Volume 117 2/2026

Rivista fondata nel 1901 da Luigi Devoto Issn 0025 - 7818

La Medicina del Lavoro

Organo della Società Italiana di Medicina del Lavoro

Work, Environment & Health

Official Journal of the Italian Society of Occupational Medicine

www.lamedicinadellavoro.it

Poste Italiane s.p.a. - Sped. in A. P. - D.L. 353/2003 (conv. in L. 27/02/2004 n. 46) art. 1, comma 1, DCB Parma - Finito di stampare nel mese di Aprile 2026

Mattioli 1885

Volume 117

Issn 0025-7818

Rivista fondata nel 1901 da Luigi Devoto

elssn 2532-1080

La Medicina del Lavoro

Organo della Società Italiana di Medicina del Lavoro

Work, Environment & Health

Official Journal of the Italian Society of Occupational Medicine

Editorial office
La Medicina del Lavoro
Clinica del Lavoro «L. Devoto»
Via San Barnaba, 8 - 20122 Milano (Italy)
Tel. 02/50320125 - Fax 02/50320103
http://www.lamedicinadellavoro.it
redazione@lamedicinadellavoro.it

Publisher
Mattioli 1885 srl - Casa Editrice
Strada di Lodesana 649/sx, Loc. Vaio - 43036 Fidenza (PR)
Tel. 0524/530383 - Fax 0524/82537
e-mail: edit@mattioli1885.com
www.mattioli1885.com

Pubblicazione bimestrale - Direttore Responsabile: Antonio Mutti
Autorizzazione del Presidente del Tribunale di Milano 10/5/1948 Reg. al N. 47

La Medicina del Lavoro è indicizzata da / La Medicina del Lavoro is indexed in:
PubMed/Medline; Embase/Excerpta Medica; Abstracts on Hygiene; Industrial Hygiene Digest; Securité et Santé au Travail Bit-CIS; Sociedad Iberoamericana de Informaciòn Cientifica (SIIC); Science Citation Index Expanded (SciSearch®); Journal Citation Report/Science Edition; ISI Web of Science; Scopus (Elsevier); Bibliovigilance

La Medicina del Lavoro

Editorial

Med. Lav. 2026; 117 (2): 18996
DOI: 10.23749/mdl.2026.18996

Artificial Intelligence in Occupational Medicine: Upholding Disciplinary Integrity and Editorial Standards

The rapid advancement of Artificial Intelligence (AI) and large language models has generated increasing interest in their potential uses in occupational medicine [1]. Like other journals, we are receiving numerous submissions that assess AI systems using occupational health datasets [2]. This trend has raised important questions about the scope of occupational medicine research and the appropriate role of AI evaluation studies in a specialized occupational health journal. In this editorial, which builds on a recent one [3], we aim to further clarify the criteria we use to evaluate these submissions, distinguish between suitable and unsuitable AI research in our field, and offer guidance for future authors.

At the core of our editorial decisions is a fundamental principle: a manuscript can be technically sound and methodologically rigorous, but if the research question is not primarily an occupational health question, it falls outside the scope of occupational medicine. This distinction is not just an editorial preference; it reflects occupational medicine's identity as a clinical and public health field focused on preventing occupational diseases and safeguarding worker health.

To elucidate, consider two types of research questions that may appear superficially similar. One might ask: "Can AI language models generate reliable and readable responses to questions about an occupational disease?" This is primarily an AI question, focusing on the technical performance of AI systems—their architecture, training, and output quality. The subject matter, being occupational health, does not by itself render it an occupational medicine question and this why the paper has been rejected. By contrast, a question such as: "Can AI assist the standardized radiographic classification of pneumoconioses in occupational health surveillance programs, and what are its limitations and challenges for implementation?" directly addresses a specific occupational health task. It contemplates integrating technology into occupational health practice and recognizes the essential role of occupational health professionals in decision-making, and this is why the manuscript has been published [4].

The distinction is not sharp but is also far from semantic. Occupational medicine is concerned with worker health outcomes, disease prevention, clinical decision-making, and the protection of workers' health. While AI tools may eventually support occupational health practice, assessing whether an AI system can perform a task is categorically different from demonstrating whether its use improves worker health or should be adopted in practice. The former is a technological evaluation; the latter is an occupational health inquiry.

Occupational medicine research typically addresses questions such as: Does this intervention prevent occupational disease? Does this diagnostic approach improve early detection of work-related illnesses? Does this surveillance method more effectively protect worker health than the current practice? Does this clinical management strategy improve outcomes for exposed workers? How can occupational health professionals enhance the protection of worker welfare? These questions are intrinsically about health outcomes and require examination of clinical workflows, integration of new tools, and occupational health practices.

Conversely, AI evaluation questions generally ask: Can this AI system complete a task? How does it compare to other AI models? What are the accuracy, reliability, or readability metrics of AI-generated outputs? How do different AI architectures perform on standard datasets? These questions focus on AI capabilities and limitations—legitimate and essential in AI research—but do not in themselves represent occupational medicine research.

Challenges emerge when AI evaluation studies are presented as occupational medicine research, merely using occupational health data without meaningful engagement with occupational health practice, outcomes, or professionals' needs. In such instances, occupational health serves as a background rather than the research focus. *La Medicina del Lavoro* will consider manuscripts involving AI if they satisfy the following editorial criteria.

(i) First, the research question must fundamentally be an occupational health question. It should address a recognized occupational health task, clinical decision, or surveillance need central to occupational medicine practice. The task must be specific to occupational health—such as diagnostic classification based on standardized occupational frameworks, hazard assessment using occupational health criteria, or disease surveillance employing occupational epidemiology methods—not generic medical applications incidentally involving occupational data. The study should be firmly situated within the context of occupational health practice and professional needs, rather than merely using data as a convenient AI testing ground.

(ii) Second, the study must demonstrate a direct, rather than theoretical, connection to occupational health practice. The AI application should address a specific, standardized task utilized in occupational medicine—for example, the International Labour Organization classification of pneumoconiosis radiographs, a globally recognized occupational health diagnostic framework. Manuscripts should discuss integration into occupational health workflows and recognize occupational health professionals as primary users or beneficiaries. Studies limited to evaluating AI responses to general questions about occupational diseases, without linking them to specific tasks or practice contexts, do not fulfill this requirement.

(iii) Third, authors must explicitly acknowledge the limitations of AI systems for occupational health applications and clearly articulate that AI is a supportive tool—not a substitute for clinical judgment or professional responsibility [5]. Manuscripts should recognize current occupational health practice and regulatory frameworks, distinguish technical performance metrics from clinical utility, and indicate whether the study is exploratory or pilot in nature. Authors must clarify that AI is not proposed for autonomous clinical decision-making and discuss how occupational physicians would integrate such tools into practice. Affirming the central role of clinical judgment and professional responsibility is essential.

(iv) Fourth, methodological rigor is imperative. Studies must use standardized, validated reference datasets or classifications (e.g., the NIOSH B Reader dataset), employ appropriate statistical analyses, including confidence intervals and significance testing, and compare AI applications to existing standards or established occupational health practices. Transparent reporting of model architectures, training data, and prompting strategies is especially important given AI's technical complexity, which can obscure methodological limitations if not carefully addressed [6].

(v) Fifth, while pilot studies are acceptable, manuscripts should convincingly propose pathways by which the AI application could enhance occupational health outcomes or explain why it might not. The study should describe the occupational health issues being addressed, detail how implementation would happen in occupational settings, identify barriers and facilitators to adoption, and compare the AI approach to current standard practices. Pilot studies should place their findings within a broader research agenda aimed at advancing occupational health practice.

To illustrate the application of these criteria, consider a recent study that evaluated generative multimodal AI models in applying the ILO classification to pneumoconiosis radiographs [4]. This work directly addresses a standardized occupational health task central to surveillance programs worldwide, uses a recognized diagnostic framework, situates the study within occupational health practice, and explicitly acknowledges AI limitations and the supportive rather than autonomous role of AI tools [5]. The methodology employed standardized datasets with certified classifications and appropriate statistical analysis. The manuscript discussed potential occupational health benefits, such as standardization and efficiency improvements, while appropriately framing the work as exploratory and emphasizing the need for prospective validation. This study meets our criteria for publication due to its clear disciplinary relevance, methodological rigor, and focus

on worker health outcomes, though AI assessments are based on an outdated imaging technique applied to a disease that is fortunately in sharp decline. An interesting aspect of the study is that the comparison among four AI software tools is not intended to produce performance rankings, but rather to confirm that none of them achieves a level of accuracy sufficient to replace a NIOSH B Reader, who remains necessary to assume diagnostic responsibility.

Authors considering submissions involving AI are urged to carefully evaluate whether their research question is fundamentally an occupational health inquiry rather than an AI performance evaluation. The research should demonstrate direct relevance to occupational health practice and address specific tasks or decisions routinely made by occupational health professionals. Explicit acknowledgment of AI system limitations in occupational health contexts and clear articulation of the role of professional judgment in clinical decision-making are required. Methodological rigor must be maintained, and manuscripts should discuss potential occupational health outcomes or clinical utility, even for exploratory studies.

Manuscripts that primarily use occupational health data as AI test datasets, without addressing specific occupational health questions or demonstrating direct relevance to practice, will be desk-rejected. We encourage authors to consult the journal's scope and these criteria prior to submission. *La Medicina del Lavoro* welcomes rigorous, discipline-grounded research on AI applications in occupational health. We recognize AI's potential to meaningfully support occupational health practice, but we uphold clear editorial standards to ensure that published research advances occupational medicine and worker protection rather than repurposing occupational data solely for AI evaluation.

The distinction between "AI evaluation using occupational health data" and "occupational medicine research involving AI" is not a semantic nuance. It reflects the fundamental mission of our discipline and journal: advancing occupational health, preventing occupational disease, and protecting worker welfare. Occupational medicine is inherently a clinical and public health field that must remain grounded in occupational health questions, professional practice, and worker protection. AI is a powerful tool in the service of this mission—not an end in itself. Maintaining these boundaries will help guide authors and reviewers toward contributions that meaningfully advance occupational medicine amidst rapid technological evolution, ensuring the journal remains a dedicated venue for occupational health research rather than a general AI technology forum.

Antonio Mutti

References

1. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29(8):1930–1940. Doi: 10.1038/s41591-023-02448-8
2. Chaudhry ZS, Choudhury A. Clinical Applications of Artificial Intelligence in Occupational Health: A Systematic Literature Review. *J Occup Environ Med.* 2024;66(12):943–955. Doi: 10.1097/JOM.0000000000003212
3. Mutti A. Where Are We Going by Applying AI Tools? Are We Going to Lose Our Jobs?. *Med Lav.* 2025;116(6):18255.
4. Baldassarre A, Padovan M, Palla A et al. Generative multimodal AI-based ILO Classification of Radiographs of Pneumoconioses. *Med Lav.* 2026;117(2):18371. Doi: 10.23749/mdl.2026.18371
5. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019;25(1):44–56. Doi: 10.1038/s41591-018-0300-7
6. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. Doi: 10.1136/bmj-2023-078378

La Medicina del Lavoro

Reviews, Commentaries, Perpectives

Med. Lav. 2026; 117 (2): 17996
DOI: 10.23749/mdl.2026.17996

The Micronucleus Test for Occupational Safety†

Ion Udroiu[1,*], Claudia Giliberti[2], Antonella Sgura[1]

[1] Dipartimento di Scienze, Università degli Studi Roma Tre, Rome, Italy
[2] INAIL, Rome, Italy

Keywords: Biomonitoring; Carcinogen; Early Damage; Erythrocyte; Howell-Jolly Body; Mutagenicity; Reticulocyte; Risk Assessment; Work

Summary

Although less commonly used than internal dose indicators, biological effect indicators can be valuable for Risk Assessment. Among the numerous biomarkers used to date, those that indicate DNA damage could be especially useful for occupational safety, as they predict the risk of carcinogenesis. The most common among these assays is the micronucleus test. Unfortunately, this test cannot be performed on human erythrocytes; to apply it to lymphocytes, it requires invasive blood sampling, limiting its use to scientific research. We have developed a new method that enables the micronucleus test to be conducted in a rapid, non-invasive, and cost-effective manner. The test is performed on immature erythrocytes (reticulocytes) in human blood smears collected via a finger prick and stained with Acridine Orange. This new protocol allows the micronucleus test to be applied to human blood samples in a manner compatible with occupational safety procedures.

1. Introduction

Biological monitoring enhances exposure assessment by incorporating fundamental measures of absorbed dose and biological effects. In Italian legislation, Legislative Decree 81/08 mandates, within the framework of risk assessment activities, not only the evaluation of exposure to chemical and physical agents but also the conduct of biological monitoring. This is compulsory for workers exposed to agents for which biological limit values have been established. In addition to "exposure monitoring", Legislative Decree 81/08 also specifies procedures and reference values for "monitoring of the absorbed dose" (Annex XXXIX). As previously noted, however, biomonitoring can encompass both the use of "internal dose indicators" and "biological effect indicators." The latter are referenced in Title IX of Legislative Decree 81/08 but lack detailed procedural or quantitative guidance. Integrating biological effect indicators into occupational health surveillance could enable the detection of early damage, allowing for preventive or protective interventions prior to the onset of occupational pathologies.

Although, as already mentioned, "biological effect indicators are not yet implemented in routine health surveillance activities, they have been extensively developed within scientific research. Among the numerous biomarkers employed, those that indicate DNA damage may be particularly relevant to workplace safety, given their predictive value for carcinogenesis [1]. Most existing literature on human biomonitoring evaluates biomarkers of genotoxic damage—i.e., DNA damage—using peripheral blood samples [2]. This approach is favored due to the lower invasive nature of blood collection

Received 26.09.2025 – Accepted 20.03.2026
* Corresponding Author: Ion Udroiu; E-mail: ion.udroiu@uniroma3.it
† A preliminary version of this work has been presented at the 13th International Symposium on Biological Monitoring (ISBM13) in Occupational and Environmental Health, held in Milan from 9th to 12th September 2025.

compared to tissue biopsies. The tests employed include detection of chromosomal aberrations, sister chromatid exchanges, DNA repair proteins, and the comet assay and micronucleus test [3].

However, these analyses often require specialized skills and resources that are not readily available outside research settings. They require non-negligible blood volumes for lymphocyte culture, preparation of metaphases for chromosomal aberration and sister chromatid exchange assessments, immunofluorescence techniques for DNA repair proteins, and gel electrophoresis for the comet assay. Consequently, although these tests are widely used in scientific studies—including investigations of cohorts with occupational exposure and significant associations between exposure and DNA damage—there is no application outside the field of scientific research. Despite the extensive scientific literature on the subject, these test are not currently employed by health surveillance or occupational safety practitioners. Nonetheless, developing accessible, non-invasive, and cost-effective biomonitoring methods could substantially advance occupational health practices.

Among these, the micronucleus test represents the most straightforward and cheapest assay for genotoxicity assessment. When performed on erythrocytes (red blood cells), it obviates the need for blood culture—necessary for lymphocyte-based tests—and can be conducted on blood smears. The micronucleus test consists in assessing the frequency of cells containing micronuclei, known in hematology as Howell-Jolly bodies. Micronuclei form when a chromosome fragment or an entire chromosome (Figure 1A) is not incorporated into one of the daughter nuclei at the end of cell division, and is therefore the product of genotoxic damage.

Therefore, it is feasible (at least in theory) in every proliferating tissue. In practice, both in human and animal studies, it is often performed on red blood cells (erythrocytes). This is due not only to the ease with which blood can be obtained, but also because, being erythrocytes enucleated (and thus containing no DNA), any nuclear material inside of them (i.e., micronuclei) can be easily recognized. In fact, during the transition from erythroblast (precursor) to erythrocyte, the nucleus is expelled, but not the possible micronucleus (Figure 1B). For these reasons, as well as for its proven sensitivity and specificity, the micronucleus test in animals (in vivo) is internationally standardized (OECD n°474) and recognized for the evaluation of induced DNA damage [4].

In certain species, however, micronuclei are extruded from erythrocytes during their passage through the spleen [5]. Consequently, the micronucleus test is not applicable in these species. This limitation prompted the Organization for Economic Cooperation and Development (OECD) to specify that "any appropriate mammalian species may be used provided it is a species in which the spleen does not remove micronucleated erythrocytes" [4]. Unfortunately, human spleens do remove micronuclei. Nonetheless, the test is feasible if only immature erythrocytes (reticulocytes) are scored, because reticulocytes have not yet passed through the spleen (where they develop into mature erythrocytes) and thus can be scored.

As noted above, many articles in the scientific literature describe occupational biomonitoring studies in which genotoxicity assays, including the micronucleus test (on lymphocytes), have been performed [3]. Nonetheless, they are not used by safety and prevention personnel and are not part of the risk assessment or health surveillance protocols. This could be mainly ascribed to three reasons: 1) invasiveness of the sampling; 2) technical skills required; 3) cost and time required for the assays. In this article, we describe a new protocol we have developed to address and resolve these issues. As a first validation, we used blood samples from X-irradiated mice to assess the method's sensitivity.

2. Methods

First, we introduce a simple and minimally invasive sampling technique. Most assays typically require several milliliters of blood, usually 5 ml, thus obtained via venipuncture. In contrast, our novel approach involves collecting a blood drop through digital puncture, specifically with a finger prick (Figure 2A-2B). While this volume is insufficient for other assays, such as those necessitating lymphocyte culture preparation, it is more than adequate for conducting the micronucleus test on reticulocytes. Indeed, human blood contains approximately

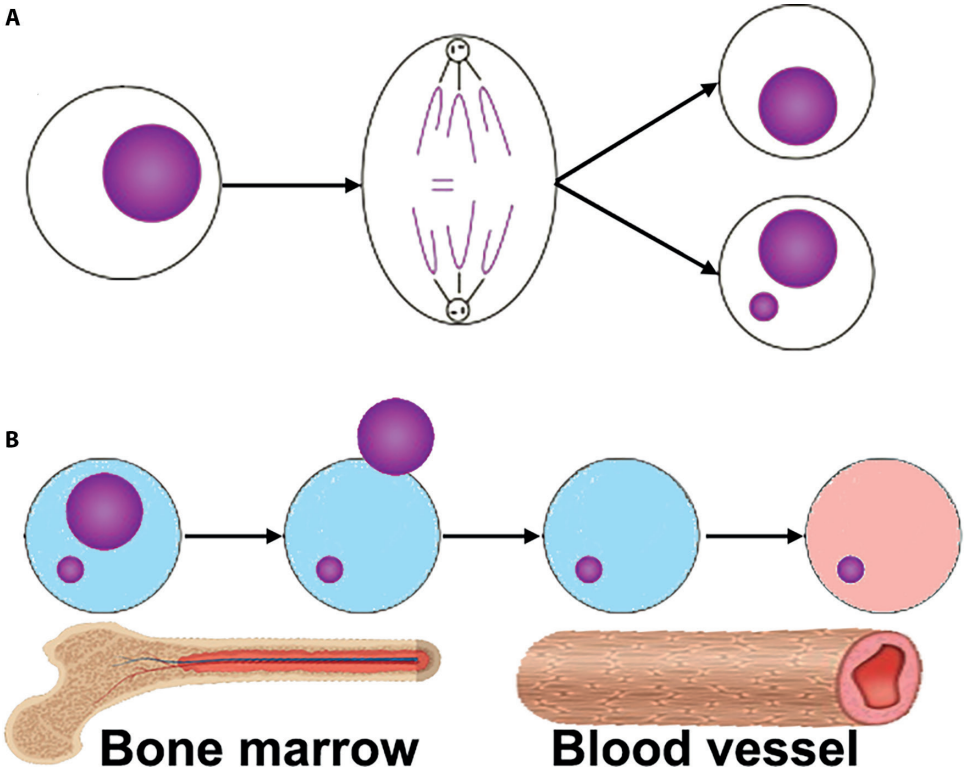

Figure 1. (**a**) During mitosis a chromosome fragment is not segregated in the two nuclei of the daughter cells, thus forming a micronucleus; (**b**) During the maturation of a micronucleated erythroblast in the bone marrow, the nucleus is expelled, resulting in a micronucleated reticulocyte released in the circulation; this will further lose all RNA and organelles, becoming a mature (and in this case micronucleated) erythrocyte.

