012807, <https://doi.org/10.11604/pamj.2016.24.66.8771>].
19. Anjum R, Raza C. Hepatoprotective effect of D-carvone against rotenone-induced liver toxicity in Swiss Albino mice. *The Journal of Basic and Applied Zoology*. 2026;87(1):4. [<https://doi.org/10.1186/s41936-026-00543-z>].
20. Ramachandran A, Jaeschke H. Mechanisms of acetaminophen hepatotoxicity and their translation to the human pathophysiology. *J Clin Transl Res*. 2017;3(Suppl 1):157-69. [PMID: 28670625, PMCID: PMC5489132, <https://doi.org/10.18053/jctres.03.2017S1.002>].
21. Dahlin DC, Miwa GT, Lu AY, Nelson SD. N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen. *Proc Natl Acad Sci U S A*. 1984;81(5):1327-31. [PMID: 6424115, PMCID: PMC344826, <https://doi.org/10.1073/pnas.81.5.1327>].
22. Jollow DJ, Mitchell JR, Potter WZ, Davis DC, Gillette JR, Brodie BB. Acetaminophen-induced hepatic necrosis. II. Role of covalent binding in vivo. *J Pharmacol Exp Ther*;187(1):195-202. [PMID: 4746327].
23. Moyer AM, Fridley BL, Jenkins GD, Batzler AJ, Pelleymounter LL, Kalari KR, et al. Acetaminophen-NAPQI hepatotoxicity: a cell line model system genome-wide association study. *Toxicol Sci*. 2011;120(1):33-41. [PMID: 21177773, PMCID: PMC3044203, <https://doi.org/10.1093/toxsci/kfq375>].
24. West M, Zimmerman HJ. Serum enzymes in hepatic disease. *Med Clin North Am*. 1959;43(2):371-400. [PMID: 13643126, [https://doi.org/10.1016/s0025-7125\(16\)34158-x](https://doi.org/10.1016/s0025-7125(16)34158-x)].
25. Zimmerman HJ. Hepatotoxicity. *Dis Mon*;39(10):675-787. [PMID: 8223092].
26. Giannini EG, Testa R, Savarino V. Liver enzyme alteration: a guide for clinicians. *CMAJ*. 2005;172(3):367-79. [PMID: 15684121, PMCID: PMC545762, <https://doi.org/10.1503/cmaj.1040752>].
27. Amresh G, Reddy GD, Rao Ch V, Singh PN. Evaluation of anti-inflammatory activity of *Cissampelos pareira* root in rats. *J Ethnopharmacol*. 2007;110(3):526-31. [PMID: 17097249, <https://doi.org/10.1016/j.jep.2006.10.009>].
28. Sahreen S, Khan MR, Khan RA. Hepatoprotective effects of methanol extract of *Carissa opaca* leaves on CCl4-induced damage in rat. *BMC Complement Altern Med*. 2011;11(1):48. [PMID: 21699742, PMCID: PMC3141600, <https://doi.org/10.1186/1472-6882-11-48>].
29. Fu L, Guldiken N, Remih K, Karl AS, Preisinger C, Strnad P. Serum/Plasma Proteome in Non-Malignant Liver Disease. *Int J Mol Sci*. 2024;25(4). [PMID: 38396688, PMCID: PMC10889128, <https://doi.org/10.3390/ijms25042008>].
30. Wang S, Zhao M, Su Z, Yang D, Mu R. Annual biological variation and personalized reference intervals of 8 serum liver enzymes and 3 noninvasive tests in fatty liver patients. *Clin Chim Acta*. 2025;569:120156. [PMID: 39870293, <https://doi.org/10.1016/j.cca.2025.120156>].
31. Carvalho JR, Verdelho Machado M. New Insights About Albumin and Liver Disease. *Ann Hepatol*. 2018;17(4):547-60. [PMID: 29893696, <https://doi.org/10.5604/01.3001.0012.0916>].
32. Tavill AS. The synthesis and degradation of liver-produced proteins. *Gut*. 1972;13(3):225-41. [PMID: 4114563, PMCID: PMC1412132, <https://doi.org/10.1136/gut.13.3.225>].
33. Badylak SF, Van Vleet JF. Sequential morphologic and clinicopathologic alterations in dogs with experimentally induced glucocorticoid hepatopathy. *Am J Vet Res*;42(8):1310-8. [PMID: 6117238].
34. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev Camb Philos Soc*. 2007;82(4):591-605. [PMID: 17944619, <https://doi.org/10.1111/j.1469-185X.2007.00027.x>].
35. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129-33. [<https://doi.org/10.1080/00031305.2016.1154108>].
36. Jaeschke H, McGill MR, Ramachandran A. Oxidant stress, mitochondria, and cell death mechanisms in drug-induced liver injury: lessons learned from acetaminophen hepatotoxicity. *Drug Metab Rev*. 2012;44(1):88-106. [PMID: 22229890, PMCID: PMC5319847, <https://doi.org/10.3109/03602532.2011.602688>].
37. Ramaiah SK. A toxicologist guide to the diagnostic interpretation of hepatic biochemical parameters. *Food Chem Toxicol*. 2007;45(9):1551-7. [PMID: 17658209, <https://doi.org/10.1016/j.fct.2007.06.007>].
38. Heinrich M, Jalil B, Abdel-Tawab M, Echeverria J, Kulic Z, McGaw LJ, et al. Best Practice in the chemical characterisation of extracts used in pharmacological and toxicological research-The ConPhyMP-Guidelines. *Front Pharmacol*. 2022;13:953205. [PMID: 36176427, PMCID: PMC9514875, <https://doi.org/10.3389/fphar.2022.953205>].