5 million erythrocytes per microliter [6], with reticulocytes constituting nearly 1% of these cells. This corresponds to tens of thousands of reticulocytes per microliter, whereas the test requires only 1000-2000 reticulocytes for scoring. A blood drop obtained by finger prick typically ranges from 30 to 50 microliters. We recommend using no more than 10 microliters of blood (see below).

For each subject, three to four blood smears can be prepared (Figures 2C-2F) to ensure technical replicates for the staining procedure. To prevent the formation of overly dense smears with piled erythrocytes that are non-scorable, it is recommended to use no more than 10 microliters of blood. The smears are subsequently allowed to dry at room temperature and then fixed in absolute methanol for 10 minutes. Once dried, they may be stored at room temperature.

A stock solution of Acridine Orange (1 mg/mL) in deionized water should be prepared and stored at +4 °C in the dark. For staining, fresh solutions are prepared by diluting aliquots of the stock solution in Sørensen's buffer at pH 6.8 to a final concentration of 20 µg/mL. Blood smears are stained with 50 microliters of this freshly prepared Acridine Orange solution at room temperature, then covered with a coverslip.

Acridine Orange binds to both DNA and RNA, emitting yellow-green fluorescence (520 nm) and red-orange fluorescence (680 nm), respectively. Thus, reticulocytes, which contain RNA, are stained in red/orange, whereas mature erythrocytes are not stained. The test is performed using a fluorescent microscope by scoring only reticulocytes (orange-stained cells). Nucleated blood cells, identifiable by their size and the yellow-stained nucleus occupying

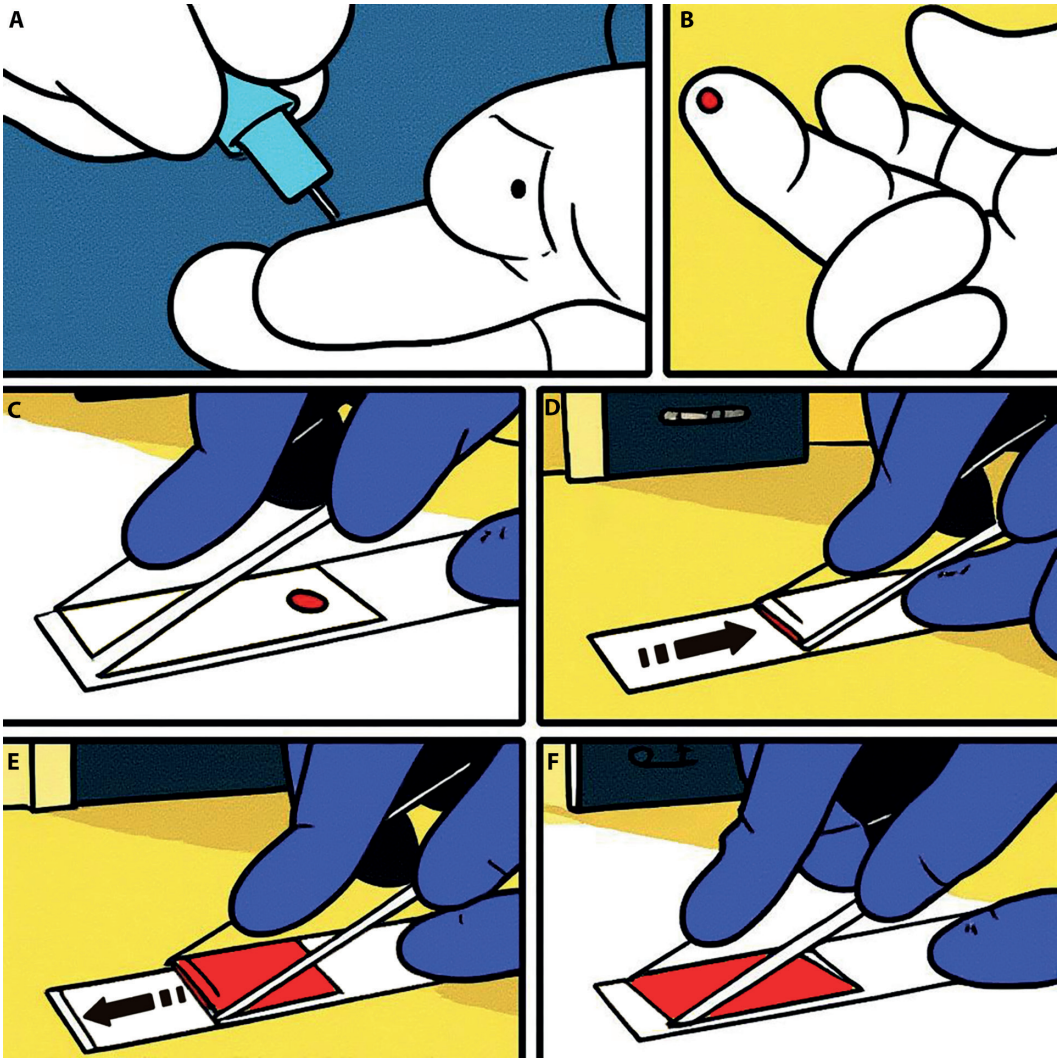

Figure 2. Non-invasive blood sampling. A finger prick (**a**) is used to collect a drop of blood (**b**), which is put on a smear glass (**c**). Another glass is put at 45° and pulled backward until it touches the drop (**d**), then it is quickly moved forward (**e**), creating the blood smear (**f**).

most of the cell, must not be scored. Micronuclei are classified as such if they are stained yellow, are clearly situated within reticulocytes, and are circular (Figure 3). Micronuclei that appear oblong or irregular must be discarded as debris or aggregates.

The assessment of micronuclei frequency then consists of scoring 1500 reticulocytes and noting how many are micronucleated. Summing up, the proposed protocol is the following:

1. Sample a drop of blood with a finger prick.
2. Smear blood on glass.
3. Air-dry the smears at room temperature.
4. Fix the smears in absolute methanol at room temperature for 10 minutes, then air-dry them (and, if needed, store the smears at room temperature).
5. Stain the smears with 50 μL of Acridine Orange in Sørensen's buffer at pH 6.8 (final concentration of 20 μg/mL) (remember Acridine Orange must be at room temperature!);
6. Close the smears with coverslips.
7. Analyze the sample with a fluorescent microscope, scoring 1500 reticulocytes.

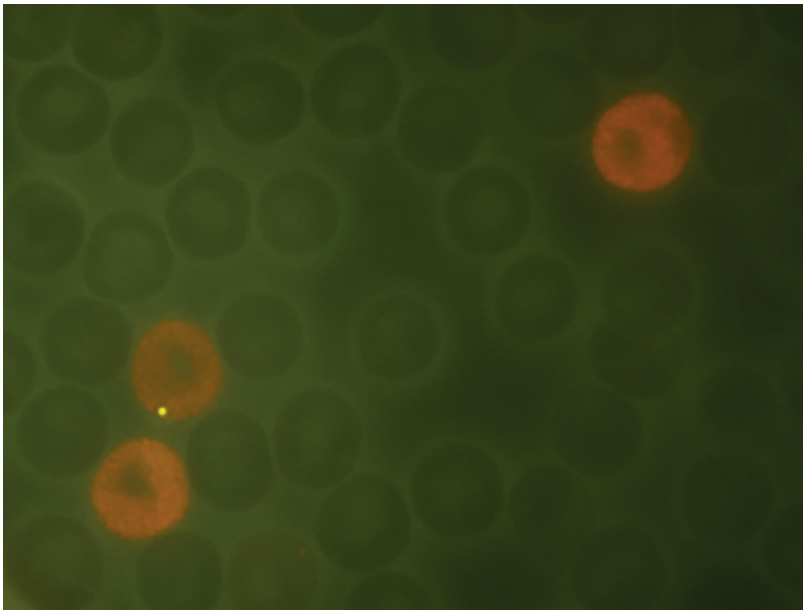

Figure 3. The erythrocyte micronucleus test with Acridine Orange staining. Mature erythrocytes (barely visible) are not stained and must not be scored. Reticulocytes are stained in orange and one of them bears a yellow-stained micronucleus.

As an initial validation of the method's feasibility and sensitivity, the erythrocyte micronucleus test coupled with Acridine Orange staining was performed on murine blood samples. These samples were derived from a prior experiment [7], approved by the Animal Research Ethical Committee of the Italian Ministry of Health (approval ID: 10.10.15). In brief, two groups of five female CD-1 Swiss mice each were subjected to X-irradiation at doses of 0.1 and 1 Gy, using a Gilardoni apparatus (Gilardoni, Italy) with 250 kV, 6 mA, 3 mm Al filter, and at a dose rate of 0.5 Gy/min. An additional group of five female mice served as unirradiated controls. Blood samples were collected via tail puncture 24 hours post-irradiation. The blood smears were fixed, stained, and analyzed as previously described. Micronucleus frequencies between the irradiated and unirradiated groups were compared using Welch's t test. Results are expressed as mean ± standard deviation.

3. Results

The mean frequency of micronucleated reticulocytes in blood samples from unirradiated mice was 1.5 ± 0.9‰ (Figure 4). In contrast, irradiated mice exhibited a mean micronucleated reticulocyte frequency of 5.6 ± 1.6‰ in the 0.1 Gy group and 23.6 ± 6.4‰ in the 1 Gy group.

These findings are consistent with results from previous studies that used Giemsa staining to evaluate micronuclei frequency in the blood of X-irradiated mice [8, 9]. Statistically significant differences were observed between the unirradiated group and the 0.1 Gy group ($p=0.0021$), as well as between the unirradiated group and the 1 Gy group ($p=0.0013$).

4. Discussion

The protocol presented herein is part of an ongoing research project funded by INAIL, the Italian National Institute for Insurance against Accidents at Work (grant number BRIC 2022 ID 54), aimed at facilitating the application of the micronucleus test within the field of occupational safety and preventive measures. During the initial validation phase, blood samples from X-irradiated mice served as positive controls to establish the sensitivity of the method. Both our findings and those of previous investigations [9] confirm that the micronucleus test

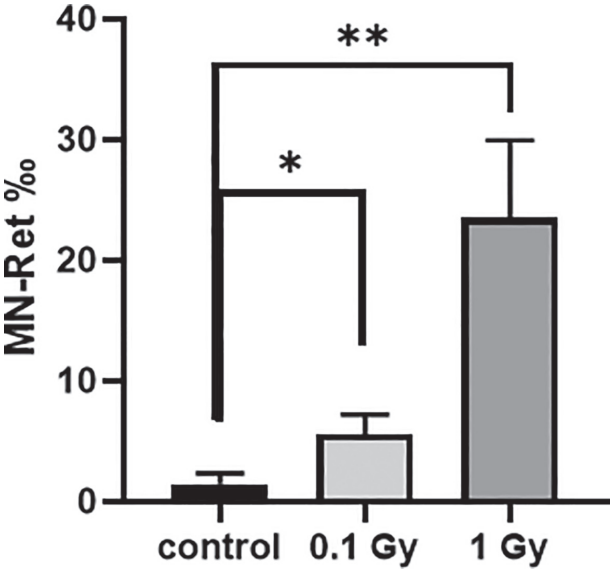

Figure 4. Frequencies of micronucleated reticulocytes (MN-Ret) in peripheral blood from unexposed (control), 0.1 Gy and 1 Gy X-irradiated mice. *: p=0.0021; **: p=0.0013.

on reticulocytes can detect genotoxic damage at low X-ray doses (0.1 Gy). Consequently, this assay is applicable in occupational safety contexts to identify exposure to various genotoxic agents beyond X-ray radiation, especially when damage levels are anticipated to be low or moderate. Although capable of detecting higher levels of genotoxicity typically associated with acute exposures, such as post-accident scenarios requiring immediate intervention rather than prevention, the test's primary utility lies in low-level exposure assessment.

The subsequent phase involves applying the test to blood samples from patients treated with iodine-131. Similar to the current study, a positive control group exposed to a known genotoxic agent will be included; however, this group will comprise human subjects. Based on our preliminary results and the extensive, decades-long application of the micronucleus assay in various contexts, we are confident that this protocol will prove to be both feasible and effective.

This methodology has the potential to be integrated into routine occupational health surveillance protocols. Although various other genotoxicity assays exist—possessing comparable sensitivity—their invasiveness, resource requirements, and need for specialized expertise render them less suitable for workplace monitoring.

Micronuclei frequency data from the worker cohort should be compared with those from a control group matched for gender, age, and habits. Alternatively, a database of unexposed individuals could be established by study personnel for use as a reference in future investigations, or comparable data could be sourced from the existing literature. A provisional threshold of >2‰ micronuclei (equivalent to three micronuclei per 1500 reticulocytes) is suggested as an indicator of concern—based on prior studies employing flow cytometry, which reported frequencies of 1.2±0.5‰ [10], 0.9±0.6‰ [11], 0.4±0.1‰ [12], and 1.3±0.5‰ [13] in unexposed populations.

This protocol is characterized by its simplicity and cost-effectiveness, requiring minimal reagents and no cell culture, antibodies, or DNA probes. The approximate cost per sample is only a few cents. Additionally, the sampling and staining procedures are straightforward and do not necessitate highly trained personnel, nor are they time-intensive. The primary time investment is in the scoring process, which typically takes 2 to 3 hours per sample. Overall, this timing is shorter than that required for other genotoxicity assays.

The ongoing development of deep neural network-based software aims to streamline the analysis of reticulocytes and micronucleated reticulocytes. Although the necessary automated microscopy equipment may be cost-prohibitive for individual facilities, this challenge can be mitigated by establishing national or regional centers. Such a centralized approach would allow users to leverage automated analysis via smear transmission, ensuring that high-standard diagnostics remain accessible and economically sustainable.

5. Conclusion

Some techniques and ideas used in this protocol are not new: Acridine Orange has been used to perform the micronucleus test on rodent samples [14], and antibody-stained human reticulocytes have been used to perform the flow cytometry micronucleus assay [10-13]. What is new is the entire procedure, combining an easy staining method with non-invasive sampling of human blood. Its minimal

invasiveness (requiring only one drop of blood per individual), ease of execution, very low cost, and potential for integration with dedicated automated software will also enable the test to be conducted by personnel responsible for the health surveillance of professionally exposed workers.

Funding: This research was funded by INAIL, grant number BRIC 2022 ID 54.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: I.U. contributed to the design and implementation of the research, A.S., C.G. and I.U. contributed to the writing of the manuscript.

References

1. Choy WN. Genetic Toxicology and Cancer Risk Assessment. CRC Press: Boca Raton, USA, 2001.
2. Albertini RJ, Kaden DA. Mutagenicity monitoring in humans: global versus specific origin of mutations. *Mut Res Rev*. 2020;786:108341. Doi: 10.1016/j.mrrev.2020.108341
3. Ladeira C, Smajdova L. The use of genotoxicity biomarkers in molecular epidemiology: applications in environmental, occupational, and dietary studies. *AIMS Genetics*. 2017;4:166-191. Doi: 10.3934/genet.2017.3.166
4. OECD Guideline for the Testing of Chemicals: Mammalian Erythrocyte Micronucleus Test, Guideline 474. OECD. Paris, France, 1997.
5. Udroiu I. Feasibility of conducting the micronucleus test in circulating erythrocytes from different mammalian species: an anatomical perspective. *Environ Mol Mutagen*. 2006;47(9):643-646. Doi:10.1002/em.20258
6. Keohane E, Smith L, Walenga J. Rodak's Hematology: Clinical Principles and Applications, 5th edition. Elsivier: Amsterdam, Netherlands, 2016.
7. Udroiu I, Antoccia A, Sgura A. Long-term genotoxic effects in the hematopoietic system of prenatally X-irradiated mice. *Int J Radiat Biol.* 2017;93:261-269. Doi: 10.1080/09553002.2017.1239137
8. Chaubey RC, Bhilwade HN, Joshi BN, Chauhan PS. Studies on the migration of micronucleated erythrocytes from bone marrow to the peripheral blood in irradiated Swiss mice. *Int J Radiat Biol.* 1993;63:239-245. Doi: 10.1080/09553009314550311
9. Lenarczyk M, Slowikowska MG. The micronucleus assay using peripheral blood reticulocytes from X-ray-exposed mice. *Mutat Res Genet Toxicol Environ Mutagen*. 1995;335:229-234. Doi: 10.1016/0165-1161(95)00025-9
10. Abramsson-Zetterberg L, Zetterberg G, Bergqvist M, Grawé J. Human cytogenetic biomonitoring using flow-cytometric analysis of micronuclei in transferrin-positive immature peripheral blood reticulocytes. *Environ Mol Mutagen.* 2000;36:22-31. Doi: 10.1002/1098-2280(2000)36:1<22::AID-EM4>3.0.CO;2-U
11. Dertinger SD, Torous DK, Hall NE, et al. Enumeration of micronucleated CD71-positive human reticulocytes with a single-laser flow cytometer. *Mutat Res.* 2002;515:3-14. Doi: 10.1016/s1383-5718(02)00009-8
12. Offer T, Ho E, Traber MG, Bruno RS, Kuypers FA, Ames BN. A simple assay for frequency of chromosome breaks and loss (micronuclei) by flow cytometry of human reticulocytes. *FASEB J.* 2005;19:485-487. Doi: 10.1096/fj.04-2729fje
13. Stopper H, Hempel K, Reiners C, et al. Pilot study for comparison of reticulocyte-micronulei with lymphocyte-micronuclei in human biomonitoring. *Toxicol Lett.* 2005; 156: 351-60. Doi: 10.1016/j.toxlet.2004.12.007
14. Tinwell H, Ashby J. Comparison of acridine orange and Giemsa stains in several mouse bone marrow micronucleus assays--including a triple dose study. *Mutagenesis*. 1989;4:476-481. Doi: 10.1093/mutage/4.6.476