Review Article

Ocular Trauma Patterns in the Middle East and North Africa: A Scoping Review with Emphasis on Iraq

Elaf Falih Abbas¹, Mohammed Mousa Imran¹, Mustafa Yahya Saleh¹, Huda Omran^{2,*}

1-Ophthalmology Department, AlSader Teaching Hospital, Basra, Iraq

2-Genetic Department, Aldafter Medical Center, Arabian Gulf University, Manama, Bahrain

ARTICLE INFO

Article history:

Received 15 Mar. 2026

Received in revised form 18 Apr. 2026

Accepted 26 Apr. 2026

Published 28 Apr. 2026

Keywords:

Ocular Trauma

War Injuries

Blindness

MENA

Iraq

Scoping Review

ABSTRACT

Background: Ocular injury is one of the leading causes of preventable monocular blindness globally. These injuries have multiple etiologies, with socioeconomic factors and geopolitical instability as the main drivers. A unique combination of these factors can be seen in the MENA region.

Methods: This scoping review used a structured analysis of published research. We searched PubMed and Scopus (2004–2025) using terms for ocular trauma and MENA countries. Two reviewers independently screened titles, abstracts, and full texts against predefined inclusion criteria that accepted original research, reviews, and case series; a third reviewer resolved disagreements. Data were extracted using a standardized form and synthesized descriptively. No formal quality appraisal was performed.

Results: The search identified 918 records, of which 25 were included. Etiological patterns differed markedly across countries. In Iraq, conflict-related mechanisms (blasts, bullets) dominated in military cohorts, alongside civilian motorcycle accidents. In other MENA countries, road traffic accidents and occupational injuries were more frequently reported. Management challenges in Iraq were compounded by limited primary eye care infrastructure and high rates of enucleation. Pediatric trauma was also a major concern. Population-based civilian data from conflict-affected settings remain scarce.

Conclusion: Ocular trauma is a major health issue in Iraq, with patterns suggesting an association with decades of conflict. However, population-based comparative data are very limited. This scoping review highlights evidence gaps, particularly the scarcity of population-based civilian data from conflict-affected settings.

1. Introduction

Ocular trauma is a critical public health problem, accounting for a substantial proportion of monocular blindness worldwide. It significantly affects individual quality of life, reducing their economic productivity and increasing healthcare costs.

The pattern of injury is largely linked to the local environment, culture, industry, and safety standards. The MENA region, with its vast geographic and political diversity, offers a critical lens for examining these relationships. While countries in the Gulf Cooperation Council (GCC) have advanced rapidly in healthcare, including ophthalmic care, others deal with the aftermath of conflict, economic instability, and limited healthcare resources. Iraq stands out within this continuum, having endured decades of war, economic sanctions, and internal conflict. The ophthalmological consequences of this are severe and distinct.

This review aims to evaluate the available scientific literature on ocular trauma in the MENA region by comparing the etiological

factors and management challenges in Iraq with those of neighboring countries. Based on the available evidence, we interpret that prolonged conflict in Iraq is associated with a distinct profile of severe ocular injuries and treatment needs, though population-based comparative data remain limited.

2. Methods

2.1. Study Design and Selection

This review was conducted as a scoping review following the framework of Arksey and O'Malley and reported in accordance with the PRISMA ScR (Preferred Reporting Items for Systematic Reviews and Meta Analyses extension for Scoping Reviews) checklist. A protocol was not registered. We searched PubMed and Scopus for articles published between January 2004 and June 2025. An initial PubMed search identified relevant keywords. These were used in a systematic search strategy that combined free-text and Boolean terms applied to both databases. These search terms for ocular trauma and MENA countries included the following:

- Ocular trauma terms: ("ocular trauma" OR "eye injury" OR "ocular injury" OR "open globe" OR "penetrating eye injury")
- Geographic terms: ("Middle East" OR "North Africa" OR MENA OR Iraq OR Iran OR "Saudi Arabia" OR Jordan OR Palestine OR Gaza OR "United Arab Emirates" OR Qatar OR Kuwait OR Oman OR Bahrain OR "Syrian Arab Republic" OR Lebanon OR Egypt OR Libya OR Tunisia OR Algeria OR Morocco OR Yemen)

*Corresponding author: Huda Omran, Genetic Department, Aldafter Medical Center Arabian Gulf University, Manama, Bahrain. Email: hudamossa.omran@gmail.com

Published and owned by PubPorta Publishing LLC. Academic oversight and scholarly guidance are provided by the American Society for Inclusion, Diversity, and Equity in Healthcare (ASIDE). ISSN (Print) 3067-8730, ISSN (Online) 3067-8749. ©2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0). Hosting by ASIDE Journals.