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18834
DOI: 10.23749/mdl.2026.18834

Reference Values for Maximal Isometric Handgrip and Pinch Strength in Healthy Italian Adults Without Occupational Biomechanical Overload

Emma Sala[1,2,*], Nicola Riolfi[2], Alessandro De Bellis[2], Andrea Bisioli[2], Cesare Tomasi[2], Francesco Romagnoli[2], Giuseppe De Palma[1,2]

[1]Unit of Occupational Health, Hygiene, Toxicology and Prevention, ASST Spedali Civili di Brescia, Brescia, Italy
[2]Unit of Occupational Health and Industrial Hygiene, Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia, Brescia, Italy

Keywords: Handgrip Strength; Pinch Strength; Reference Values; Occupational Health; Biomechanical Load

Abstract

Background*: Handgrip strength (HGS) and pinch strength are key indicators of hand function with relevant clinical and ergonomic implications. In Italy, normative reference values for combined grip and pinch strength measurements in healthy adults not exposed to upper-limb biomechanical overload are currently lacking. The primary aim of this cross-sectional study was to provide descriptive reference values for HGS and three pinch types, stratified by sex and age group, in a healthy Italian adult population (18–65 years) not exposed to occupational biomechanical overload. The secondary aim was to investigate the influence of sex, age, body mass index (BMI), and hand dominance, as well as the relationships among strength measures.* **Methods***: A total of 813 participants (319 men and 494 women) were evaluated. Measurements were performed using a calibrated Baseline® hand dynamometer and pinch gauge according to the standardized protocol of the American Society of Hand Therapists (ASHT).* **Results***: Men showed higher strength values than women across all measures. The association between body mass index (BMI) and strength was more pronounced and consistent in women than in men, with statistical significance observed only for selected pinch measures in men. Strength values showed a non-linear distribution across age groups, with peak levels observed between 30 and 49 years in men and between 40 and 49 years in women. Hand dominance showed a selective effect in both sexes, influencing specific pinch tasks but not handgrip strength.* **Conclusions***: This study provides the first national joint normative dataset for handgrip and pinch strength in healthy Italian adults aged 18–65 years not exposed to occupational biomechanical overload. These findings may support clinical and occupational assessments and contribute to the development of broader Italian normative reference tables.*

1. Introduction

Handgrip strength (HGS) represents the maximal voluntary force generated by the combined contraction of extrinsic and intrinsic muscles acting across the joints of the hand [1, 2]. It is a practical, safe, non-invasive, reliable, and feasible method for assessing muscle strength across all age groups. The test can be easily administered by minimally trained personnel and is readily interpretable in clinical

Received 17.03.2026 – Accepted 30.03.2026
*Corresponding Author: Emma Sala; E-mail: emma.sala@unibs.it

settings [3, 4]. Accordingly, it is routinely used by occupational physicians, occupational therapists, and other trained professionals in a wide range of clinical contexts. Evidence indicates that grip strength is an appropriate proxy for overall muscular strength [5], reflects the impact of neuromuscular and musculoskeletal disorders as well as cardiovascular diseases, and represents an important indicator in the diagnosis of sarcopenia [6] and frailty [7], as well as a predictor of mortality [8]. Pinch strength refers to the force generated by thumb opposition against the distal phalanges of other fingers and can be evaluated using standardized tasks, including pulp pinch, lateral (key) pinch, and tip pinch [9].

A widely accepted approach for interpreting HGS test results is to compare them with normative reference values. Such norms allow individual performance to be evaluated relative to peers of the same sex and age, facilitating the identification of individuals with reduced muscle strength who may be at increased health risk or require intervention, as well as those with higher strength levels who may exhibit superior physical or occupational performance. Normative values can also be used to monitor physiological ageing by assessing longitudinal changes in strength capacity [10]. In occupational medicine, they are additionally useful for evaluating outcomes of surgical or rehabilitative interventions, monitoring recovery after injury or illness, and supporting fitness-for-work assessments. In occupational health practice, grip and pinch strength measurements are also frequently used to support functional evaluations and return-to-work assessments. Therefore, reliable normative data for grip and pinch strength are essential for multiple clinical and occupational applications.

Normative values for HGS and pinch strength have been widely reported for decades; however, most available datasets derive from local or regional samples rather than nationally representative populations, and frequently cover restricted age ranges rather than the entire adult lifespan [11–29]. Numerous factors are known to influence HGS, including age, height, body weight, and body mass index (BMI). Several studies have shown that HGS is consistently higher in men than in women across comparable age groups throughout adulthood (20–69 years). Peak HGS has been reported between 25 and 50 years [36], whereas a recent systematic review suggests a peak between 30 and 39 years [10]. Variability in HGS has been shown to correlate negatively with age and positively with wrist, arm, and hand circumference, palm length, and arm muscle mass. Other potential determinants include BMI and height, although findings regarding these variables remain inconsistent [20, 27]. Differences between dominant and non-dominant hands have also been described, with the dominant hand generally reported to be approximately 10% stronger [36].

Normative reference values should ideally be derived from representative samples collected within the last 15–20 years [13]. Comparisons across populations indicate regional variability [16, 24], particularly between developed and developing countries [31, 32]. Consequently, despite the recent publication of international normative values by Tomkinson et al. in 2024 [10] based on a very large dataset (approximately 2.4 million individuals), such variability may limit the direct applicability of international norms to specific national populations [12]. To the best of our knowledge, in the Italian context, although recent studies have proposed normative values for grip strength [33] and pinch strength [34], a comprehensive dataset providing combined reference values for both grip and pinch strength in healthy adults aged 18–65 years not exposed to upper-limb biomechanical overload is still lacking.

The present study was designed to address this gap. The primary objective was to estimate reference values for HGS and three pinch types (lateral, three-point, and tip-to-tip) stratified by sex and age group in a sample of healthy adults aged 18–65 years not occupationally exposed to biomechanical overload of the upper limbs and without a history of substantial non-occupational upper-limb overload. Secondary objectives were to investigate the influence of sex, age, BMI, and hand dominance on strength measures, to assess relationships between grip and pinch strength, and to compare the findings with normative data available from other countries.

2. Methods

2.1. Study Design and Population

A cross-sectional study was conducted in a sample of 813 healthy adults (319 men and 494 women) aged 18–65 years who were not occupationally exposed to upper-limb biomechanical overload. Participants were recruited during routine occupational health surveillance examinations performed by occupational physicians at two large hospital organizations in Northern Italy (Brescia and Trento) between November 2020 and June 2025. Written informed consent was obtained from all participants prior to data collection. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. All data were anonymized prior to analysis and processed in accordance with the EU General Data Protection Regulation (GDPR 2016/679) and national privacy regulations.

Exclusion criteria included any current or previous musculoskeletal, neurological, or metabolic condition potentially affecting upper-limb strength. Individuals self-reporting secondary occupations or leisure activities (e.g., sports or manual hobbies) involving relevant biomechanical loading of the upper limbs were also excluded.

For analytical purposes, participants were stratified into five age groups: 18–29 years (Group 1), 30–39 years (Group 2), 40–49 years (Group 3), 50–59 years (Group 4), and 60–65 years (Group 5).

2.2. Anthropometric and Descriptive Variables

For each participant, the following anthropometric and descriptive variables were recorded: sex, age, height, body weight, and body mass index (BMI), calculated as weight in kilograms divided by height in meters squared (kg/m^2). Hand dominance was assessed by self-report and classified into four categories: right-handed, left-handed, ambidextrous, and corrected left-handed. These variables were collected to enable stratified analyses by sex and age group and to examine potential associations between anthropometric characteristics and grip and pinch strength outcomes.

2.3. Strength Assessment Protocol

Handgrip strength (HGS) and pinch strength were measured using a Baseline® hydraulic hand dynamometer and a Baseline® mechanical pinch gauge, respectively. All assessments were performed in accordance with the standardized protocol of the American Society of Hand Therapists (ASHT). Participants were evaluated in a seated position with the shoulder adducted and neutrally rotated, the elbow flexed at 90°, the forearm in neutral position, and the wrist positioned between 0–15° of extension and 0–15° of ulnar deviation. For handgrip testing, the dynamometer handle was set at the second position for all participants.

Both instruments were calibrated by the manufacturer and periodically recalibrated throughout the study in accordance with technical specifications to ensure measurement accuracy. All measurements were performed by two trained assessors to minimize inter-rater variability. When necessary, a third trained operator assisted with data recording and participant management.

Pinch strength was assessed bilaterally using three pinch types in the following order:

1. Lateral pinch (key pinch) (Figure 1).
2. Three-point pinch (Figure 2).
3. Tip-to-tip pinch (Figure 3).

For pinch strength assessment, two trials per hand were performed in alternating sequence. Handgrip strength was subsequently evaluated with three trials per hand. For each strength variable, the final recorded value was calculated as the arithmetic mean of valid trials.

A trial was considered valid when the variability between repeated measurements did not exceed 10%. If this criterion was not satisfied, an additional trial was performed. A standardized rest interval of approximately 15 seconds was allowed between consecutive trials to minimize fatigue effects. Standardized verbal instructions were provided without motivational encouragement. Participants were instructed to progressively increase force to a maximal voluntary contraction over approximately 1 second

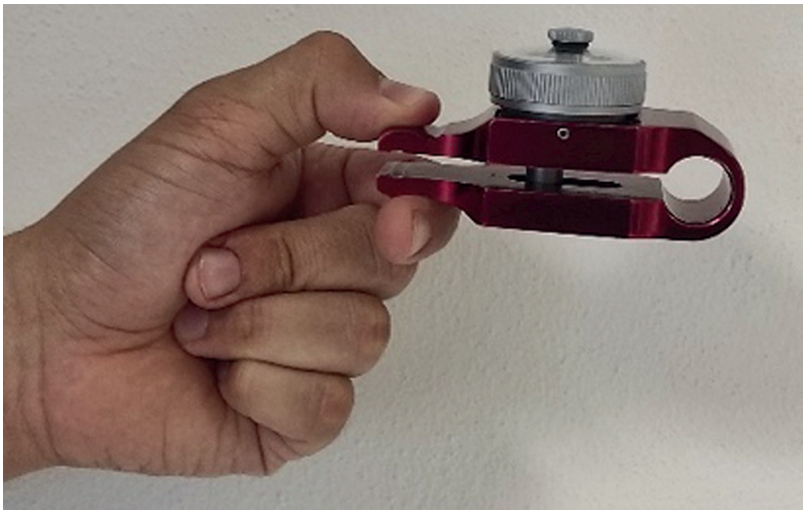

Figure 1. SEQ Figura * ARABIC 1: Lateral pinch (Key pinch).

Figure 2. Three point pinch (TTT).

and to maintain the contraction for approximately 4 seconds.

Figure 4 shows the standardized testing position for handgrip strength assessment according to the ASHT protocol. Participants were seated with the shoulder adducted and neutrally rotated, the elbow flexed at 90°, the forearm in neutral position, and the wrist positioned between 0–15° of extension and 0–15° of ulnar deviation. The figure illustrates hand positioning for handgrip strength and for the three pinch types (lateral pinch, three-point pinch, and tip-to-tip pinch).

2.4. Statistical Analysis

Data were initially entered into Microsoft Excel® and subsequently analyzed using IBM SPSS® Statistics (version 26.0.1; IBM Corp., Armonk, NY, USA).

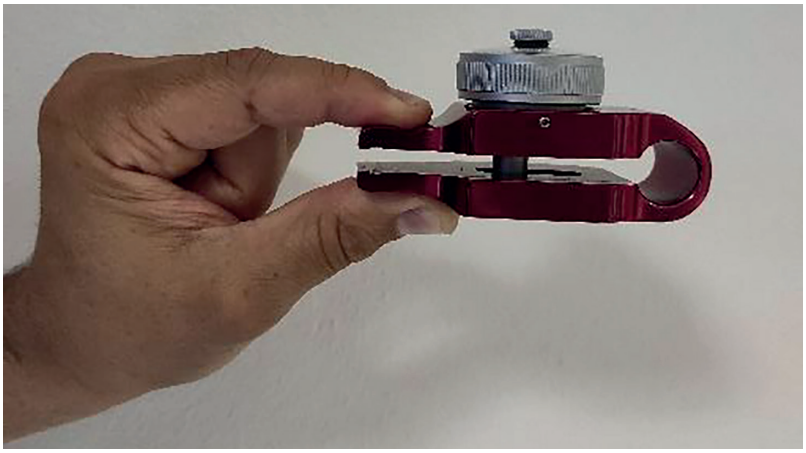

Figure 3. Two point pinch (tip to tip).

Figure 4. Hand Grip Strenght Test.

The distribution of continuous variables was assessed using the Kolmogorov–Smirnov test with Lilliefors correction. Because grip and pinch strength variables were not normally distributed, nonparametric statistical methods were used. Descriptive statistics are presented as medians and ranges (minimum–maximum). Differences between sexes were evaluated using the Mann–Whitney U test. Comparisons across age groups (18–29, 30–39, 40–49, 50–59, and 60–65 years) and hand dominance categories were performed using the Kruskal–Wallis test.

Associations between categorical variables (e.g., sex and hand dominance) were examined using Pearson's chi-square test. To investigate relationships between anthropometric continuous

variables and strength outcomes, Spearman's rank correlation coefficients were calculated between age, BMI, and each grip and pinch strength measure. Correlation analyses were stratified by sex to explore potential sex-specific patterns in the association between BMI and strength. Age-related trends were assessed both through correlation analyses and through comparisons across predefined age groups, allowing identification of potential non-linear (non-monotonic) patterns across the adult lifespan rather than assuming a single linear decline. All statistical tests were two-tailed, and p-values were calculated.

3. RESULTS

3.1. Study Population

A total of 813 participants were included in the analysis, comprising 319 men (39.2%) and 494 women (60.8%). The demographic and anthropometric characteristics of the study population are summarized in Table 1.

3.2. Normative Handgrip and Pinch Strength Values

Sex- and age-specific normative values for handgrip strength and pinch strength (lateral pinch, three-point pinch, and tip-to-tip pinch), measured bilaterally, are reported in Table 2.

Across all age groups, men consistently showed higher median values than women for all grip and pinch strength measures. In both sexes, strength values followed a non-linear age-related pattern, with higher values observed in early and mid-adulthood and a progressive decline in older age groups. Overall, the distribution of strength values across age groups suggested peak levels in early to mid-adulthood.

3.3. Grip and Pinch Strength Differences By Sex

Comparisons between men and women showed significantly higher grip and pinch strength in men across all measures. These differences were confirmed by the Mann–Whitney U test for all grip and pinch variables. No significant difference in age distribution between sexes was observed (P = 0.462).

3.4. Hand Dominance Distribution

The distribution of hand dominance by sex is reported in Supplementary Table S1. A statistically significant association between sex and hand dominance was observed (χ^2 = 8.313; df = 3; P = 0.040). Analysis of adjusted residuals indicated that this association was driven exclusively by the category of corrected left-handedness, which was overrepresented among men and underrepresented among women. No significant deviations from expected frequencies were observed for right-handed, non-corrected left-handed, or ambidextrous participants.

3.5. Correlations Between Anthropometric Variables and Strength Measures

Spearman correlation analyses stratified by sex are reported in Supplementary Tables S2 (men) and S3 (women). In men, age showed weak correlations with strength measures, including a weak negative correlation with right-hand grip strength and weak positive correlations with selected pinch measures. Body mass index showed weak but statistically significant positive correlations with selected pinch measures, whereas correlations with grip strength

Table 1. Demographic and Anthropometric Characteristics of the Study Population.

Sex	Participants (n)	% of Total	Age, years, median (range)	BMI, kg/m^2 (median [range])
Men	319	39.2	40 (18–65)	24.38 (16.5–38.8)
Women	494	60.8	43 (18–65)	22.04 (15.9–47.8)
Total	813	100	42 (18–65)	23.14 (15.9–47.8)

Table 2a. Normative Handgrip and Pinch Strength Values in Men by Age Group (Median and Range, kg).

Age Group (years)	n Men	Grip R (Men)	Grip L (Men)	Lat. Pinch R (Men)	Lat. Pinch L (Men)	TPP R (Men)	TPP L (Men)	Tip-to-Tip R (Men)	Tip-to-Tip L (Men)
18–29	93	51.00 (29.00–71.00)	48.33 (23.33–71.00)	10.50 (6.50–16.75)	10.00 (6.25–15.50)	9.00 (5.50–13.25)	8.50 (5.50–14.75)	6.75 (3.75–11.50)	6.50 (3.75–10.50)
30–39	66	53.00 (14.25–92.66)	49.66 (14.50–80.66)	11.25 (7.00–17.00)	10.44 (7.25–16.00)	9.25 (5.75–15.00)	9.00 (5.50–14.50)	7.25 (4.50–10.25)	7.00 (4.00–9.50)
40–49	54	52.50 (35.33–73.00)	49.66 (33.33–75.00)	11.00 (5.50–15.50)	10.19 (7.00–14.25)	9.50 (5.25–13.00)	9.00 (5.00–13.75)	7.25 (4.75–11.00)	7.44 (3.75–11.50)
50–59	73	48.33 (35.00–72.33)	46.00 (32.00–64.00)	10.75 (6.50–15.10)	10.25 (6.25–15.25)	9.25 (6.00–14.87)	8.75 (6.00–14.12)	7.00 (4.13–12.00)	7.00 (4.50–11.75)
60–65	33	48.66 (25.00–65.00)	46.00 (21.33–61.66)	10.50 (7.50–12.50)	10.25 (7.38–12.75)	9.00 (6.50–12.00)	8.50 (5.25–12.00)	7.00 (4.88–10.25)	6.50 (5.38–9.38)
Total	319	50.66 (14.25–92.66)	48.00 (14.50–80.66)	10.75 (5.50–17.00)	10.25 (6.25–16.00)	9.00 (5.25–15.00)	8.75 (5.00–14.75)	7.00 (3.75–12.00)	7.00 (3.75–11.75)

Table 2b. Normative Handgrip and Pinch Strength Values in Women by Age Group (Median and Range, kg).

Age Group (years)	n Women	Grip R (Women)	Grip L (Women)	Lat. Pinch R (Women)	Lat. Pinch L (Women)	TPP R (Women)	TPP L (Women)	Tip-to-Tip R (Women)	Tip-to-Tip L (Women)
18–29	115	31.33 (20.00–54.33)	28.33 (15.66–51.00)	7.50 (4.75–10.50)	7.00 (4.00–15.75)	6.00 (3.25–11.25)	5.75 (2.50–10.00)	5.00 (2.80–7.75)	4.50 (2.80–7.00)
30–39	104	32.00 (22.00–45.00)	29.00 (16.33–43.00)	7.75 (4.00–10.50)	7.25 (3.50–12.00)	6.75 (4.00–11.00)	6.38 (4.00–10.00)	5.00 (3.25–7.00)	4.75 (3.00–7.50)
40–49	114	33.33 (18.66–56.66)	30.83 (15.66–51.66)	7.75 (3.50–11.00)	7.50 (3.00–11.00)	7.00 (2.75–11.75)	6.50 (2.75–12.00)	5.25 (2.50–10.38)	5.00 (2.50–9.00)
50–59	124	30.66 (13.33–43.33)	28.50 (12.00–40.99)	7.50 (3.75–10.50)	7.00 (3.25–11.50)	6.63 (2.75–12.00)	6.25 (2.50–13.75)	5.13 (2.25–7.75)	5.00 (2.50–8.00)
60–65	37	27.66 (18.66–43.33)	25.00 (18.00–40.66)	7.00 (4.75–9.75)	6.25 (4.50–10.00)	6.00 (4.50–8.50)	6.00 (4.25–9.00)	4.75 (3.50–7.25)	4.25 (3.00–7.00)
Total	494	31.00 (13.33–56.66)	28.67 (12.00–51.66)	7.50 (3.50–11.00)	7.00 (3.00–15.75)	6.50 (2.75–12.00)	6.00 (2.50–13.75)	5.00 (2.25–10.38)	4.75 (2.50–9.00)

were weaker and not consistently significant. All grip and pinch strength measures were positively intercorrelated.

In women, age was weakly negatively correlated with right-hand grip strength and weakly positively correlated with selected pinch measures. Body mass index showed weak but statistically significant positive correlations with all grip and pinch strength measures. As observed in men, all strength measures were positively intercorrelated.

3.6. Effect of Hand Dominance On Strength Measures

Differences in grip and pinch strength across hand dominance categories were evaluated separately in men and women using the Kruskal–Wallis test (Supplementary Table S4).

In men, statistically significant differences across dominance categories were observed only for left-hand lateral pinch and right-hand tip-to-tip pinch, whereas no significant differences were detected for handgrip strength or the remaining pinch measures. In women, statistically significant differences across dominance categories were observed only for right-hand three-point pinch, with no significant differences for handgrip strength or other pinch measures.

3.7. Differences in Strength Measures Across Age Groups

Differences in grip and pinch strength across age groups were assessed separately for men and women using the Kruskal–Wallis test (Supplementary Table S5).

In men, statistically significant differences across age groups were observed for body mass index, bilateral handgrip strength, left-hand three-point pinch, and bilateral tip-to-tip pinch, whereas no significant differences were found for lateral pinch strength or right-hand three-point pinch. In women, statistically significant differences across age groups were observed for all tested variables, including body mass index, bilateral handgrip strength, and all pinch strength measures.

4. Discussion

In this study, men exhibited higher handgrip and pinch strength values than women across all measured variables. This finding is consistent with large population-based cohorts and recent international syntheses reporting higher peak strength and greater values across the adult lifespan in men compared with women [10, 13, 23, 28, 29]. These differences are plausibly related to well-established biological factors, including differences in muscle mass, hormonal profiles, and body composition, and further support the use of sex-specific normative reference values.

The association between body mass index (BMI) and strength differed by sex. In women, BMI showed weak but consistent positive correlations with all grip and pinch strength measures, whereas in men significant associations were limited to selected pinch tasks (Supplementary Tables S2 and S3). This pattern suggests a more consistent association between BMI and manual strength in women, potentially related to sex-specific differences in body composition and fat–lean mass distribution. Overall, these findings are consistent with previous studies reporting positive associations between BMI and strength [12, 14, 17], while also aligning with evidence indicating non-linear or heterogeneous relationships between BMI and hand strength [18, 27]. Further investigations in larger and more diverse samples are warranted to clarify these sex-specific associations.

The combined analysis of sex- and age-stratified normative values (Table 2), age-group comparisons, and correlation analyses indicates that strength trajectories across adulthood are non-linear. For most measures, strength peaked in young to mid-adulthood, followed by a gradual decline in older age groups. In some pinch measures, particularly in men, strength profiles appeared relatively stable across age groups, suggesting lower sensitivity to age-related gradients. This observation is consistent with findings first reported by Mathiowetz et al. in 1985. [15].

Overall, the age-related pattern observed in this study follows a non-linear (bell-shaped) trajectory

corresponding to the expected profile described in international reference curves, with a peak in early adulthood and a progressive decline thereafter [10, 13, 29]. A slight shift in peak strength was observed between sexes, with peak values occurring at 30–49 years in men and 40–49 years in women. This minor discrepancy compared with large transnational datasets reporting a peak at 30–39 years [10] may reflect sample-specific characteristics. The weak Spearman correlation coefficients observed are likely explained by the non-monotonic and age-dependent nature of the age–strength relationship.

In the present sample, the effect of hand dominance on strength measures was selective rather than generalized. In men, significant differences between dominance categories were observed only for left-hand lateral pinch and right-hand tip-to-tip pinch, whereas in women a significant effect was detected only for right-hand three-point pinch (Supplementary Table S4). No consistent dominance-related differences were observed for handgrip strength (HGS).

These findings partially contrast with reports describing a general dominance-related advantage for both handgrip and pinch strength [13, 14, 18, 34, 36], but are consistent with studies reporting no systematic differences between dominant and non-dominant hands [37]. Methodological differences, particularly the use of between-group dominance analyses rather than within-subject comparisons, may have attenuated small individual asymmetries and contributed to the observed discrepancies.

Finally, all grip and pinch strength measures were positively and strongly intercorrelated, indicating a coherent functional profile of manual strength. Positive correlations were also observed between contralateral measures (e.g., right vs. left handgrip), supporting the internal consistency and reliability of the measurement protocol (Supplementary Tables S2 and S3).

Beyond the occupational medicine framework, the availability of national normative reference values for handgrip and pinch strength has relevant implications across several medical disciplines. Handgrip strength is widely recognised as a sensitive and early marker of neuromuscular function and is routinely applied in neurology and neurorehabilitation for monitoring both peripheral and central nervous system disorders, including multiple sclerosis, amyotrophic lateral sclerosis, peripheral neuropathies, and post-stroke functional outcomes.

In rheumatology and hand orthopaedics, grip and pinch strength represent core functional outcomes in conditions directly affecting hand performance, such as rheumatoid arthritis, hand osteoarthritis, Raynaud's phenomenon, and connective tissue diseases, as well as in the evaluation of recovery following surgical procedures including distal radius fracture fixation, tendon reconstruction, carpal tunnel release, and wrist arthroplasty.

In geriatrics and ageing medicine, handgrip strength is a recognised diagnostic criterion for sarcopenia according to the EWGSOP2 consensus and a robust predictor of mortality, disability, hospitalisation, and cognitive decline. Although the present study includes individuals up to 65 years of age, the proposed reference values may represent a useful baseline for future longitudinal studies investigating muscle strength trajectories in older Italian populations.

Further applications extend to internal medicine and oncology, where grip strength is increasingly used as a proxy for muscle mass and nutritional status in patients with chronic diseases and cancer, and is incorporated into frailty indices and preoperative risk stratification tools. In these settings, comparison with national reference values may improve the accuracy of functional assessment and risk classification. Additional fields of application include endocrinology and diabetology, sports medicine and motor sciences, paediatrics, and psychosomatic medicine, where grip and pinch strength are increasingly employed as integrative functional and health-related indicators.

From a methodological perspective, the availability of sex- and age-stratified normative values derived from an Italian population reduces the bias associated with the use of North American or Northern European reference data, which have been shown to systematically overestimate strength in Mediterranean populations. This aspect represents one of the most robust and clinically transferable contributions of the present study and supports the

use of these reference values across both occupational and non-occupational clinical contexts.

5. Study limitations and future perspectives

Several limitations should be acknowledged. First, specific upper-limb anthropometric measures (e.g., limb circumferences, hand dimensions, and forearm muscle mass), which are known to influence grip and pinch strength [14, 16, 21], were not collected. Inclusion of these variables in future studies may help explain part of the observed variability and allow a more precise normalization of strength values.

Second, maximal voluntary contraction is inherently effort-dependent and may be influenced by participant motivation and operator-related factors. Although standardized ASHT procedures, calibrated instruments, and trained assessors were used, a residual degree of measurement variability cannot be excluded. Future studies may benefit from standardized pre-test familiarization procedures, uniform verbal instructions, and additional practice trials.

Third, the study population was limited to two geographical areas and to adults aged 18–65 years not occupationally exposed to upper-limb biomechanical overload. Consequently, the results cannot be generalized to younger individuals, adults over 65 years of age, or workers exposed to occupational upper-limb biomechanical risk factors. Expanding both the sample size and the age range represents an important objective for future research.

In addition, recruitment during occupational health surveillance may introduce a potential healthy worker effect.

Finally, this study did not include formal statistical comparisons with large international datasets. Such analyses will be addressed in future work to explore transnational and regional differences and to contribute to the development of broader normative reference tables.

Despite these limitations, the present sample represents one of the largest Italian cohorts currently available for the normative assessment of hand strength and provides the first combined national reference dataset for both grip and pinch strength in healthy Italian adults.

6. Conclusions

This study provides the first Italian national joint normative reference values for handgrip strength and three pinch strength measures in healthy adults aged 18–65 years. Measurements were conducted according to the standardized ASHT protocol using calibrated instruments and trained assessors in a large sample comprising 319 men and 494 women not occupationally exposed to upper-limb biomechanical overload.

These reference values are directly applicable in both clinical practice and occupational medicine, including the evaluation of surgical and rehabilitative outcomes, monitoring of injury- and disease-related functional impairment, assessment of fitness for work, and prognostic evaluation in frail individuals. More broadly, normative strength data support health, safety, comfort, productivity, and the ergonomic design of tools and tasks. In occupational settings, handgrip strength assessment is particularly relevant for pre-employment screening, workload estimation, and return-to-work evaluations.

Although international and some Italian reference data on handgrip or pinch strength are available, this study uniquely provides joint normative values for both grip and pinch strength in a healthy Italian population. The resulting dataset represents a valuable quantitative contribution to future Italian and international reviews and meta-analyses and provides a robust foundation for the development of more comprehensive national normative reference tables.

Institutional Review Board Statement: All evaluations performed in the study were included in occupational risk assessment and health surveillance activities that are mandatory by law under the Italian D.Lgs. 81/08 (Testo Unico sulla Salute e Sicurezza sul Lavoro). Consistent with the criteria established by the Italian Ministry of Health Decree of 30 January 2023 for observational studies conducted in the context of occupational health monitoring, formal approval by an Ethics Committee was not required. All participants provided written informed consent prior to enrollment. The study was conducted in accordance with the WMA Declaration of Helsinki and the General Data Protection Regulation (GDPR, Regulation EU 679/2016).

Declaration of Interests: The authors declare no conflict of interest.

Author Contributions: E. Sala conceived and designed the study and drafted the manuscript; F. Romagnoli conceived the study; N. Riolfi, A. De Bellis, and A. Bisioli collected the data; C. Tomasi performed the statistical analysis; G. De Palma conceived the study and critically revised the manuscript. All authors critically revised the manuscript and approved the final version.

References

1. Schlüssel MM, dos Anjos LA, de Vasconcellos MTL, Kac G. Reference values of handgrip dynamometry of healthy adults: a population-based study. *Clin Nutr*. 2008;27(4):601–607. Doi: 10.1016/j.clnu.2008.04.004.
2. Mitsionis G, Pakos EE, Stafilas KS, Paschos N, Papakostas T, Beris AE. Normative data on hand grip strength in a Greek adult population. *Int Orthop*. 2009;33(3): 713–717. Doi: 10.1007/s00264-008-0551-x.
3. Cuenca-Garcia M, Marin-Jimenez N, Perez-Bey A, et al. Reliability of field-based fitness tests in adults: a systematic review. *Sports Med*. 2022;52(8):1961–1979. Doi: 10.1007/s40279-021-01635-2.
4. Suni JH, Miilunpalo S I, Asikainen T M, et al. Safety and feasibility of a health-related fitness test battery for adults. *Phys Ther*. 1998;78(2):134–148. Doi:10.1093 /ptj/78.2.134.
5. Bohannon RW. Hand-grip dynamometry predicts future outcomes in aging adults. *J Geriatr Phys Ther*. 2008.
6. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48(1):16–31. Doi: 10.1093/ageing/afy169.
7. Fried LP, Tangen C M, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146–M156. Doi: 10.1093/gerona /56.3.M146.
8. Bohannon RW. Hand-grip dynamometry predicts future outcomes in aging adults. *J Geriatr Phys Ther*. 2008;31(1):3-10. Doi: 10.1519/00139143-200831010 -00002. PMID: 18489802.
9. Torri D, Sala E, Orizio C, Apostoli P. Stima della presa in pinch: approfondimento degli aspetti metodologici ed applicativi mediante l'impiego di un nuovo apparato portatile, P force Met [Pinch assessing in bio mechanical analysis: methodological and applicative aspects dealt by using the new P force Met portable instrument]. *G Ital Med Lav Ergon*. 2011;33(1):63-73. Italian. PMID: 21425633.
10. Tomkinson GR, J Lang J, Rubín L,et al. International norms for adult handgrip strength: a systematic review of data on 2.4 million adults aged 20 to 100+ years from 69 countries and regions. *J Sport Health Sci*. 2024;14:101014. Doi: 10.1016/j.jshs.2024.101014
11. Lam NW, Goh HT, Kamaruzzaman SB, Chin AV, Poi PJH, Tan MP. Normative data for hand grip strength and key pinch strength stratified by age and gender for a multiethnic Asian population. *Singapore Med J*. 2016;57(10):578–584. Doi: 10.11622/smedj.2015164.
12. Werle S, Goldhahn J, Drerup S, Simmen BR, Sprott H, Herren DB. Age- and gender-specific normative data of grip and pinch strength in a healthy adult Swiss population. *J Hand Surg Eur*. 2009;34(1):76–84. Doi: 10.1177/1753193408096763.
13. Wang YC, Bohannon RW, Li X, Sindhu B, Kapellusch J. Hand-grip strength: normative reference values and equations for individuals 18 to 85 years of age residing in the United States. *J Orthop Sports Phys Ther*. 2018;48(9):685–693. Doi: 10.2519/jospt.2018.7851.
14. Ekşioğlu M. Normative static grip strength of the population of Turkey: effects of various factors and comparison with international norms. *Appl Ergon*. 2016;52:8–17. Doi: 10.1016/j.apergo.2015.06.023.
15. Mathiowetz V, Kashman N, Volland G, Weber K, Dowe M, Rogers S. Grip and pinch strength: normative data for adults. *Arch Phys Med Rehabil*. 1985;66(2):69-74. PMID: 3970660.
16. Macdermid JC, Fehr L, Lindsay K. The effect of physical factors on grip strength and dexterity. *Br J Hand Ther*. 2002;7(4):112–118. Doi: 10.1177/175899830 200700401.
17. Xiao G, Lei L, Dempsey PG, Lu B, Liang Y. Isometric muscle strength and anthropometric characteristics of a Chinese sample. *Int J Ind Ergon*. 2005;35(7):674–679. Doi: 10.1016/j.ergon.2005.02.003.
18. Kamarul T, Ahmad TS, Loh WY. Hand grip strength in the adult Malaysian population. *J Orthop Surg* (Hong Kong). 2006;14(2):172–177. Doi: 10.1177/230949900 601400213.
19. Bohannon RW, Peolsson A, Massy-Westropp N, Desrosiers J, Bear-Lehman J. Reference values for adult grip strength measured with a Jamar dynamometer: a descriptive meta-analysis. *Physiotherapy*. 2006;92(1): 11–15. Doi: 10.1016/j.physio.2005.05.003.
20. Anakwe RE, Huntley JS, McEachan JE. Grip strength and forearm circumference in a healthy population. *J Hand Surg Eur*. 2007;32(2):203–209. Doi: 10.1016 /j.jhsb.2006.11.003.
21. Shim JH, Roh SY, Kim JS, et al. Normative measurements of grip and pinch strengths of the 21st century Korean population. *Arch Plast Surg*. 2013;40(1):52–56. Doi:10.5999/aps.2013.40.1.52.
22. Michael AI, Iyun AO, Olawoye OA, Ademola SA, Nnabuko RE, Oluwatosin OM. Normal values of key pinch strength in a healthy Nigerian population. *Ann Ib Postgrad Med*. 2015;13(2):84.
23. Dodds RM, Syddall HE, Cooper R, et al. Grip strength across the life course: normative data from twelve British studies. *PLoS One.* 2014;9(12). Doi: 10.1371/journal .pone.0113637.
24. Lo VEW, Chiu YC, Tu HH, Liu CW, Yu CY. A pilot study of five types of maximum hand strength among manufacturing industry workers in Taiwan. *Int J*

EnvironResPublicHealth.2019;16(23):4742.Doi:10.3390/ijerph16234742.
25. Alrashdan A, Ghaleb AM, Almobarek M. Normative static grip strength of Saudi Arabia's population and factors influencing grip strength. *Healthcare*. 2021;9(12):1647. Doi: 10.3390/healthcare9121647.
26. Luna-Heredia E, Martín-Peña G, Ruiz-Galiana J. Handgrip dynamometry in healthy adults. *Clin Nutr*. 2005;24(2):250–258. Doi: 10.1016/j.clnu.2004.10.007.
27. Massy-Westropp NM, Gill TK, Taylor AW, Bohannon RW, Hill CL. Hand grip strength: age and gender stratified normative data in a population-based study. *BMC Res Notes*. 2011;4:127. Doi: 10.1186/1756-0500-4-127.
28. He H, Wang D, Liu F, et al. Normative values of hand grip strength in a large unselected Chinese population: evidence from the China National Health Survey. *J Cachexia Sarcopenia Muscle*. 2023;14(3):1312–1321. Doi: 10.1002/jcsm.13223.
29. Perna FM, Coa K, Troiano RP, Lawman HG, Wang CY, Li Y, Moser RP, Ciccolo JT, Comstock BA, Kraemer WJ. Muscular Grip Strength Estimates of the U.S. Population from the National Health and Nutrition Examination Survey 2011-2012. *J Strength Cond Res*. 2016 Mar;30(3):867-74. Doi: 10.1519/JSC.0000000000001104. PMID: 26196662; PMCID: PMC7197498.
30. Larson CC, Ye Z. Development of an updated normative data table for hand grip and pinch strength: a pilot study. *Comput Biol Med*. 2017;86:40–46. Doi: 10.1016/j.compbiomed.2017.01.021. 30
31. Dodds RM, Syddall HE, Cooper R, Kuh D, Cooper C, Sayer AA. Global variation in grip strength: a systematic review and meta-analysis of normative data. *Age Ageing*. 2016;45(2):209–216. Doi: 10.1093/ageing/afv192.
32. Leong DP, Teo KK, Rangarajan S, et al. Reference ranges of handgrip strength from 125,462 healthy adults in 21 countries: the PURE study. *J Cachexia Sarcopenia Muscle*. 2016;7(5):535–546. Doi: 10.1002/jcsm.12112.
33. Landi F, Calvani R, Martone AM, et al. Normative values of muscle strength across ages in a real-world population: results from the Longevity Check-up 7+ project. *J Cachexia Sarcopenia Muscle*. 2020;11(6):1562–1569. Doi: 10.1002/jcsm.12610.
34. Dottor A, Sansone LG, Battista S, Mori L, Testa M. Flexion-extension strength of the index-thumb system in the Italian population: a cross-sectional study to gather normative data. *J Hand Ther*. 2023;36(1):85–96. Doi: 10.1016/j.jht.2021.05.004.
35. American Society of Hand Therapists. Clinical Assessment Recommendations. 3rd ed. Chicago: American Society of Hand Therapists; 2015.
36. Sidarta A, Soh LJ, Lie E, et al. Establishing normative pinch and grip strengths across adult age groups in Singapore. *BMC Sports Sci Med Rehabil*. 2025;17(1):84. Doi: 10.1186/s13102-025-01140-3.
37. Härkönen R, Piirtomaa M, Alaranta H. Grip strength and hand position of the dynamometer in 204 Finnish adults. *J Hand Surg Br*. 1993;18(1):129–132. Doi: 10.1016/0266-7681(93)90212-X.
38. Yim SY, Cho JR, Lee IY. Normative data and developmental characteristics of hand function for elementary school children in Suwon area of Korea: grip, pinch and dexterity study. *J Korean Med Sci*. 2003;18(4):552–558. Doi: 10.3346/jkms.2003.18.4.552.
39. Sayadizadeh M, Daliri M, Sadeghi M, Azimi MA, Mozafari JK, Moradi A. Normative values of grip and pinch strength and their predictor factors: Persian cohort study of healthcare staff. *Int Orthop*. 2025;49(3):549–557. Doi: 10.1007/s00264-025-06409-3.