Citation: Abbas EF, Mousa Imran M, Saleh MY, Omran H. Ocular Trauma Patterns in the Middle East and North Africa: A Scoping Review with Emphasis on Iraq. ASIDE Health Sci. 2026;2(2):40-46, doi:10.71079/ASIDE.HS.042826665

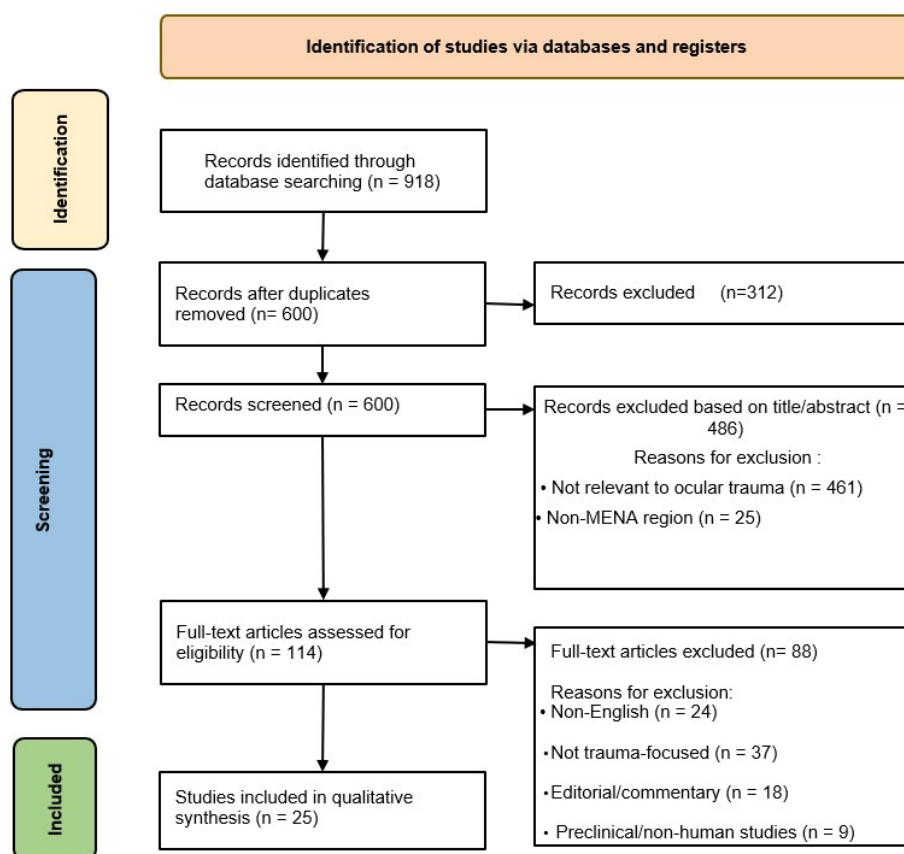


Figure 1: PRISMA flow diagram for the identification and selection of studies eligible for this review.

- Supplementary terms: "war injury," "blast injury," "occupational injury," "road traffic accident."

Two reviewers independently screened titles and abstracts, followed by full-text review. A third reviewer resolved disagreements. The inclusion criteria were original research, reviews, and case series in English reporting on the etiology, epidemiology, or management of ocular trauma in MENA countries. Exclusion criteria: non-English publications, studies not focused on trauma, editorials, preclinical studies, and studies conducted outside the MENA region.

Articles were systematically grouped by their primary country or region of focus, primary mechanism of injury (e.g., blast, road traffic accidents RTA, occupational), and key clinical themes (e.g., pediatric trauma, surgical outcomes, rehabilitation). Data relevant to Iraq were extracted and analyzed in parallel with data from other MENA nations, like Iran, Saudi Arabia, and Jordan, to enable a direct comparative analysis of epidemiological patterns, management protocols, and public health implications.

2.2. Data Extraction and Synthesis

Microsoft Excel was used to manage extracted data. Authors, year, country, study design, population (civilian, military, or mixed), primary injury mechanism, and key clinical findings were collected. Data were grouped by country (**Figure 1**). Given heterogeneity in study designs and population, no meta-analysis was performed. Synthesis is descriptive and comparative. All included studies were summarized in a single table (**Table 1**). Risk of bias assessment was not performed, consistent with scoping review methodology. Findings were presented as descriptive patterns rather than definitive comparative epidemiology.

3. Results

The initial search yielded 918 records. After duplications had been removed, 600 titles and abstracts were screened, and 486 were excluded. The full text of 114 articles was assessed for eligibility. A total of 25 studies from the MENA region met the inclusion criteria and were included in the qualitative synthesis (**Figure 1**). Of the 25 included studies, 11 focused on Iraq (including those that covered both Iraq and Afghanistan), 3 on Iran, 1 on Saudi Arabia, 1 on Jordan, 2 on Palestine/Gaza, 1 on the Levant, 1 on Turkey, and 1 examined the MENA region as a whole. Several studies involved military personnel, while others focused on civilian or pediatric populations. To ensure transparency, the evidence is presented in two tiers: primary epidemiologic studies (cross sectional, cohort, and case series reporting original trauma data) form the basis for all comparative statements about injury mechanisms, severity, and outcomes. Reviews, meta-analyses, and service-oriented papers are cited only for contextual background (e.g., infrastructure or rehabilitation discussions) and are not used for direct epidemiological comparisons.

Epidemiological patterns and etiological comparison The etiology of ocular trauma in MENA is a direct reflection of local realities. The provided literature reveals a clear contrast between conflict-affected zones like Iraq and more stable regions. As shown in Table 1.

While Iraq represents an extreme within the conflict injury spectrum, a gradient of instability and resource scarcity exists across the region.

Table 1: Comparison of Ocular Trauma Epidemiology Across Countries/Regions in the MENA Region (n = 25) Tier 1 (primary epidemiologic) studies are marked with †.

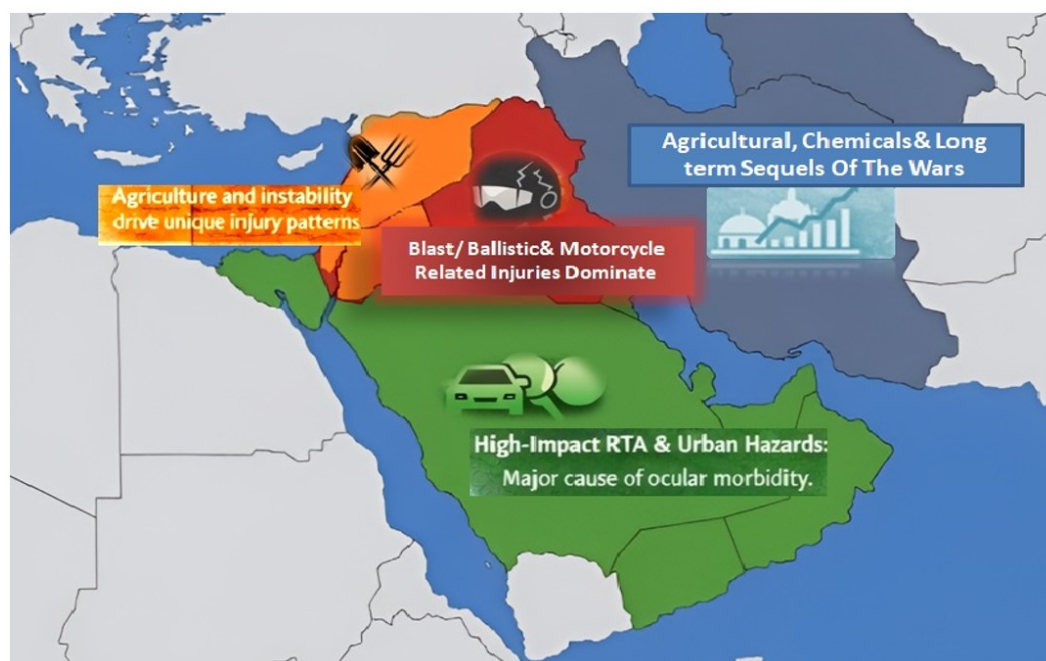
Author(s), Year	Country/Region	Study Design	Sample Size (N)	Study Setting	Denominator / Population at Risk	Population	Primary Mechanism(s)	Key Findings
Sulaiman II et al., 2024	Iraq	Review/met analysis (Tier 2)	688	Military & civilian hospitals	Not applicable	Civilian/military	Ballistic	High enucleation rates; penetrating orbital trauma
Blanch RJ et al., 2011†	Iraq	Case series	63	British military hospital (Iraq/Afghanistan)	All combat eye injuries referred	Military personnel	Blast/ fragmentation	52% open globe
Justin GA et al., 2020†	Iraq	Cohort	652	U.S. military tertiary hospital	All orbital fractures referred	Military personnel	Blast	31.3% NLP
Thach AB et al., 2008†	Iraq	Case series	797 severe eye injuries	U.S. military registry	Combat zone injuries 2003–2005	Military personnel	Blast/ fragmentation	116 enucleations; 438 open globe
Hashim ZA et al., 2024†	Iraq	Cross sectional	335	Iraqi tertiary hospitals	Consecutive motorcycle accident patients	Civilian	Motorcycle accidents	17% severe; lid/corneoscleral injuries
Yulish M et al., 2012†	Levant	Case series	119	Regional hospital	Agricultural workers presenting with eye injury	Agricultural workers	Olive harvesting	Corneal perforations
Hashemi H et al., 2015†	Iran	Cross sectional	5190	Population based	General population aged 40–64	Civilian	Mixed	Lifetime trauma prevalence 8.57% (blunt 3.91%, sharp 3.82%, chemical 1.93%)
Challa NK et al., 2024†	Saudi Arabia	Cross sectional	300 truck drivers	Truck driving company	All active drivers	Civilian	RTA	Refractive error linked to RTA risk
Odat TAM et al., 2014†	Jordan	Case series	21	Tertiary eye center	Consecutive cases	Civilian	Prickly pear spine	Keratoconjunctivitis
Al Barqi M et al., 2022†	Saudi Arabia	Case series	3	Tertiary eye center	Fireworks injuries during celebrations	Civilian	Fireworks	Severe open globe
Al Amry M et al., 2021†	Saudi Arabia	Case series	36	Tertiary eye center, Riyadh	Consecutive terror-related trauma	Civilian	Terror-related (blast)	77.8% NLP; devastating outcomes
Harris JP et al., 2021†	Iraq	Cohort	890 eyes	U.S. military registry	All combat open globe injuries	Military personnel	Blast	32.02% open globe injuries
Shakarchi F., 2025	Iraq	Cross-sectional (service)	2805 primary health centers	National primary care survey	All PHCs in Iraq	Civilian	Mixed (context)	Only 2%(n=63) offer eye services
Bahremani E et al., 2023	MENA region	Cross-sectional (burden study)	Not specified	Global Burden of Disease	MENA population	Civilian	Mixed (all cause)	Iraq age-standardized prevalence 7188.6/100,000
Holmes CJ et al., 2019†	Iraq	Cohort	19	British military registry	All combat eye injuries requiring surgery	Military personnel	Blast/ fragmentation	Outcomes of evisceration/ enucleation
Keenan TDL et al., 2011†	Palestinian territories	Case series	32	Tertiary hospital, Gaza	Consecutive enucleations/ eviscerations	Civilian/ mixed	Conflict related	High rates of enucleation
Ciftci MD et al., 2023	Turkey	Case series (surgical technique)	9	Tertiary eye center	Corneal perforations	Civilian	Corneal perforations	Sandwich technique + amniotic membrane is effective
Arevalo JF et al., 2012	Saudi Arabia	Review (Tier 2)	Not applicable	Not applicable	Not applicable	Civilian	Penetrating trauma	Sympathetic ophthalmia prevention
Hayder Abdul Al Husain et al., 2023†	Iraq	Cross sectional	120 pediatric patients	Pediatric ophthalmology clinic	Consecutive clinic attendees	Pediatric	Mixed	Home-occurring open globe wounds are preventable
Hosseini H et al., 2011†	Iran	Cross sectional	278	Tertiary hospital (southern Iran)	All pediatric ocular trauma	Pediatric	Sharp/blunt	Boys predominantly affected

MENA, Middle East and North Africa; NLP, No Light Perception; PHC, Primary Health Center; RTA, Road Traffic Accident. † Denotes studies included in the primary ocular trauma analysis.