APPENDIX

Supplementary Table S1. Hand Dominance by Sex (Chi-Square Analysis).

Test	Value	df	p value
Pearson χ^2	8.31	3	0.040
Likelihood ratio	8.557	3	0.036

Note: The association was driven by corrected left-handedness.

Supplementary Table S2. Spearman Correlation Matrix – Men.

	Age	BMI	Grip R	Grip L	Lateral Pinch R	Lateral Pinch L	TPP R	TPP L	Tip-to-tip R
BMI	0.28**								
Grip R	-0.12*	0.08							
Grip L	-0.11	0.04	0.89**						
Lateral Pinch R	-0.04	0.12*	0.45**	0.40**					
Lateral Pinch L	-0.03	0.10	0.44**	0.47**	0.82**				
TPP R	0.06	0.09	0.42**	0.39**	0.62**	0.59**			
TPP L	0.11*	0.10	0.40**	0.41**	0.58**	0.67**	0.85**		
Tip-to-tip R	0.12*	0.21**	0.47**	0.47**	0.52**	0.47**	0.63**	0.60**	
Tip-to-tip L	0.18**	0.22**	0.42**	0.46**	0.41**	0.50**	0.51**	0.60**	0.82**

***. Correlation is significant at the 0.01 level (two-tailed).*
**. Correlation is significant at the 0.05 level (two-tailed).*
**Spearman's rho. *p < 0.05; *p < 0.01.*
R = right hand.
L = left hand.

Supplementary Table S3. Spearman Correlation Matrix – Women.

	Age	BMI	Grip R	Grip L	Lateral Pinch R	Lateral Pinch L	TPP R	TPP L	Tip-to-tip R
BMI	0.28**								
Grip R	-0.11*	0.13**							
Grip L	-0.06	0.13**	0.89**						
Lateral Pinch R	-0.04	0.16**	0.46**	0.40**					
Lateral Pinch L	-0.03	0.21**	0.47**	0.47**	0.82**				
TPP R	0.08	0.14**	0.40**	0.31**	0.57**	0.55**			
TPP L	0.09*	0.19**	0.40**	0.37**	0.51**	0.60**	0.88**		
Tip-to-tip R	0.04	0.22**	0.49**	0.45**	0.54**	0.56**	0.63**	0.60**	
Tip-to-tip L	0.06	0.22**	0.54**	0.55**	0.55**	0.63**	0.55**	0.61**	0.81**

***. Correlation is significant at the 0.01 level (two-tailed).*
**. Correlation is significant at the 0.05 level (two-tailed).*

Supplementary Table S4. Effect of Hand Dominance on Strength Measures (Kruskal–Wallis Test).

Sex	Significant variables ($p < 0.05$)
Men	Lateral Pinch Left; Tip-to-Tip Right
Women	Three-Point Pinch Right

Supplementary Table S5. Differences in Strength Measures Across Age Groups (Kruskal–Wallis Test).

Sex	Variables with significant differences
Men	BMI; Grip R; Grip L; TPP L; Tip-to-Tip R/L
Women	All tested variables

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18371
DOI: 10.23749/mdl.2026.18371

Artificial Intelligence in Occupational Health Surveillance: Evaluating AI-Assisted ILO Classification of Radiographs of Pneumoconioses

Antonio Baldassarre[1,2,*], Martina Padovan[3], Alessandro Palla[4], Augusto Quercia[5], Rita Leonori[6], Stefano Dugheri[7], Nicola Mucci[1,2], Veronica Traversini[1,2]

[1]Department of Experimental and Clinical Medicine, University of Florence, Florence, Italy
[2]Occupational Medicine, Careggi University Hospital, Florence, Italy
[3]Preventive Medicine, Tuscany North-West Health Local Unit, Italy
[4]Intel Corporation, Santa Clara, USA
[5]Past Chief Workplace Prevention and Safety Unit, Viterbo Health Local Unit, Italy
[6]Workplace Prevention and Safety Unit, Viterbo Health Local Unit, Italy
[7]Department of Life Science, Health, and Health Professions, Link Campus University, 00165 Rome, Italy

Keywords: ILO International Classification of Radiographs of Pneumoconioses; Occupational Medicine; Artificial intelligence

Abstract

Background: *Pneumoconioses remain an important occupational health issue, particularly in low- and middle-income countries. The International Labour Organization (ILO) Classification standardizes chest radiograph interpretation but requires trained readers and is affected by inter-reader variability. This study evaluated whether generative multimodal artificial intelligence (AI) models can approximate ILO-based diagnostic reasoning.* **Methods:** *Eighty-two chest radiographs from the official NIOSH B Readers syllabus were analysed using four AI systems (GPT-4o, GPT-5, MedGemma-4B, MedGemma-27B). Each image was evaluated with a standardized prompt based on the 2022 revised ILO guidelines using deterministic settings. Model outputs were mapped to ILO codes and compared with the official answer keys of the ILO Standard Radiograph Set used for B Reader training and examination. Performance metrics included balanced accuracy, sensitivity, specificity, precision, and Matthews correlation coefficient (MCC). Bootstrap 95% confidence intervals, McNemar's test, and Cohen's κ assessed performance variability and agreement.* **Results:** *All four AI models showed moderate diagnostic performance, with balanced accuracy ranging from 60.8% to 70.3%. Sensitivity remained limited (35.5%–54.9%), while specificity was consistently high (84.6%–86.2%). MedGemma-27B performed best for small opacities, GPT-5 for pleural abnormalities and for technical quality. Large opacities and rare findings were systematically under-detected. Statistical comparisons showed significant differences between models, although agreement patterns were broadly similar.* **Conclusion:** *All AI models partially followed structured ILO radiographic criteria but did not achieve expert-level performance, confirming that they cannot replace certified B Readers. Larger, real-world datasets are needed to assess their potential clinical utility as supportive tools in occupational health surveillance programs.*

1. Introduction

Pneumoconioses refers to a group of occupational interstitial lung diseases caused by prolonged inhalation of mineral dusts such as silica, coal, and asbestos [1]. Despite advances in safety standards and dust-exposure control, these diseases remain a major global public health concern. According to

Received 10.12.2025 – Accepted 08.04.2026
* Corresponding Author: Antonio Baldassare; E-mail: antonio.baldassarre@unifi.it

the Global Burden of Disease 2023 data, approximately 18700 deaths were attributed to pneumoconioses worldwide in 2023, representing an 18% increase in absolute mortality since 2000, although the age-standardised death rate declined by 37.6% over the same period [2]. This discrepancy highlights the persistent risk of occupational exposure, particularly in the mining, construction, and manufacturing sectors in low- and middle-income countries, where effective dust mitigation and surveillance systems remain limited [3]. Given the complex radiographic presentation of pneumoconiosis [4] and the need for diagnostic consistency across clinical and surveillance settings, the International Labour Organization (ILO) developed the International Classification of Radiographs of Pneumoconioses to standardize the description and grading of chest radiographic abnormalities caused by occupational dust exposure.

This system provides a structured framework for evaluating image quality, parenchymal opacities (small and large), pleural abnormalities, and other findings, ensuring reproducibility and uniform interpretation across readers and countries [5]. In the United States, the National Institute for Occupational Safety and Health (NIOSH) adopted this classification as the reference standard for its B Reader Program, which certifies physicians who demonstrate proficiency in the ILO system through structured training and testing. This certification ensures high intra- and inter-reader reliability and remains the cornerstone of pneumoconioses surveillance and compensation programs worldwide [6, 7]. This methodological rigor reflects a broader scientific pursuit: the standardization of observation and reasoning as a foundation of reproducible knowledge.

Throughout history, humans have sought to digitize and standardize processes of reasoning and calculation. From Charles Babbage's differential engine [8, 9] and punched-card systems [10], early mechanical algorithms that inspired modern computing to contemporary data-driven models [11], the pursuit of reproducibility and efficiency has been central to scientific progress. In occupational medicine, the same principle of structured reasoning underpins the ILO classification, which can itself be regarded as a diagnostic algorithm, defining explicit decision steps for classifying radiographic findings related to dust exposure. Since the First International Conference of Experts on Pneumoconioses, held in 1930 in Johannesburg, the classification of radiographs of pneumoconioses has undergone successive revisions in 1938 (American), 1944 (Eck and Hanaut's), 1948 (Hasselt), 1949 (British) 1950 (Sydney), 1958 (Geneva), 1968, 1971, 1980, 2000, 2011, and most recently in 2022. The initial versions primarily addressed silicosis, but by 1958 the system had been broadened to include all types and profusions of linear markings (Geneva classification).

A major advancement occurred in 1968, when the ILO classification was harmonized with the International Union Against Cancer (UICC) system, thereby extending its scope to encompass all dust-induced pneumoconioses, including irregular opacities characteristic of asbestos-related disease. This classification system was developed to provide a simple, systematic, and reproducible method for codifying radiographic abnormalities associated with pneumoconioses, thereby enabling reliable international comparisons of data and supporting epidemiological studies and research, reflecting the ongoing evolution of scientific methodology and technology toward greater objectivity and precision [12, 5, 13]. Since 1950 ILO delivers Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconioses in Occupational Safety and Health Series (No. 22).

The NIOSH began the B Reader program (named after the Black lung or Coal Workers' X-ray Surveillance Program) in 1974, with the intent to train and certify physicians in the ILO Classification system.

Within this historical trajectory, the emergence of artificial intelligence (AI) and large language models (LLMs) represents a natural continuation of the same goal: to enhance accuracy, reproducibility, and diagnostic reasoning through digital cognition [14, 15]. Building on the foundations of LLMs, which process and generate textual information, the multimodal design allows AI systems to analyze and reason across both textual and visual data, bridging structured diagnostic frameworks such as the ILO classification with image-based interpretation [16]. These generative systems, capable of integrating text

and image interpretation, could represent a valuable supportive tool in radiology within occupational medicine, potentially assisting occupational physicians in applying the classical rule-based framework of the ILO classification more consistently. Previous research has explored the use of machine learning and deep learning algorithms for the automated detection of pneumoconioses on chest radiographs, achieving promising diagnostic accuracy [17]. Most of these models have relied on convolutional neural networks (CNNs) trained on labelled datasets [18, 19, 20]. Although some studies have incorporated the ILO International Classification of Radiographs of Pneumoconioses as a reference standard to improve grading consistency [21, 22], few have fully aligned with its structured coding system [23]. More recently, studies have begun to evaluate also LLMs for general radiographic interpretation tasks [24, 25, 26, 27], but no published work to date has assessed multimodal generative models like ChatGPT or MedGemma using the ILO 2022 classification as a diagnostic framework. This pilot study aimed to comparatively evaluate the diagnostic accuracy of four generative multimodal AI systems (GPT-4o, GPT-5, MedGemma-4B, and MedGemma-27B) in classifying chest radiographs according to the 2022 revised ILO International Classification of Radiographs of Pneumoconioses. The study assessed whether these models could approximate B Reader level performance and identified the domains in which current generative AI systems perform best or remain limited.

2. Methods

2.1. Dataset and Reference Classification

A total of 82 chest radiographs were analysed, retrieved in DICOM (Digital Imaging and Communications in Medicine) format, the international standard to transmit, store, retrieve, print, process, and display medical imaging information. The sample size reflects the pilot nature of the study, which was designed to provide an initial proof-of-concept assessment rather than definitive statistical power. All images were obtained from the NIOSH B Readers syllabus, which includes the official ILO Standard Radiograph Set used for B Reader training and examination. The images were viewed using the NIOSH B Reader software, which enables standardized visualization for the detection of pneumoconioses. Each image had an established reference classification based on consensus readings by certified B Readers, following the ILO guidelines. This classification provides a standardized framework for describing and quantifying radiographic abnormalities caused by occupational dust exposure. Each radiograph is first assessed for technical quality (Grades 1–4), with notation of specific defects such as overexposure, underexposure, poor contrast, improper positioning, artifacts, scapular superimposition, or inadequate lung inflation. Once image quality is deemed adequate, the reader evaluates parenchymal and pleural abnormalities following the ILO's structured format. Parenchymal findings are documented on small opacities, which are classified by shape, size, profusion, and anatomical distribution. Shape and size follow the ILO's six-category coding system: rounded opacities (p, q, r) and irregular opacities (s, t, u), each defined by size ranges illustrated in the ILO standard radiographs (p/s ≤ 1.5 mm; q/t > 1.5–3 mm; r/u > 3–10 mm). The opacities are recorded across upper, middle, and lower lung zones on each side. Profusion is assigned using the twelve ordered ILO subcategories (0/– to 3/+), based on comparison with the official reference radiographs to ensure reproducibility. When present, large opacities are coded as A, B, or C, according to their maximal dimension and the extent of involvement. Pleural abnormalities are coded for pleural plaques, diffuse pleural thickening, and costophrenic angle obliteration. For plaques, the reader records their site (in profile, face-on, diaphragmatic, or chest wall), laterality, extent, thickness category (a, b, or c), and calcification. Diffuse pleural thickening is documented using parallel criteria. Additional standardized ILO symbols allow relevant associated findings such as coalescence (ax), emphysema (em), or evidence of prior tuberculosis (tb) to be noted. In addition to these morphological domains, the ILO reading form includes a structured field related to the clinical relevance of radiographic abnormalities. Specifically, Section 4C requires the reader to indicate whether the worker should be advised to

consult a physician because of findings noted on the radiograph. In this study, this item was analysed as the "clinical decision" category, reflecting the integrated judgement that combines technical, parenchymal, and pleural assessments to support occupational health management. All radiographs were anonymized and used exclusively for research and educational purposes in compliance with NIOSH policies. For each radiograph, a reading sheet consistent with the ILO classification scheme was available and had been previously completed by certified human readers. To facilitate direct comparison between the human reference classifications (provided in the official answer keys) and the AI-generated outputs, all results were transcribed into a structured Excel database. Each row corresponded to one radiograph and contained the reference values. This structure allowed a systematic case-by-case evaluation of concordance between human and AI readings across all classification domains.

2.2. Generative AI Models

Four LLMs were evaluated. One of the most widely known examples of LLMs is ChatGPT, developed by OpenAI and first introduced in 2022. Unlike traditional AI systems, ChatGPT is based on deep learning transformer architectures [28] trained on vast text corpora, enabling the model to analyse, synthesize, and generate natural language. Over time, OpenAI has continuously improved its models [29]. The GPT-4 series, released in March 2023, introduced multimodal capabilities allowing simultaneous interpretation of text and images. The enhanced GPT-4o ("omni"), officially released on May 13, 2024, unified text, vision, and audio processing within a single model architecture, greatly improving reasoning speed and visual comprehension (https://openai.com/index/hello-gpt-4o/). In August 7, 2025, OpenAI announced GPT-5, its latest multimodal model, expanding contextual depth, visual alignment, and multi-turn reasoning performance (https://openai.com/index/introducing-gpt-5/). For this study, the GPT-4o and GPT-5 multimodal models were tested, both capable of processing textual and visual inputs within a unified reasoning framework. Also, two variants of the MedGemma model were employed: MedGemma-4B and MedGemma-27B. MedGemma is an open, multimodal foundation model designed for health-related AI applications and released by Google DeepMind in 2024. The 4B model is a multimodal system (~4 billion parameters) that combines text and image understanding through a SigLIP (Sigmoid Loss for Language–Image Pretraining)-based visual encoder pretrained on medical images (chest radiography, dermatology, pathology). The 27B model (~27 billion parameters) is a larger-scale architecture optimized for medical reasoning; in this study, we specifically used its multimodal version, capable of processing both textual and visual inputs. Both models are openly available for research through the official DeepMind repository (https://deepmind.google/models/gemma/medgemma/). All LLMs evaluations were conducted on September 22, 2025.

2.3. Input Data and Prompting Procedure

All models (GPT-4o, GPT-5, MedGemma-4B and MedGemma-27B) were applied without additional fine-tuning, using a standardized prompt and textual context from the ILO classification guidelines to ensure consistent task interpretation [30]. The primary prompt was formulated as follows:

> *"Assume the role of an occupational physician specialized in chest radiography. You are trained according to the 2022 revised edition of the ILO International Classification of Radiographs of Pneumoconioses. Based on this classification, analyse and report the chest X-ray image provided. Structure your report strictly following the ILO 2022 criteria".*

This prompt was then integrated with the full 2022 revised ILO International Classification guidelines. For all model runs, the following system parameters were used to ensure consistency and reproducibility of outputs. The temperature was set to 0.0, making every response fully deterministic, so identical cases yield identical results regardless of when or how often the prompt is run. The maximum token limit was 4096, which allowed for detailed and complete classification reports. A greedy

decoding strategy (argmax) was applied, and top-p (nucleus sampling) was not utilized due to the zero-temperature setting. These settings ensured that the output for each prompt was stable and not subject to random sampling variation.

The prompt was crafted to maximize consistency and clarity across all model outputs, reflecting practical priorities in occupational radiology. By requiring outputs in a strict JSON schema mapped to the ILO coding system, the structure and format of each model's answer align directly with official criteria, eliminating ambiguity and simplifying downstream analysis. Assigning an explicit role (certified B Reader) instructs the model to adopt the mindset and approach of a trained occupational physician, which guides its interpretation toward domain-specific standards and reasoning.