Table 1: Comparison of Ocular Trauma Epidemiology Across Countries/Regions in the MENA Region (n = 25) Tier 1 (primary epidemiologic) studies are marked with †.

Author(s), Year	Country/Region	Study Design	Sample Size (N)	Study Setting	Denominator / Population at Risk	Population	Primary Mechanism(s)	Key Findings
Al Dahan D et al., 2019†	Saudi Arabia	Cross sectional	199 pediatric prosthesis patients	Tertiary referral center	All pediatric prosthesis fittings	Pediatric	Mixed	Prosthesis indications after trauma
Masrur A et al., 2025†	Gaza (Palestine)	Observation cohort	23	4 hospitals in Gaza	Consecutive CNN registry cases (March–Sept 2024)	Pediatric (<20 years)	Explosive / penetrating (conflict-related)	11/23 (48%) ocular trauma; severe visual loss including NLP; communication barriers main challenge
Mousavi B et al., 2009	Iran	Cross-sectional (chemical warfare)	147	Chemical warfare survivor's registry	Iranian chemical warfare survivors	Civilian	Chemical weapons	Reduced quality of life
Geiling J et al., 2012	Iraq	Review (Tier 2)	Not applicable	Not applicable	Not applicable	Military veterans	Conflict-related	Long-term care costs
Ramadhan HA et al., 2025†	Iraq	Cross-sectional	71	Halabja population survey	Chemically-exposed population	Civilian	Chemical weapons	Corneal injuries

MENA, Middle East and North Africa; NLP, No Light Perception; PHC, Primary Health Center; RTA, Road Traffic Accident. † Denotes studies included in the primary ocular trauma analysis.

**Figure 2:** Schematic illustration of key etiological factors and patterns of ocular trauma.

The data from Iraq is affected by the consequences of modern conflicts. Studies on military personnel from Operation Iraqi Freedom detail high rates of open globe wounds from blast and fragmentation injuries [1, 2]. A study by Justin and his coworkers of U.S. personnel evacuated to a tertiary military hospital from Iraq (2001 – 2011) found that orbital fractures occurred in 34.2% (304/890) of eye injuries, with 31.3% (95/304) of those cases resulting in no light perception [3]. Earlier data from the 2003 – 2005 periods documented 797 severe eye injuries, which included 438 open globe injuries (49 bilateral) and 116 eyes requiring removal

(enucleation/evisceration) [4]. Concurrently, civilian studies point to motorcycle accidents as a major cause [5] of ocular morbidity, suggesting a dual burden of conflict-related and conventional trauma. Hashim and his collaborators documented that Iraqi tertiary hospitals reported 335 ophthalmological accidents from motorcycle crashes, predominantly affecting young males (mean age 27.8 years) [5]. Injuries were classified as mild in 55.82%, moderate in 27.16%, and severe in 17.01% of cases, with lid lacerations (60.9%) and corneoscleral injuries (22.1%) being most common [5].

In contrast, studies from other regions highlight different priorities. Occupational hazards are prominent, including industrial accidents and olive-harvesting injuries in the Levant [6]. RTA is a universal cause, detailed in studies from Iran [7] and Saudi Arabia [8]. Unique environmental injuries include prickly pear spine keratoconjunctivitis in Jordan [9] and fireworks injuries during celebrations in Saudi Arabia [10]. Terror related ocular trauma is reported from a tertiary center in Saudi Arabia [11], indicating the high rate of ocular trauma in conflict affected settings; similar patterns are observed in Gaza [12, 13]. Challa and Al Rasheed linked refractive error and abnormal stereopsis among truck drivers in Saudi Arabia to the risk of road traffic accidents, adding depth to the understanding of RTA etiologies [8]. This variation in the etiology of ocular trauma, from high-velocity conflict-related injuries to occupational and environmental mechanisms, directly influences the spectrum of injury severity, clinical presentation, and subsequent management challenges encountered across the region. These relationships between cause and clinical consequence are explored in the following section. The variation in the etiology of ocular trauma is illustrated in (Figure 2).

3.1. Disparities in Clinical Management and Visual Prognosis

The nature of trauma determines its severity. Iraqi data emphasize high energy, devastating injuries often presenting with extensive tissue loss, multiple intraocular foreign bodies, and associated maxillofacial trauma [2, 14]. In Iraq, the capacity to manage this trauma burden is severely constrained. A study revealed that only 2% of Iraq's primary health centers offer eye-related services, which contributes to tertiary hospitals being overwhelmed; more than half of specialist hospital consultations are for conditions that are manageable at the primary level [15]. This systemic gap takes place in the context of a high overall burden of eyesight loss in Iraq (7.4% of Iraqis, all causes combined, including cataract, glaucoma, and trauma [15, 16]), which is provided here only as background on the general eye health challenge.

In other MENA settings, presentations, while serious, may more frequently involve isolated corneal abrasions, lid lacerations, or blunt trauma. (Figure 3) shows the different kinds of trauma reported in many studies included in this review and reflects the variation in clinical courses and, consequently, the treatment needed.

The surgical burden differs accordingly. Evisceration and enucleation rates are significantly reported in the context of military trauma from Iraq [4, 17] and in Palestinian territories [12], reflecting the severity of initial injuries and possibly delayed presentation. Surgical innovation is directed at complex reconstruction, such as managing orbital bullet injuries [2] and repairing corneal perforations using advanced techniques like the "sandwich" and amniotic membrane transplantation [18]. Managing sequelae like traumatic cataract and preventing sympathetic ophthalmia are critical concerns shared across the region [19]. An important and recurrent theme is the high incidence of trauma among children, a vulnerable demographic across the MENA region. Ocular injury is a leading cause of preventable blindness in children all over the world. Studies from Iraq [20], Iran [21], and Saudi Arabia [22] consistently show that children represent a vulnerable demographic, suffering from domestic accidents and stray projectile injuries, often leading to lifelong visual disability and a need for prosthetic fitting.

The demand for pediatric prostheses following enucleation and evisceration was highlighted with this permanent need [13]. The challenges of delivering pediatric ophthalmic care in Gaza highlight how systemic collapse in conflict zones exacerbates trauma outcomes [13]. Furthermore, a study by Bahremani and colleagues confirmed

that, while the prevalence of vision loss in MENA decreased by 11.1% from 1990 to 2019, the absolute number remains high at 32.5 million cases, with Iraq's age-standardized prevalence rate at 7188.6 per 100,000 [16].

Furthermore, research on Iranian chemical warfare survivors provides clear evidence of the significant reduction in quality of life associated with ophthalmologic war injuries [23], a finding with direct relevance to the Iraqi civilian population [24, 25].

3.2. Public Health Implications and Challenges

A distinguished variation exists between the advanced, well-resourced tertiary eye centers in the Gulf countries and the overwhelmed or fragmented systems in conflict zones. The Iraqi healthcare system bears a double burden: managing acute, complex war injuries requiring multi-specialty input and providing lifelong rehabilitation for survivors [2]. This includes not only prosthetic services but also low vision aid, psychological support for disfigurement and vision loss, and surgical management of late complications like glaucoma and retinal detachment. The strain on specialized ophthalmic surgical capacity is huge, Gaza, for example [12]. This disparity affects everything from timely primary repair to access to complex vitreoretinal surgery.

3.3. Etiology Specific Prevention Strategies

For Iraq, the primary preventive strategy is primarily political, focusing on conflict mitigation, mine clearance, and weapons control. Secondary prevention includes disseminating basic ocular first-aid knowledge to first responders and the public, and strengthening emergency trauma systems.

For other countries in the MENA region, effective strategies include enforcing occupational safety laws (e.g., mandatory eye protection in agriculture/industry), strict RTA prevention campaigns, public awareness on domestic hazards (fireworks, hazardous toys, chemicals), and community education on the dangers of traditional eye remedies.

3.4. Limitations and Future Directions

The nature of its source data limits this review. The dataset consists of English-language, predominantly hospital-based studies, which introduces publication bias and underrepresents cases that do not reach tertiary centers. There is an obvious scarcity of population-based incidence and prevalence studies from Iraq.

Future research efforts must prioritize several key directions to address the identified gaps. First, establishing standardized ocular trauma registries in Iraq is essential for capturing accurate national epidemiological data. Building on this foundation, conducting cost-effectiveness analyses of prevention strategies tailored to the Iraqi context, compared with other MENA countries, would help guide efficient resource allocation. Additionally, research is needed to develop and evaluate integrated care models capable of delivering proper ophthalmic trauma and rehabilitation services within resource-constrained, post-conflict primary health systems.

4. Conclusion

The etiology of ocular trauma in the MENA region is a potent indicator of its socioeconomic and political conditions. From workplace accidents to high-velocity traumas, Iraq presents itself as a significant case study in this review. The patterns observed suggest the long-term ophthalmological consequences of decades of conflict manifesting as a high volume of severe, sight-depriving injuries and a legacy of rehabilitation needs. Without population-based data, the