Inclusion of the full ILO guidelines in the prompt removes the need for the model to rely on memory or incomplete information, ensuring that every classification is made with direct reference to the current standard. This comprehensive context supports both accuracy and fairness in evaluation, allowing each model to perform as if it were consulting the primary source material, just as a human expert would.

2.4. AI-Based Image Analysis

For each image, a new independent conversation was initiated with the LLM. This design simulated the workflow of an occupational physician performing a B Reader evaluation, allowing the model to generate a case-specific report based solely on the information contained in the image and the official ILO framework. By isolating each case in a distinct conversational environment, the study minimized inter-case contamination and ensured that the model's reasoning and descriptive output reflected a *de-novo* evaluation for every radiograph. As outlined above, each classification produced by the models was stored in a structured JSON output for subsequent analysis.

2.5. Evaluation and Comparison

The AI-based classifications were compared with the reference (human) classifications across all 82 images. For each model and each image, the predicted outputs were aligned with the corresponding ILO 2022 reference codes. The following performance metrics were calculated [31, 32, 33]:

- Sensitivity (Recall): proportion of pathological cases correctly identified by the model.
- Specificity: proportion of normal cases correctly identified as negative.
- Precision (Positive Predictive Value, PPV): probability that a case predicted as pathological is truly positive.
- Balanced Accuracy: mean of sensitivity and specificity, giving equal weight to both classes and minimizing class imbalance bias.
- Matthews Correlation Coefficient (MCC): overall correlation between predicted and true classifications, ranging from –1 (inverse correlation) to +1 (perfect agreement).

Additionally, category-specific analyses were conducted according to the ILO coding structure (technical quality, small and large parenchymal opacities, pleural abnormalities, and other findings) to identify the areas of highest and lowest consistency across models. To examine the effect of class imbalance on model behaviour, we also computed the Class Imbalance Ratio for each ILO category, defined as the ratio between the number of negative and positive cases predicted by the model. This metric provides an indication of whether the model tends to over-predict normal findings or pathological findings. Lower ratios indicate a greater tendency to detect positive cases, which is clinically relevant for screening applications. This analysis was exploratory and descriptive, and no inferential statistical testing was performed. All metrics were computed using full numerical precision.

2.6. Statistical Analysis

To assess performance differences among models, we conducted a comparative statistical analysis across all predictions. McNemar's test was applied for pairwise comparisons of classification errors between models, with a χ^2 statistic and a significance threshold set at $p<0.05$. In addition, bootstrap

resampling with 1,000 iterations was used to estimate 95% confidence intervals for balanced accuracy, sensitivity, specificity, PPV and MCC, providing a robust assessment of performance variability. Inter-model agreement was further evaluated using Cohen's kappa (κ), interpreted according to standard qualitative thresholds (from poor to almost perfect agreement). All analyses were performed on paired predictions derived from the same radiographic dataset, ensuring a valid and direct comparison across AI systems. All performance metrics and all statistical analysis were computed using Python (SciPy v1.11).

3. Results

3.1. Overall Model Performance

The comparative performance of the four generative AI models is summarized in Table 1. Among them, MedGemma-27B achieved the highest overall balanced accuracy (70.28%), followed by GPT-4o (69.43%), GPT-5 (65.45%), and MedGemma-4B (60.84%). MedGemma-27B also obtained the highest sensitivity (54.91%) and precision (PPV) (48.34%), while MedGemma-4B recorded the best specificity (86.22%) but the lowest sensitivity (35.45%), indicating a more conservative classification pattern. The MCC ranged from 0.206 (MedGemma-4B) to 0.387 (MedGemma-27B), suggesting moderate diagnostic correlation with the human reference standard.

3.2. Category-Specific Analysis

The analysis by ILO category revealed marked heterogeneity in diagnostic performance across the different models. When assessing technical quality, GPT-4o achieved the highest balanced accuracy (62.0%), slightly outperforming GPT-5 (60.8%) and MedGemma-27B (59.6%), indicating a relatively better ability to recognize image adequacy and exposure-related artifacts. For small parenchymal opacities, MedGemma-27B demonstrated the most consistent performance, with a balanced accuracy of 59.0% and a sensitivity of 80.2%, suggesting a greater capacity to detect subtle parenchymal changes compared with the other models. All models struggled with large opacity classification. GPT-4o, GPT-5, and MedGemma-27B showed balanced accuracies of 50% due to complete failure to detect any large opacities. MedGemma-4B showed improved performance (75% balanced accuracy), but its sensitivity of 50% still indicates limited clinical reliability. In contrast, pleural abnormalities were identified with comparatively higher accuracy, particularly by GPT-5, which achieved the best-balanced accuracy (71.8%). This finding is of clinical

Table 1. Overall diagnostic performance of four generative AI models for ILO pneumoconioses classification. Balanced Accuracy, Sensitivity (Recall), Specificity, Precision (PPV), Matthews Correlation Coefficient (MCC) are reported for each model, along with their corresponding 95% confidence intervals. All metrics are reported as percentages, except the Matthews Correlation Coefficient (MCC), which is reported on its natural scale (−1 to +1), with confidence intervals expressed on the same scale.

Metric	GPT-4o	GPT-5	MedGemma-27B	MedGemma-4B
Balanced Accuracy (95% CI)	69.43% (67.24-71.81)	65.45% (63.29-67.80)	70.28% (68.16-72.69)	60.84 (58.60-63.97)
Sensitivity (Recall) (95% CI)	54.08% (50.00-58.39)	46.31% (42.56-50.40)	54.91% (50.79-59.37)	35.45% (30.81-44.17)
Specificity (95% CI)	84.78% (83.84-85.76)	84.59% (83.62-85.58)	85.66% (84.96-86.40)	86.22% (82.22-87.76)
Precision PPV (95% CI)	46.28% (44.10-48.50)	41.95% (39.69-44.12)	48.34% (45.90-50.71)	31.43% (28.98-34.52)
MCC (95% CI)	0.367 (0.330-0.410)	0.298 (0.257-0.337)	0.387 (0.349-0.426)	0.206 (0.168-0.251)

relevance, as pleural irregularities often represent early or coexisting manifestations of asbestos-related disease. Performance for other abnormalities was uniformly poor across all models, with sensitivity values below 11%, indicating that atypical or less frequent radiological signs are still poorly recognized by current generative AI. Finally, the clinical decision category, which integrates multiple features into an overall interpretative judgment, showed only moderate consistency across systems, with balanced accuracies ranging from 50% to 55%. GPT-5 shows the best performance in this category with 54.83% balanced accuracy. However, sensitivity is concerning (21.43%), indicating potential missed cases. No model clearly outperformed the others in this integrative diagnostic domain, suggesting that while AI systems can approximate human scoring in individual parameters, the holistic synthesis required for final classification remains challenging.

3.3. Class Balance and Overall Reliability

Technical quality yields uniformly low sensitivity and precision, while MedGemma-27B achieves the best recall for small opacities, though with only moderate precision. Performance collapses for large opacities, with all models showing perfect sensitivity but no specificity. Pleural abnormalities represent the only domain with a more favorable precision–recall balance, with GPT-5 and MedGemma-27B performing best.

3.4. Statistical Validation of Model Performance

Pairwise McNemar's tests showed statistically significant differences across all model comparisons (all $p < 0.001$), confirming non-equivalent error distributions. GPT-4o significantly outperformed GPT-5 and MedGemma-4B, while MedGemma-27B significantly outperformed GPT-5 and MedGemma-4B. Bootstrap analysis (1,000 resamples) demonstrated robust stability of performance estimates across all models, with narrow confidence intervals indicating limited variability in resampled performance metrics. Among the evaluated systems, MedGemma-27B achieved the highest balanced accuracy (70.28%, 95% CI: 68.16–72.69) and MCC (0.387, 95% CI: 0.349–0.426), suggesting a more favourable trade-off between sensitivity and specificity. GPT-4o ranked second overall, exhibiting slightly lower sensitivity but comparable specificity, while GPT-5 showed moderate degradation across all metrics, particularly in recall (46.31%, 95% CI: 42.56–50.40). In contrast, MedGemma-4B displayed the lowest discriminative capacity, consistent with its reduced sensitivity and MCC. Cohen's κ agreement scores confirmed high consistency in model decision patterns, with almost-perfect agreement between GPT-4o and MedGemma-27B (κ = 0.9191), and substantial agreement between all remaining pairs (κ = 0.6659–0.8691). These findings show that although MedGemma-27B and GPT-4o were statistically superior, all models tended to produce similar judgment trends, with disagreement mostly concentrated in borderline ILO categories. Statistical analysis is shown in Table 2.

4. Discussion

Crucially, these results reaffirm that the ultimate diagnostic responsibility and clinical judgment must remain with the occupational physician. AI systems should be viewed strictly as supportive tools

Table 2. Statistical comparison between generative AI models on ILO pneumoconioses classification performance.

Model Comparison	McNemar χ^2	p-value	Cohen's κ
GPT-4o vs GPT-5	93.582	<0.001	0.8691
GPT-4o vs MedGemma-27B	11.753	<0.001	0.9191
GPT-4o vs MedGemma-4B	245.263	<0.001	0.7002
GPT-5 vs MedGemma-27B	67.189	<0.001	0.8484
GPT-5 vs MedGemma-4B	33.914	<0.001	0.7101
MedGemma-27B vs MedGemma-4B	141.103	<0.001	0.6659

designed to assist, rather than replace, human expertise in occupational health practice. While models were prompted to follow the structured logic of the ILO classification, we did not evaluate internal reasoning pathways; no claims are made regarding their ability to reproduce human cognitive mechanisms.

Recent studies have explored automated detection of pneumoconiosis using deep learning applied to chest radiographs. Conventional CNN-based approaches have reported strong performance, typically achieving accuracy between ~90% and 98% or AUC values above 0.90 in binary detection tasks and simplified multi-class staging [19,20,34]. Advanced attention-based architectures have further improved feature extraction from lesion-specific regions[35]. Earlier work relying on handcrafted features demonstrated feasibility but was limited by small datasets and non-standard classification schemes [36]. Hybrid pipelines combining lung segmentation with classical machine-learning classifiers showed promising performance when incorporating ILO guidance, although they did not reproduce full ILO scoring [22]. In parallel, early initiatives using LLM-based strategies for pneumoconioses imaging suggest emerging opportunities for multimodal AI in occupational radiology [24], while other studies emphasize the need for robust, scalable AI solutions, especially in low-resource settings [37].

In contrast to these studies, our work is, to our knowledge, the first to systematically evaluate generative multimodal AI systems in a zero-shot setting using the complete revised 2022 ILO classification. This approach reflects realistic clinical deployment scenarios, in which pre-trained models are used without re-training. Certified B Readers remain the gold standard for radiographic diagnosis of pneumoconioses according to ILO guidelines. The NIOSH certification examination using a digital set (revised 2022) requires candidates to classify 72 chest radiographs within four hours covering the full range of technical quality and pneumoconiotic findings, with performance compared against expert reference standards. Despite formal certification, substantial intra- and inter-reader variability has been documented among B Readers [38]. Such variability reflects the perceptual and cognitive demands of applying the full ILO rubric, particularly in borderline profusion grades, subtle pleural abnormalities, and distinction between dust-related and non-occupational changes. As highlighted by the Italian experience with ILO certification courses [39] and consistent with NIOSH data, the mean passing rate for initial B Reader certification between 1987 and 2018 was only about 40% [7]. This confirms the intrinsic perceptual and cognitive complexity of radiographic diagnosis according to the ILO classification system. In this context, the balanced accuracy observed for the AI models in this study, although obtained on a smaller dataset of 82 images, may approximate the performance typically achieved by less experienced human readers. Variability across investigated models may also parallel the inter-operator variability observed among human readers. By potentially reducing inter-reader variability and serving as a 'second reader' or prescreening tool, such AI applications could ultimately enhance the protection of worker health through more consistent and timely disease detection in occupational surveillance programs [7].

4.1. Limitations and Future Perspectives

This pilot study has several limitations that should be acknowledged. The relatively small dataset and the reliance on standardized training images may limit the generalizability of results. Differences in image quality as well as variability in resolution and inspiratory phase, could have influenced both human and AI interpretations. From an algorithmic standpoint, generative AI systems may exhibit hallucinations or reasoning biases depending on their training data which likely constrained diagnostic precision. Methodologically, each image was analysed through a standardized prompt and the 2022 ILO guidelines, simulating a structured B Reader workflow. Although this ensured consistency, future research could test alternative paradigms such as interactive or iterative readings that simulate real-time reasoning. Furthermore, future evaluations should aim to reproduce the operational conditions of the official NIOSH B Reader certification examination, which historically required candidates to classify

approximately 72 chest radiographs within four hours [7]. Replicating these temporal and quantitative parameters would allow a more realistic benchmarking of multimodal AI systems against human performance, assessing not only diagnostic accuracy but also efficiency and cognitive consistency under standardized testing constraints. Expanding the evaluation to additional multimodal architectures beyond GPT and MedGemma could also clarify whether diagnostic reliability depends more on model design or prompt structure. Statistical power was limited by the small sample size, which may have reduced the ability to detect subtle between-model differences. Additionally, no human B Reader re-assessment of the reference labels was performed; the study relied on the original NIOSH-certified readings as the diagnostic gold standard. While this approach reflects real-world reference conditions, subsequent work should include dual-reader adjudication to evaluate model-human agreement in parallel with human-human reproducibility. Further research should include larger and more heterogeneous datasets, explore domain-specific fine-tuning, and assess intra- and inter-reader variability to benchmark AI reproducibility against human performance. In this context, multimodal generative AI could become a valuable tool for supporting training, pre-screening, and comparative research, contributing to greater consistency and harmonization in the radiological diagnosis of pneumoconioses.

5. Conclusion

The ILO Classification of Pneumoconioses and the NIOSH B Reader Program represent decades of coordinated international efforts to standardize occupational lung disease surveillance. While the B Reader certification program has long served to enhance standardization and reduce inter-reader variability, its effectiveness remains incomplete - particularly in complex or borderline cases where subjective judgment plays a substantial role, underscoring the need for transparent and impartial diagnostic processes. In this context, AI integration holds promise not only for improving diagnostic accuracy but also for reducing variability and increasing efficiency across large-scale screening programs.

This study presents the first evaluation of generative AI applied to the 2022 ILO International Classification of Radiographs of Pneumoconioses. By testing four LLMs (GPT-4o, GPT-5, MedGemma-4B, and MedGemma-27B), the research explored the feasibility of using multimodal AI within a standardized diagnostic framework. Although current performance remains limited, the results suggest that these models may begin to approximate the structured reasoning process underlying the ILO system. These findings highlight both the potential and the current limitations of multimodal AI in radiology applied to occupational medicine, particularly in the diagnostic assessment of pneumoconioses and provide a methodological basis for future work aimed at improving accuracy, consistency, and interpretive transparency. As AI systems continue to evolve, the most promising future lies in a human-AI interface – where technology augments the expertise of B Readers, improving both efficiency and consistency while retaining essential human oversight and clinical responsibility. Achieving this vision will require sustained collaboration among AI developers, occupational health professionals, regulatory authorities, and worker communities to ensure that these powerful tools fulfill their ultimate purpose: safeguarding worker health through timely and accurate disease detection, firmly grounded in the principles of occupational medicine, without neglecting ethical aspects.

Funding: This research received no external funding.

Acknowledgments: Intel Corporation Mindshare Program 2024.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: A.B. designed the study and reviewed the manuscript, M.P. carried out statistical analysis, drafted and reviewed the manuscript, A.P. designed the AI system and carried out the experiment, A.Q. and R.L. carried out classification benchmarks, S.D. retrieved datasets, N.M. reviewed the manuscript, and V.T. reviewed the literature.

References

1. Hou X, Wei Z, Jiang X, et al. A comprehensive retrospect on the current perspectives and future prospects of pneumoconiosis. *Front Public Health*. 2025;12:1435840.
2. Naghavi M, Hmwe Hmwe K, Bhoomadevi A, et al. Global burden of 292 causes of death in 204 countries and territories and 660 subnational locations, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. *The Lancet*. 2025, vol. 406.
3. Zhang JS, Xiong X, Ruan, et al. Global burden of pneumoconiosis from 1990 to 2021: a comprehensive analysis of incidence, mortality, and socio-demographic inequalities in 204 countries and territories. *Front Public Health*. 2025;13:1579851.
4. Matyga AW, Chelala L, Chung JH. Occupational Lung Diseases: Spectrum of Common Imaging Manifestations. *Korean J Radiol*. 2023;24:8.
5. Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconioses Revised edition 2022. Available online at: https://www.ilo.org/resource/ilo-international-classification-radiographs-pneumoconioses-1 (Last Accessed 3-10-25).
6. Morgan RH. Proficiency examination of physicians for classifying pneumoconiosis chest films. *AJR Am J Roentgenol*. 1979;132(5).
7. Halldin, CN, Hale JM, Weissman, et al. The National Institute for Occupational Safety and Health B Reader Certification Program – An Update Report (1987 to 2018) and Future Directions. *J Occup Environ Med*. 2019;61(12):1045-1051.
8. Lewis MA. Charles Babbage: Reclaiming an operations management pioneer. *J Oper Manag*. 2007;25(2):248-259.
9. Grattan-Guinness I. Charles Babbage as an Algorithmic Thinker. *IEEE Ann Hist Comput*. 1992;14(3):34-48.
10. Strawn G. Masterminds of Punched Card Data Processing: Herman Hollerith and John Billings. *IT Prof*. 2023;25(6):90-93
11. Ziv L, Nakash M. Behind the Algorithm: International Insights into Data-Driven AI Model Development. *Mach Learn Knowl Extr*. 2025;7(4):122.
12. Y. Hosoda. ILO international classifications of radiographs of pneumoconioses – Past, present and future. *International Classification of HRCT for Occupational and Environmental Respiratory Diseases*, 2005.
13. Muszyńska-Graca M, Dąbkowska B, Brewczyński, PZ. Guidelines for the use of the International Classification of Radiographs of Pneumoconioses of the International Labour Office (ILO): Substantial changes in the currrent edition. *Med Pr*. 2016;67(6):833-837.
14. Buess L, Keicher M, Navab N, et al. From large language models to multimodal AI: A scoping review on the potential of generative AI in medicine. *Biomed Eng*. 2025;15:845–863.
15. Nam Y, Kim DY, Kyung S, et al. Multimodal Large Language Models in Medical Imaging: Current State and Future Directions. *Korean J Radiol*. 2025;26(10):900-923.
16. Soni N, Ora M, Agarwal A, et al. A Review of the Opportunities and Challenges with Large Language Models in Radiology: The Road Ahead. *AJNR Am. Neuroradiol*, 2025;46(7):1292-1299.
17. Sun W, Wu D, Luo Y, et al. A Fully Deep Learning Paradigm for Pneumoconiosis Staging on Chest Radiographs. *IEEE J Biomed Health Inform*. 2022;26(10):5154-5164.
18. Zheng R, Deng K, Jin H, et al. An improved CNN-based pneumoconiosis diagnosis method on X-ray chest film. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019.
19. Zhang L, Rong R, Li Q, et al. A deep learning-based model for screening and staging pneumoconiosis. *Sci Rep*. 11,2201(2021).
20. Zhang Y, Zheng B, Zeng F, et al. Potential of digital chest radiography-based deep learning in screening and diagnosing pneumoconiosis: An observational study. *Medicine*. 2024;103(25):e38478.
21. Yang F, Tang ZR, Chen J, et al. Pneumoconiosis computer aided diagnosis system based on X-rays and deep learning. *BMC Med Imaging*. 2021;(1):2201.
22. Hanampa V, Astete J, Castaneda B , Romero S. Diagnosis of Pneumoconiosis with Machine Learning. in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024.
23. Li X, Liu CF, Guan L, et al. Deep Learning in Chest Radiography: Detection of Pneumoconiosis. *Biomed Enviro Sci*. 2021;34(10):842-845.
24. Song M, Wang J, Yu Z, et al. PneumoLLM: Harnessing the power of large language model for pneumoconiosis diagnosis. *Med Image Anal*. 2024;97:103248.
25. Tian D, Jiang S, Zhang L, Lu X, Xu Y. The role of large language models in medical image processing: a narrative review. *Quant Imaging Med Surg*. 2024,14(1), 1108-1121.
26. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. *Diagn Interv Radiol*. 2024,30(2):80-90.
27. Lanzafame LRM, Gulli C, Mazziotti S, et al. Chatbots in Radiology: Current Applications, Limitations and Future Directions of ChatGPT in Medical Imaging. *Diagnostics*. 2025;15(13):1635.
28. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need, *Advances in Neural Information Processing Systems*, 2017.
29. OpenAI, Achiam J, Adler S, et al. GPT-4 Technical Report. 2024 Available online at: http://arxiv.org/abs/2303.08774 (Last Accessed on 20-10-25).
30. Vilakati S. Prompt engineering for accurate statistical reasoning with large language models in medical research. *Front Artif Intell*. 2025;8:1658316.