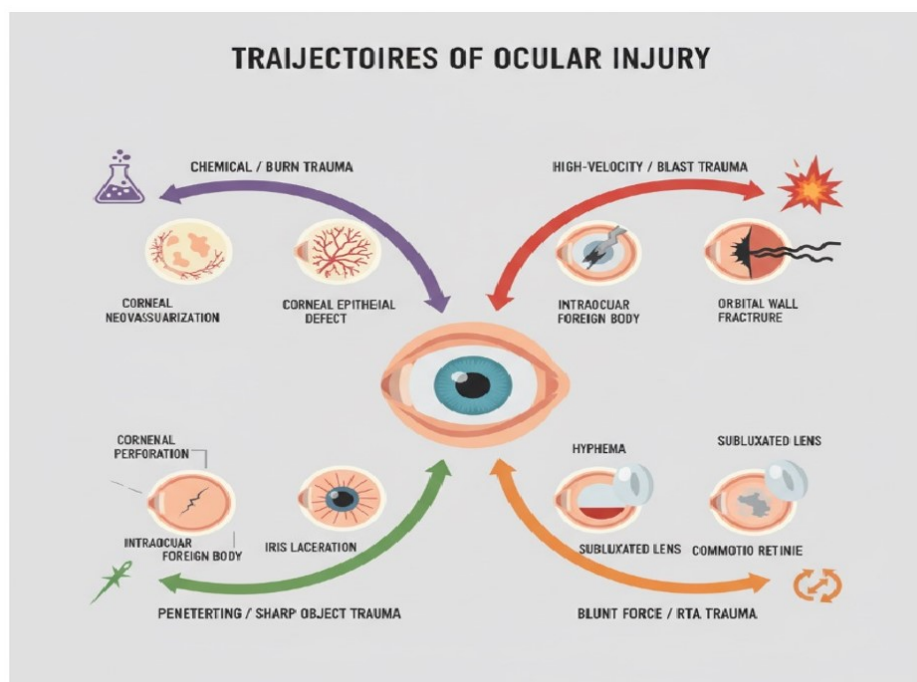


Figure 3: Schematic illustration of trajectories and clinical sequelae of ocular injury.

full clinical burden of ocular trauma in conflict-affected communities remains undetermined.

Conflicts of Interest

The authors declare that they have no known financial, personal, professional, or institutional relationships that could have influenced, or could be perceived to have influenced, the work reported in this manuscript. The authors further confirm that there are no competing interests related to the study design, literature selection, interpretation of findings, manuscript preparation, or decision to submit this article for publication.

Funding Source

The authors declare that no specific grant or funding was received for this research from any public, commercial, or not-for-profit funding agency.

Acknowledgments

None.

Institutional Review Board (IRB)

This study is a scoping review of previously published literature and did not involve direct interaction with human participants, collection of identifiable private information, intervention on patients, or use of unpublished patient-level data. Therefore, institutional review board approval was not required. Informed consent was also not applicable.

Large Language Model

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist in refining English-language writing and enhancing (Figure 2) and (Figure 3). The tool was used solely for

language improvement and visual clarity; AI generated no scientific content, data, or conclusions. The authors take full responsibility for the integrity and accuracy of the manuscript.

Author Contributions

EA contributed to conceptualization, methodology, validation, formal analysis, resources, writing review and editing, and supervision. MI contributed to conceptualization, methodology, validation, formal analysis, resources, writing original draft preparation, writing review and editing, and visualization. MS contributed to methodology, validation, formal analysis, resources, and writing review and editing. HO contributed to conceptualization, validation, formal analysis, resources, writing original draft preparation, writing review and editing, and visualization. All authors have read and agreed to the published version of the manuscript.

Data Availability

No new primary data were generated in this scoping review. All data synthesized in this article were derived from previously published studies identified through PubMed and Scopus searches and from the sources cited in the reference list.

References

1. Blanch RJ, Bindra MS, Jacks AS, Scott RA. Ophthalmic injuries in British Armed Forces in Iraq and Afghanistan. *Eye (Lond)*. 2011;25(2):218-23. [PMID: 21164529, PMCID: PMC3169213, <https://doi.org/10.1038/eye.2010.190>].
2. Sulaiman I, Al-Wassiti ASA, Bani Saad M, Mutar MT, Abdalridha RH, Al-Badri SG, et al. Penetrating orbital trauma: Comprehensive review and meta-analysis of bullet injuries. *Surg Neurol Int*. 2024;15:465. [PMID: 39777188, PMCID: PMC11704445, https://doi.org/10.25259/SNI_632_2024].

3. Justin GA, Turnage WA, Brooks DI, Davies BW, Ryan DS, Eiseman AS, et al. Orbital Fractures and Associated Ocular Injuries in Operation Iraqi Freedom and Operation Enduring Freedom Referred to a Tertiary Care Military Hospital and the Effect on Final Visual Acuity. *Ophthalmic Plast Reconstr Surg*. 2020;36(1):55-60. [PMID: 31567916, <https://doi.org/10.1097/IOP.0000000000001461>].
4. Thach AB, Johnson AJ, Carroll RB, Huchun A, Ainbinder DJ, Stutzman RD, et al. Severe eye injuries in the war in Iraq, 2003-2005. *Ophthalmology*. 2008;115(2):377-82. [PMID: 17904224, <https://doi.org/10.1016/j.ophtha.2007.04.032>].
5. Hashim ZA, Mohammed SK, Abdulla MY, Fawzi HA. Presentations and Incidence of Ocular Injuries Caused by Motorcycle Accidents in Iraq. *F1000Res*. 2024;13:183. [PMID: 40183006, PMCID: PMC11966094, <https://doi.org/10.12688/f1000research.142871.2>].
6. Yulish M, Pikkel J. Olive-harvesting eye injuries. *Middle East Afr J Ophthalmol*. 2012;19(3):320-2. [PMID: 22837627, PMCID: PMC3401803, <https://doi.org/10.4103/0974-9233.97939>].
7. Hashemi H, Khabazkhoob M, Emamian MH, Shariati M, Mohazzab-Torabi S, Fotouhi A. Past History of Ocular Trauma in an Iranian Population-Based Study: Prevalence and its Associated Factors. *Middle East Afr J Ophthalmol*. 2015;22(3):377-82. [PMID: 26180480, PMCID: PMC4502185, <https://doi.org/10.4103/0974-9233.159766>].
8. Challa NK, Alrasheed SH. Refractive error and abnormal stereopsis association with road traffic accidents in Saudi Arabian truck drivers. *Saudi J Ophthalmol*. 2024;38(2):157-62. [PMID: 38988784, PMCID: PMC11232750, https://doi.org/10.4103/sjopt.sjopt_83_23].
9. Odat TA, Al-Tawara MJ, Hammouri EH. Prickly pear spine keratoconjunctivitis. *Middle East Afr J Ophthalmol*. 2014;21(1):61-5. [PMID: 24669148, PMCID: PMC3959044, <https://doi.org/10.4103/0974-9233.124100>].
10. Al Barqi M, Al Ghadeer H. Severe Ocular Trauma Caused by Fireworks Injury: A Case Series. *Middle East Afr J Ophthalmol*. 2022;29(2):103-5. [PMID: 37123429, PMCID: PMC10138128, https://doi.org/10.4103/meajo.meajo_217_22].
11. Al-Amry M, Chaudhry IA, Al-Kahatni E, Al-Ghadeer H. Terror-Related Ocular Trauma in Patients Presenting to a Tertiary Eye Center in the Middle East. *Middle East Afr J Ophthalmol*. 2021;28(4):226-9. [PMID: 35719291, PMCID: PMC9198528, https://doi.org/10.4103/meajo.meajo_316_21].
12. Keenan TD, Sargent NJ. Enucleation and evisceration in the palestinian territories. *Middle East Afr J Ophthalmol*. 2011;18(2):170-2. [PMID: 21731330, PMCID: PMC3119288, <https://doi.org/10.4103/0974-9233.80708>].
13. Masrur A, Yasin YAY, Alhalis EO, Abu Muaileq EAA. Challenges in providing ophthalmic care to children in Gaza. *East Mediterr Health J*. 2025;31(4):262-7. [PMID: 40448491, <https://doi.org/10.26719/2025.31.4.262>].
14. Harris JP, Justin GA, Brooks DI, Woreta FA, Agrawal RV, Ryan DS, et al. Open-globe wounds in operation Iraqi Freedom and Operation Enduring Freedom: risk factors for poor visual outcomes and enucleation. *Acta Ophthalmol*. 2021;99(8):904-8. [PMID: 33565256, <https://doi.org/10.1111/aos.14790>].
15. Shakarchi F. Eye health in post-conflict and climate-vulnerable settings: Iraq as a case study for equitable, integrated primary eye care. *AJO International*. 2025;2(3). [<https://doi.org/10.1016/j.ajoint.2025.100148>].
16. Bahremani E, Alizadeh M, Nejadghaderi SA, Noori M, Sullman MJM, Kolahi AA, et al. The burden of vision loss in the Middle East and North Africa region, 1990-2019. *Arch Public Health*. 2023;81(1):172. [PMID: 37749704, PMCID: PMC10521494, <https://doi.org/10.1186/s13690-023-01188-y>].
17. Holmes CJ, McLaughlin A, Farooq T, Awad J, Murray A, Scott R. Outcomes of ocular evisceration and enucleation in the British Armed Forces from Iraq and Afghanistan. *Eye (Lond)*. 2019;33(11):1748-55. [PMID: 31165770, PMCID: PMC7002516, <https://doi.org/10.1038/s41433-019-0480-5>].
18. Ciftci MD, Selver OB. Sandwich Technique and Amniotic Membrane Transplantation as an Effective Method in Repairing Corneal Perforations. *Middle East Afr J Ophthalmol*. 2023;30(4):246-9. [PMID: 39959590, PMCID: PMC11823530, https://doi.org/10.4103/meajo.meajo_191_23].
19. Arevalo JF, Garcia RA, Al-Dhibi HA, Sanchez JG, Suarez-Tata L. Update on sympathetic ophthalmia. *Middle East Afr J Ophthalmol*. 2012;19(1):13-21. [PMID: 22346110, PMCID: PMC3277011, <https://doi.org/10.4103/0974-9233.92111>].
20. Abdul Al-Redha HAA. Epidemiological Trends of Pediatric Eye Diseases in Iraq: A Cross-Sectional Study. *American Journal of Pediatric Medicine and Health Sciences*. 2023;3.
21. Hosseini H, Masoumpour M, Keshavarz-Fazl F, Razeghinejad MR, Salouti R, Nowroozzadeh MH. Clinical and epidemiologic characteristics of severe childhood ocular injuries in southern iran. *Middle East Afr J Ophthalmol*. 2011;18(2):136-40. [PMID: 21731324, PMCID: PMC3119282, <https://doi.org/10.4103/0974-9233.80702>].
22. Al-Dahan D, Khan AO. Indications for Pediatric Ocular Prosthesis Fitting at a Referral Center in the Middle East. *Middle East Afr J Ophthalmol*. 2019;26(2):107-9. [PMID: 31543669, PMCID: PMC6737787, https://doi.org/10.4103/meajo.MEAJO_274_18].
23. Mousavi B, Soroush MR, Montazeri A. Quality of life in chemical warfare survivors with ophthalmologic injuries: the first results form Iran Chemical Warfare Victims Health Assessment Study. *Health Qual Life Outcomes*. 2009;7:2. [PMID: 19152700, PMCID: PMC2647906, <https://doi.org/10.1186/1477-7525-7-2>].
24. Geiling J, Rosen JM, Edwards RD. Medical costs of war in 2035: long-term care challenges for veterans of Iraq and Afghanistan. *Mil Med*. 2012;177(11):1235-44. [PMID: 23198496, <https://doi.org/10.7205/milmed-d-12-00031>].
25. Ramadhan HA, Mardan RE, Fattah FH, Mustafa AM, Habibullah IJ, Mohammed MS, et al. Prevalence of corneal injuries among Halabja populations exposed to chemical weapons: a cross-sectional study. *Critical Public Health*. 2025;35(1). [<https://doi.org/10.1080/09581596.2025.2487206>].