31. Devnath L, Fan Z, Luo S, et al. Detection and Visualisation of Pneumoconiosis Using an Ensemble of Multi-Dimensional Deep Features Learned from Chest X-rays. *Int J Environ Res Public Health*. 2022;19(18):11193.
32. Holzmann H, Klar B. Robust performance metrics for imbalanced classification problems. *arXiv preprint*. 2024,2404.07661.
33. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. *20th International Conference on Pattern Recognition*, 2010, pp. 3121-3124.
34. Li X, Xu M, Yan Z, et al. Deep convolutional network-based chest radiographs screening model for pneumoconiosis. *Front Med.* 2024:11:1290729.
35. Alam MS, Wang D, Sowmya A. DLA-Net: dual lesion attention network for classification of pneumoconiosis using chest X-ray images. *Sci Rep*, 2024;14:11616.
36. Okumura E, Kawashita I, Ishida T. Computerized Classification of Pneumoconiosis on Digital Chest Radiography Artificial Neural Network with Three Stages. *J Digit Imaging*. 2017;30(4):413-426.
37. Akhter Y, Ranjan R, Singh R, et al. On AI-Assisted Pneumoconiosis Detection from Chest X-rays. *IJCAI International Joint Conference on Artificial Intelligence*, 2023.
38. Halldin CN, Blackley DJ, Petsonk EL, Laney AS. Pneumoconioses radiographs in a large population of U.S. coal workers: Variability in a reader and B Reader classifications by using the international labour office classification. *Radiology*. 2017;284(3):870-876.
39. Leonori R, Cardona E, Napoli G. Risultati di un'esperienza di formazione sulle linee guida ilo per le pneumoconiosi, *Giornale Italiano Di Medicina Del Lavoro Ed Ergonomia*, 2025, vol. XLVII.

Mattioli 1885

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18756
DOI: 10.23749/mdl.2026.18756

Alcohol Consumption at Work in Construction Workers Employed in Small Italian Companies

Ivan Borrelli[1,*], Luca Fontana[2], Antongiulio Perrotta[3], Paolo Zita[1], Mauro Fedele[2], Ivo Iavicoli[1-4]

[1]Dipartimento di Sicurezza e Bioetica, Catholic University of Sacred Heart, Rome, Italy
[2]Department of Public Health, Section of Occupational Medicine, University of Naples "Federico II", Naples, Italy
[3]Department of Prevention, U.O.S.T. Interdistrettuale Ambienti di Lavoro Ambito Sud, Asl Salerno, Italy
[4]Fondazione Policlinico Universitario A. Gemelli, IRCCS, Rome, Italy

Keywords: AUDIT-C; Construction Workers; Occupational Health Surveillance; Alcohol Use; Risk Perception; Workplace Safety; Injury Prevention

Abstract

Background: *Alcohol consumption is an important occupational risk factor, especially in safety-sensitive sectors such as construction. Alcohol-related psychomotor impairment may increase the risk of workplace injuries and may also affect the safety of third parties. However, data on alcohol-related behaviors, workers' risk perception, and alcohol-focused health surveillance among Italian construction workers are still limited.* **Methods:** *We conducted a cross-sectional survey between September 2023 and June 2024 in 315 construction workers employed in small Italian companies. All participants performed work activities for which local legislation mandates alcohol-use and alcohol-dependence checks as part of occupational health surveillance. Data were collected through an anonymous 27-item questionnaire exploring sociodemographic characteristics, alcohol consumption (AUDIT-C; cut-off ≥5 for men and ≥4 for women), alcohol use at work, knowledge of national regulations, alcohol-related risk perception, and accident/injury indicators. Nonparametric tests and multivariable linear regression were used to identify factors associated with AUDIT-C scores. Logistic regression analyses examined associations between perceived alcohol-related risks and awareness of, and exposure to, alcohol-dependence checks during health surveillance.* **Results:** *The sample was predominantly male (274/315; 86.98%), and 26.03% of workers were aged 46–55 years. The mean AUDIT-C score was 2.85±2.49 (3.13±2.47 in men; 0.95±1.75 in women); 32.7% of workers screened positive, mainly men. Alcohol consumption during the work shift was reported by 1.9% of participants (7.94% occasionally), and during breaks by 5.08% (19.37% occasionally). Most workers were aware of the workplace alcohol ban (90.48%), and 83.49% reported having undergone alcohol-related checks. Higher AUDIT-C scores were associated with alcohol consumption during the work shift and occasional drinking during breaks.* **Conclusions:** *Although average AUDIT-C scores were not high, a relevant subgroup of workers showed risky drinking patterns and some alcohol consumption at work. These findings highlight the need to strengthen prevention strategies and ensure consistent alcohol-related health surveillance in construction settings.*

1. Introduction

Alcohol consumption is currently one of the biggest challenges in public health, both because of its effects on human health and its potential social consequences [1]. From a health perspective, scientific evidence has demonstrated that alcohol consumption is associated with a causal link to over

Received 03.03.2026 – Accepted 08.04.2026
*Corresponding Author: Ivan Borrelli; E-mail: ivan.borrelli@unicatt.it

200 three-digit disease and injury codes in the International Statistical Classification of Diseases and Related Health Problems –10th Revision (ICD-10) [2], the most important of which are malignant neoplasms, mental, behavioral and neurological disorders, cardiovascular and gastrointestinal diseases [3]. In this regard, the Global Burden of Disease Study 2016 showed that, globally, 2.8 million deaths and 1.6%-6% (for females and males, respectively) of total Disability-Adjusted Life Years (DALYs) might be attributed to alcohol use [4]. On the other hand, alcohol consumption represents a significant impediment to the full achievement of several goals (SDGs) of the 2030 Agenda for Sustainable Development such as halve the number of global deaths and injuries from road traffic accidents (SDG 3.6), eliminate all forms of violence against all women and girls in the public and private spheres (SDG 5.2), reduce all forms of violence and related death rates everywhere (SDG 16.1) and end abuse, exploitation, trafficking and all forms of violence against and torture of children (SDG 16.2). Therefore, taking into account these considerations, it is not surprising that the World Health Organization (WHO), pointing out the urgent need for comprehensive strategies and policies to address alcohol-related issues and raise awareness about alcohol-related risks, developed the "Global Alcohol Action Plan 2022–2030" to define and boost the effective implementation of goals and actions to reduce the harmful use of alcohol and related morbidity, mortality and social consequences [3]. Interestingly, the previous document also highlighted how the harms deriving from alcohol consumption can have significant consequences not only for the drinker but also for their social and work relationships (e.g., family members, friends and co-workers) and, in particular, in this context, special emphasis is placed on road traffic injuries [3]. In this context, WHO estimates indicated that in 2016, of the 0.9 million injury deaths attributable to alcohol (accounting for 28.7% of all alcohol-attributable deaths worldwide), 370,000 were due to road injuries, and of these, 187,000 alcohol-attributable deaths occurred among people other than drivers [5]. This aspect is particularly relevant from an occupational medicine perspective, since the psychoactive effects of alcohol (e.g., drowsiness, impaired judgment and decision-making, slowed thinking and reaction time) [6] increase the risk of accidents at work, particularly in certain specific working activities (i.e., professional driving) or work environments (i.e., construction or agricultural sectors) where maintaining optimal perception, attention, and psychomotor coordination is essential [7-10]. In this regard, it is worth noting that, although it is particularly complex and difficult to determine the role of alcohol consumption in the occurrence of work-related injuries, it has been estimated that 10–30% of workplace accidents are alcohol-related, and that problem drinkers face a 2–4 times higher risk of experiencing a workplace accident than non-drinkers [11]. Therefore, although precise figures for the number of these alcohol-attributable adverse events are unavailable, it is widely recognized that alcohol consumption might play a significant causal role, especially in working activities with a higher inherent risk of injury [12-16]. In Italy, occupational health surveillance primarily focuses on protecting workers' health and safety while also considering the potential risks to third parties from alcohol-impaired work-related conduct. However, across Europe, approaches to alcohol-related issues in the workplace remain highly fragmented and heterogeneous, reflecting substantial differences in legislative frameworks, regulatory strategies, and organizational practices [17].

Therefore, it is quite evident that addressing alcohol consumption in the workplace (whether habitual or occasional) clearly constitutes a fundamental pillar of strategies and policies aimed at preventing occupational accidents. For example, in Italy, the framework law on alcohol and alcohol-related problems (Law No. 125/2001) introduced a ban on the consumption and serving of alcoholic beverages and spirits for certain work activities that entail a high risk of workplace accidents and include, among others, the healthcare professions, teaching activities, road vehicle driving, and work at heights exceeding two meters listed in the resolution adopted by the Permanent Conference for Relations between the State and the Regions. In addition, the main framework on occupational health and safety (Legislative Decree 81/08) provides that, in conducting different health-surveillance medical examinations (for those

workers who effectively perform the aforementioned working tasks) the occupational physician (OP) must also assess the alcohol dependence and the use of psychotropic and narcotic substances, in order to protect both workers' health and safety and the safety of third parties. However, screening employees for alcohol use in the workplace is a complex and contentious topic involving moral, ethical, and practical issues. Regarding this latter aspect, although assessing the absence of alcohol consumption is mandatory, it is important to highlight that clear, unequivocal guidance on the conditions and methods for conducting this assessment remains lacking.

In this context, the present study examined risk perception of workplace accidents associated with alcohol use among a sample of construction workers undergoing targeted occupational health surveillance designed to monitor alcohol consumption at work and to identify alcohol dependence-related problems, to pinpoint gaps and critical issues in current practice and consequently to inform improvements to existing measures for the prevention of, and protection from, alcohol-related occupational accidents.

2. Methods

2.1. Setting and Participants

This cross-sectional study was conducted between September 2023 and June 2024 in a convenience sample of Italian construction workers employed by several small companies (10–30 employees) operating in the construction sector in a region of Southern Italy. Eligibility for participation in the survey, and consequent inclusion in the study, required: (i) employment as a construction worker, (ii) performance of at least one work activity included in the resolution of the Permanent Conference for Relations between the State and the Regions, (iii) age ≥18 years and (iv) provision of written informed consent to participate and to the potential future publication of the results in anonymized, aggregated form.

2.2. Recruitment Strategy and Data Collection

A convenience sampling strategy was adopted. Initially, OPs (as well as other professionals involved in the occupational safety and health system) working locally with construction companies (where activities and tasks were performed that, for selected groups of workers, required mandatory assessment of alcohol use) were contacted. After being informed of the study's purpose, sampling strategy, and data collection and analysis procedures, the professionals were asked to assist in recruiting workers from their companies. Subsequently, in the companies identified through this process, workers potentially eligible for the study (according to the above-mentioned inclusion criteria) were invited to participate in the survey. Workers were informed about the aims, study protocol, and methodologies during routine occupational health surveillance or training activities. They were invited to participate voluntarily, and the questionnaire was completed anonymously. A link to the online questionnaire (hosted on the Microsoft platform) was sent to all subjects who expressed interest in participating in the study. In this regard, it should be noted that access to the survey questionnaire was conditional on reading an information cover letter, which reiterated the details previously provided to potential participants regarding the aims of the study and the related data collection and data processing procedures. Likewise, the actual filling in of the questionnaire was contingent upon participants' confirmation of informed consent, both to take part in the study and to the potential future publication of the results in anonymized and aggregated form. All data were collected anonymously, and no information that could identify individual participants was recorded at any stage. The study received ethical approval from the local Ethics Committee for Scientific Research (CERS) (Comitato Etico Territoriale Lazio Area 3, protocol ID 6014/2023) and was conducted under the principles of the Declaration of Helsinki.

2.3. Questionnaire

In line with the aims of the present study, a structured questionnaire was developed to obtain valuable information on issues related to the topic under investigation.

The questionnaire consisted of 5 sections, totaling 27 items. The first section (7 items) investigated the

participants' socio-demographic characteristics. The second part is related to the habits regarding alcohol consumption with the use of the Alcohol Use Disorders Identification Test-Consumption (AUDIT-C, the cut-off level was set to ≥4 for women and ≥5 for men) test [18] and three further questions on alcohol consumption habits at work and the possible presence of alcohol-related problems within the family. In this regard, it should be noted that the AUDIT-C questionnaire was completed assuming that one alcoholic drink corresponds to 12 grams of alcohol, which is equivalent to 330 ml of beer, a standard glass of wine, or a serving of spirits. The third section (5 items) was elaborated to assess knowledge of current national laws and regulations on alcohol consumption in the workplace, whereas the fourth section (3 items) asked about alcohol-related risk perception. The fifth and final section (5 items) investigated any injuries or accidents that may be potentially alcohol related. With the exception of the items included in the short version of the AUDIT tool, the other survey items and questions could not be directly extrapolated or adapted from any existing questionnaire and consequently they were developed specifically for this study, drawing on prior experience, the available literature, and the current Italian regulatory framework for the control of alcohol use in workplaces, with the aim of obtaining reliable data on the variables of interest [19].

2.4. Statistical Analysis

We performed descriptive statistics to characterize the socio-demographic aspects of participants and the characteristics of the study variables, which were presented as frequencies and percentages. To study the correlation between the AUDIT-C score and the categorical variables, we performed nonparametric tests: for the dichotomic variable, we used the Mann-Whitney U test, and for the variable with more than two groups, we used the Kruskal-Wallis test. In a second stage, the variables showing significant associations in bivariate analyses were entered together in a final multiple linear regression. No formal correction for multiple comparisons was applied, as the analyses were primarily exploratory and aimed at identifying potential associations between alcohol consumption patterns and occupational safety indicators. To analyze the collected data, we used STATA 16.

3. Results

3.1. Socio-Demographic Characteristics of Participants

The population surveyed consisted of a final sample of 315 construction workers, predominantly male (n=274, 86.98%). Twelve persons (3.81%) had completed primary school; most participants had completed secondary or lower secondary school (n=216, 68.57%; n=70, 22.22%, respectively), while only 5.40% (n=17) of recruited workers held a university degree. With regard to the specific occupational task (namely, work activities for which health surveillance is also mandated with respect to alcohol consumption), most workers were employed as bricklayers (n=83, 26.35%), followed by general construction laborers (n=72, 22.86%). In this regard, it is noteworthy that approximately two-thirds of the study population had a long length of service (33.97% and 32.38% in the range of 5-15 years and 16-30 years, respectively), a finding of particular relevance given that limited professional experience has been identified as an important determinant of work-related injury [20].

3.2. Alcohol Use Disorders Identification Test-Consumption (AUDIT-C) and Alcohol Consumption Habits in the Workplace

In Table 1, the information obtained from the administration of the AUDIT-C tool and the findings concerning the alcohol consumption habits in the workplace and alcohol-related issues in the family are reported. Overall, in the entire study population, the mean AUDIT-C score was 2.85±2.49, thus indicating that alcohol consumption was not generally indicative of problematic drinking patterns. However, the results also showed a marked gender difference, as the mean score was higher in male than in female workers (3.13±2.47 *vs* 0.95±1.75), suggesting a greater tendency toward hazardous drinking among men. This interpretation would be further

Table 1. Results on alcohol consumption habits (section 2 of the questionnaire).

AUDIT-C mean scores – Entire study population (n=315)		
Reference groups	**Mean score (± SD)**	**AUDIT-C range of scores**
All workers	2.85±2.49	
Male workers (n=274)	3.13±2.47	0-12
Female workers (n=41)	0.95±1.75	
AUDIT-C mean scores – Participants with a positive AUDIT-C score* (n=103, 32.7%)		
Reference groups	**Mean score (± SD)**	**AUDIT-C range of scores**
All workers	5.82±1.43	4-12
Male workers (n=99)	5.84±1.44	5-12
Female workers (n=4)	5.5±1.29	4-12
Alcohol consumption habits in the workplace and alcohol-related issues in the family		
Alcohol consumption during the work shift – n (%)	Yes	6 (1.9%)
	No	284 (90.16%)
	Occasionally	25 (7.94%)
Alcohol consumption during work breaks – n (%)	Yes	16 (5.08%)
	No	238 (75.56%)
	Occasionally	61 (19.37%)
Family history of alcohol-related issues – n (%)	Yes	66 (20.95%)
	No	249 (79.05%)

** A positive AUDIT-C score (indicative of potentially hazardous alcohol consumption) is defined as a score ≥5 for men and ≥4 for women.*

corroborated by data among participants with a positive AUDIT-C score, since the percentage of male workers who tested positive in the AUDIT-C was 32.7%, while this percentage fell to 9.75% among female colleagues (however this gender-specific result should be considered with caution given the substantial imbalance in the number of male (n=274) and female (n=41) participants which may affect the precision and comparability of the data). Interestingly, the average scores observed in the different groups of workers (total, males and females) who tested positive in AUDIT-C were essentially comparable to each other.

Regarding the workplace context, alcohol consumption during the work shift was rarely reported (1.9% and 7.94% of participants answered "yes" and "occasionally", respectively) but higher frequencies were observed during work breaks (5.08% and 19.37% for "yes" and "occasionally" options, respectively). This finding has been particularly surprising because Italian Law No. 125/2001 explicitly prohibits both the consumption and the serving of alcoholic beverages in the workplace for the working activities associated with a higher risk of occupational injuries (such as those performed by the workers included in this study). At the same time, this data is also particularly worrying as a non-negligible proportion of workers reported alcohol consumption in the workplace, which may have implications for safety, job performance, and the risk of work-related accidents.

Finally, concerning the last item in this section (that evaluated whether participants had relatives with known alcohol-related issues), 20.95% of the sample reported a family history of alcohol-related problems, which might represent a vulnerability or hyper-susceptibility (if only psychological) factor

for a greater tendency toward possible risky alcohol consumption.

3.3. Knowledge of Current National Laws About Alcohol Consumption in the Workplace

The items in this section of the survey explored participants' level of knowledge regarding the main Italian laws regulating alcohol consumption in the workplace (Law No. 125/2001) as well as the related control measures implemented through occupational health surveillance as provided for by Legislative Decree 81/08. In this context, previous information on occasional alcohol consumption in the workplace appears even more remarkable in light of the fact that most workers (90.48%) were aware that Italian Law No. 125/2001 prohibits alcohol consumption at work, and a similarly high proportion (89.84%) knew that occupational health surveillance may include controls to check alcohol dependence.

Importantly, 83.49% of workers were actually subjected (during health surveillance medical examinations) to tests aimed at assessing alcohol dependence, which effectively demonstrates that such control measures are not merely theoretical but are commonly implemented in practice by OPs. Nevertheless, it is perplexing that a sizeable proportion of workers (16.51%) did not report undergoing such checks, given that the professional tasks they perform are legally expected to be covered by alcohol-related controls.

With regard to participation in information and training courses on alcohol-related workplace issues, the majority of participants stated that they had received training on this topic (87.30%), even if the issue concerning the risk to third parties (i.e., involvement of and consequences for other people such as family members, friends and co-workers) seems to have been insufficiently addressed (22.22% not addressed).

3.4. Alcohol-Related Risk Perception

The results of the survey fourth section (Table 2) are focused on the workers' alcohol-related risk perception. In this regard, concerning the measures that could be taken by the employer if alcohol-related issues were detected, most respondents reported that they would expect disciplinary sanctions to be imposed (52.38%) or, alternatively, the implementation of supportive interventions (i.e., specific counseling) (19.05%), but a non-negligible proportion of workers (21.27%) were unable to answer this question. With respect to the level of agreement among workers regarding the usefulness of prevention programs in improving alcohol-related issues in the workplace, the data showed that an overwhelming majority of workers were "Strongly agree" and "Somewhat agree" (24.13% and 57.78%, respectively) with this sentence. Similar trends were also observed for the perceived safety implications, with response percentages strongly skewed toward complete or partial agreement that alcohol consumption in the workplace might actually pose a risk to one's own safety, to that of colleagues but also of third parties, thus suggesting a broad awareness on this topic. Indeed, this perception is further supported by the data of the latest question, which unequivocally indicated that most of participants perceive that alcohol consumption may play a role in adversely influencing a wide range of outcomes across occupational, family, social, and individual domains.

3.5. Road Traffic Accidents and Work-Related Injuries

In Table 3 are reported the anamnestic information regarding road traffic accidents and work-related injuries together with possible indicators of their potential association with alcohol use (e.g., driving licence restrictions or alcohol test results). In the study sample, nearly one-fifth of participants (17.78%) had been victims of a work-related injury in the three years preceding the survey, while an even smaller percentage of workers (9.52%) had been involved in a road traffic accident in the previous five years. Concerning the alcohol-related indicators, although relatively few workers had their driving licences restricted (13.33%) or had tested positive on a breathalyzer (n=28), these data still suggest further attention, as targeted prevention strategies and policies (also in the workplace) could help in minimizing the risk of alcohol-related road traffic accidents and/or work-related injuries.

Table 2. Results on alcohol-related risk perception.

Items	Response options - n (%)				
How would the company respond if alcohol-related issues were identified?	**Disciplinary actions**	**Assistance programs**	**The company wouldn't take any action**	**I don't know**	
	165 (52.38%)	60 (19.05%)	23 (7.30%)	67 (21.27%)	
Indicate how much you agree with the following statement: prevention programs improve alcohol-related issues in the workplace	**Strongly agree**	**Somewhat agree**	**Somewhat disagree**	**Strongly disagree**	
	76 (24.13%)	182 (57.78%)	38 (12.06%)	19 (6.03%)	
Indicate how much you agree with the following statement: alcohol consumption in the workplace can be a risk factor for …	**Strongly agree**	**Somewhat agree**	**Somewhat disagree**	**Strongly disagree**	
your own workplace safety	223 (70.79%)	68 (21.59%)	16 (5.08%)	8 (2.54%)	
your co-workers' safety	208 (66.03%)	82 (26.03%)	19 (6.03%)	6 (1.90%)	
others' workplace safety	226 (71.75%)	59 (18.73%)	22 (6.98%)	8 (2.54%)	
In your opinion, how much can alcohol consumption affect the following conditions?	**Very much**	**Considerably**	**Fairly**	**Slightly**	**Not at all**
Road traffic accidents	205 (65.08%)	56 (17.78%)	18 (5.72%)	6 (1.90%)	30 (9.52%)
Workplace injuries	191 (60.63%)	59 (18.73%)	39 (12.38%)	12 (3.81%)	14 (4.44%)
Decreased attention	157 (49.84%)	91 (28.89%)	43 (13.65%)	11 (3.49%)	13 (4.13%)
Reduced work performance	147 (46.67%)	83 (26.35%)	60 (19.04%)	11 (3.49%)	14 (4.44%)
Disciplinary actions	171 (54.29%)	77 (24.44%)	38 (12.06%)	18 (5.71%)	11 (3.49%)
Onset of disease	192 (60.95%)	72 (22.86%)	32 (10.16%)	9 (2.86%)	10 (3.17%)
Domestic violence	211 (66.98%)	51 (16.19%)	34 (10.79%)	7 (2.22%)	12 (3.81%)
Workplace disputes	175 (55.56%)	55 (17.46%)	58 (18.41%)	13 (4.13%)	14 (4.44%)
Poor workplace relationships	156 (49.52%)	81 (25.71%)	50 (15.87%)	12 (3.81%)	16 (5.08%)
Increased costs for companies	153 (48.57%)	63 (20.00%)	51 (16.9%)	21 (6.67%)	27 (8.57%)
Workplace absenteeism	164 (52.06%)	64 (20.32%)	51 (16.9%)	18 (5.71%)	18 (5.71%)

3.6. Associations Between AUDIT-C Scores and Socio-Demographic Characteristics, Occupational Role and Alcohol-Related Workplace Behaviors

The linear regression analysis examined the potential links between AUDIT-C scores and various variables surveyed (Table 4). Most socio-demographic characteristics and occupational features were not significantly related to AUDIT-C data. However, a statistically significant connection was found for secondary school qualification, and overall, the analysis indicated a possible trend toward

Table 3. Results on road traffic accidents and work-related injuries.

	Response options - n (%)	
Items	**Yes**	**No**
Have you experienced any work-related injuries in the past three years?	56 (17.78%)	259 (82.22%)
Have you experienced any road traffic accidents in the past five years?	30 (9.52%)	285 (90.48%)
In the past five years, have you been subject to any measures affecting your driving licence such as points deducted (due to alcohol-related issues)?	42 (13.33%)	273 (86.67%)
Have you ever undergone a breathalyzer test following a police roadside check?	85 (26.98%)	230 (73.02%)
… if yes, was the result positive for alcohol?	28 (32.94%)	57 (67.06%)

Table 4. Linear regression analysis with mean scores AUDIT-C.

Variables	**Coeff.**	**St. Err.**	**CI 95%**	***p* – value**	
Gender	Male *	0	-	-	-
	Female	-0.363	0.345	[-1.043, 0.316]	0.293
Age	18-25 *	0	-	-	-
	26-35	0.322	0.409	[-0.483, 1.128]	0.432
	36-45	0.072	0.527	[-0.966, 1.11]	0.891
	46-55	-0.463	0.625	[-1.693, 0.767]	0.46
	>55	-0.275	0.744	[-1.74, 1.19]	0.712
Educational Level	Primary School *	0	-	-	-
	Lower secondary school	0.482	0.557	[-0.614, 1.577]	0.388
	Secondary school	1.155	0.553	[-0.67, 2.244]	**0.038**
	Degree	1.22	0.7	[-0.159, 2.598]	0.083
Specific job role	Bricklayer *	0	-	-	-
	Construction laborer	0.153	0.327	[-0.490, 0.796]	0.64
	Truck driver	-0.033	0.362	[-0.745, 0.680]	0.928
	Site manager	0.199	0.393	[-0.574, 0.972]	0.613
	Tower crane operator	-0.437	0.501	[-1.423, 0.548]	0.383
	Other	-0.225	0.326	[-0.867, 0.417]	0.491
Length of service	<5 *	0	-	-	-
	5-15	0.291	0.408	[-0.512, 1.094]	0.476
	16-30	1.15	0.546	[0.074, 2.225]	**0.036**
	>30	0.66	0.716	[-0.749, 2.068]	0.357
Alcohol consumption during the work shift	No *	0	-	-	-
	Yes	3.334	0.931	[1.501, 5.167]	**<0.001**
	Occasionally	0.571	0.487	[-0.387, 1.529]	0.242
Alcohol consumption during work breaks	No *	0	-	-	-
	Yes	0.918	0.576	[-0.217, 2.052]	0.113
	Occasionally	0.93	0.384	[0.174, 1.687]	**0.016**

Variables	Coeff.	St. Err.	CI 95%	*p* – value	
Are you aware that there is an Italian law (Law No. 125/2001) prohibiting alcohol consumption in the workplace?	Yes *	0	-	-	-
	No	0.739	0.455	[-0.157, 1.636]	0.106
During the health surveillance medical examination, were you checked for alcohol dependence?	Yes *	0	-	-	-
	No	0.487	0.462	[-0.423, 1.396]	0.293
Have you received information and training courses on alcohol-related workplace issues?	Yes *	0	-	-	-
	No	-0.486	0.401	[-1.275, 0.303]	0.226
During courses, was the risk to third parties associated to alcohol consumption at work addressed?	Yes *	0	-	-	-
	No	0.927	0.337	[0.264, 1.590]	**0.006**
How would the company respond if alcohol-related issues were identified?	The company wouldn't take any action *	0	-	-	-
	I don't know	-1.068	0.436	[-1.927, -0.210]	**0.015**
	Assistance programs	-0.616	0.443	[-1.488, 0.257]	0.166
	Disciplinary actions	0.327	0.435	[-0.530, 1.184]	0.453
Indicate how much you agree with the statement: prevention programs improve alcohol-related issues in the workplace	Strongly disagree *	0	-	-	-
	Somewhat disagree	-0.661	0.537	[-1.718, 0.396]	0.220
	Somewhat agree	-0.486	0.488	[-1.447, 0.475]	0.321
	Strongly agree	-1.02	0.515	[-2.034, -0.006]	**0.049**

** reference category used in the regression models.*

higher AUDIT-C scores with increasing education level. Additionally, regarding professional covariates, workers with 16-30 years of service had significantly higher AUDIT-C scores compared to colleagues with less than 5 years of seniority, while the other groups showed small, non-significant differences.

Interestingly, a strong association (p<0.001) was found between AUDIT-C scores and alcohol consumption habits, as participants who reported drinking alcohol during the work shift or occasionally during breaks had significantly higher AUDIT-C scores (β=3.334; 95% CI 1.501, 5.167, p<0.001 and β=0.93; 95% CI 0.174, 1.687, p=0.016, respectively). A possible disconnect between risk communication and risky drinking behaviors is suggested by the statistically significant association observed in those who reported that information and training courses on alcohol-related workplace issues did not address the risk to third parties. Finally, the analysis supports the hypothesis that workers who had lower AUDIT-C scores consider workplace initiatives and measures to prevent alcohol-related problems to be particularly important.

3.7. Perceived Alcohol-Related Risk and Health Examinations for Alcohol Dependence

The logistic regression analysis data showed that the workers' perception of the alcohol contributing role to road traffic accidents was strongly associated

with greater awareness that health surveillance medical examinations may include (for specific working tasks) control for alcohol dependence (Table 5).

Similar results were also observed with regard to the self-reported execution of alcohol-dependence targeted checks during the medical examination. Moreover, the same statistically significant correlation was found for the reduced work performance. In this context, a higher level of awareness would have been expected to be also associated with a greater perception of alcohol influence on workplace injuries, while instead the findings would suggest the opposite.

4. Discussion

Intending to ensure the highest level of workers' health and safety, the definition and implementation of multidisciplinary policies and strategies that integrate health promotion, accident prevention, and measures to control, prevent, mitigate, and address alcohol use and dependence in the workplace still remain one of the most important challenges in occupational medicine today. The magnitude of this problem is underscored by recent International Labour Organization figures reporting that every year, globally, 395 million workers suffer work-related injuries, of which (again according to ILO estimates) between 10 and 30% are more or less directly connected to alcohol consumption [11, 21]. These data are unsurprising, given that the potential contribution of risky alcohol consumption to various adverse outcomes, such as work-related injuries or road traffic and domestic accidents, is well established [22-24], and it aligns with alcohol's well-recognized effects on psychomotor and cognitive functions, including impaired coordination, perception, attention, and judgment [25]. It is important to highlight that, in terms of increased risk of injury at work, these adverse effects of alcohol are especially significant in activities where workers are exposed to hazards and safety-critical tasks (e.g., operating heavy machinery, working at heights, professional driving, using power tools, or handling dangerous equipment such as explosives), where maintaining psycho-physical fitness (and thus avoiding alcohol-related impairment) is essential for performing duties safely [26-27].

In this context, the findings of the present cross-sectional survey, conducted on Italian construction workers undergoing targeted health surveillance to also monitor alcohol dependence, offer a detailed overview of workplace alcohol use behaviors, risky alcohol consumption, perceptions of alcohol-related risk, health surveillance activities, and the relationships between these dimensions. These findings provide noteworthy insights for discussion and help identify priority areas for intervention to further reduce the risk of alcohol-related accidents at work. First, an interesting initial finding emerged from the analysis of AUDIT-C data, since even though the average score in the study population was not particularly high (thus not suggesting a problem of possible hazardous alcohol consumption), nearly one-third of participants (almost all male workers) tested positive. This is a highly important and serious concern, especially in the construction sector where, even an occasional impairment while performing safety-sensitive tasks and activities (e.g., working at height, operating heavy machinery, driving) could cause disproportionate and potentially catastrophic consequences for workers' health and safety [28]. Therefore, it seems clear that there is an urgent need to improve, strengthen, and expand the application of context-specific workplace policies and strategies. These should combine periodic health surveillance medical examinations with evidence-based education, health promotion programs, and targeted support interventions aimed at discouraging hazardous alcohol use and reducing alcohol-related risks, particularly among male workers.

Furthermore, it should also be considered that, although the AUDIT-C tool is one of the few validated tests for assessing alcohol consumption in occupational settings, its reliability might be affected in non-anonymous contexts such as workplace health surveillance medical exams. Indeed, concerns about professional consequences and implications can lead to response bias and intentional underreporting, increasing the risk of false-negative results and limiting its usefulness in occupational history [29-30]. In this regard, the occupational safety and health system should be involved in promoting honest disclosure during AUDIT-C screening within health surveillance to reduce underreporting

Table 5. Associations (Logistic Regression analysis) between perceived alcohol-related risk and (A) awareness of and (B) exposure to health surveillance medical examinations aimed at assessing alcohol dependence.

A - Do you know that health surveillance may include (for specific working tasks) control for alcohol dependence?		**Odds Ratio**	**St. Err.**	**95% CI**	***p* – value**
In your opinion, how much can alcohol consumption affect the following conditions?	Road traffic accidents	1.738	0.464	[1.030, 2.934]	**0.038**
	Workplace injuries	0.579	0.142	[0.359, 0.935]	**0.025**
	Decreased attention	0.487	0.201	[0.218, 1.092]	0.081
	Reduced work performance	1.548	0.746	[0.603, 3.979]	0.364
	Disciplinary actions	0.533	0.152	[0.305, 0.932]	**0.027**
	Onset of disease	1.851	0.647	[0.934, 3.671]	0.078
	Domestic violence	0.546	0.229	[0.241, 1.240]	0.148
	Workplace disputes	0.9	0.346	[0.423, 1.913]	0.784
	Poor workplace relationships	1.007	0.336	[0.524, 1.938]	0.982
	Increased costs for companies	1.003	0.344	[0.512, 1.966]	0.993
	Workplace absenteeism	1.375	0.501	[0.673, 2.808]	0.383
B - In this regard, during the health surveillance medical examination, were you checked for alcohol dependence?					
In your opinion, how much can alcohol consumption affect the following conditions?	Road traffic accidents	4.024	1.422	[2.013, 8.043]	**<0.01**
	Workplace injuries	0.882	0.218	[0.544, 1.431]	0.611
	Decreased attention	0.42	0.141	[0.218, 0.810]	**0.01**
	Reduced work performance	2.436	0.947	[1.137, 5.219]	**0.022**
	Disciplinary actions	0.463	0.118	[0.280, 0.764]	**0.003**
	Onset of disease	1.121	0.313	[0.648, 1.937]	0.684
	Domestic violence	0.404	0.127	[0.218, 0.747]	**0.004**
	Workplace disputes	1.174	0.299	[0.713, 1.933]	0.529
	Poor workplace relationships	0.956	0.284	[0.534, 1.710]	0.878
	Increased costs for companies	0.882	0.267	[0.488, 1.596]	0.679
	Workplace absenteeism	0.912	0.285	[0.495, 1.682]	0.769

Multivariable logistic regression model. Odds ratios represent the change in the odds of the outcome associated with a one-unit increase in each predictor measured on a five-point Likert scale.

and enable effective prevention strategies. Possible approaches to explore include integrating screening within a clearly non-punitive prevention and support framework, training occupational health professionals in stigma-free communication, and most importantly, providing workers with clear and comprehensive information about legal requirements regarding alcohol consumption at work and the potential consequences of non-compliance. These considerations become even more relevant when analyzing the results concerning alcohol consumption habits at the workplace. Although the vast majority of respondents reported being aware that alcohol consumption is prohibited during certain safety-sensitive activities and that health surveillance may include alcohol-dependence assessments, an alarmingly high percentage of workers (many with higher average AUDIT-C scores, Table 4) admitted to drinking alcohol—either regularly or occasionally—during work hours, especially during breaks. This data, which cannot be easily explained by low awareness of legal obligations or poor perception of alcohol-related risks (both of which are notably high in the survey, Tables 3 and 4), certainly

warrants further investigation and calls for a careful review of how workers are trained and informed about this issue in the workplace. Knowledge alone is clearly insufficient; occupational-specific guidelines and workplace organizational culture are likely to play a key role [31-32].

Regarding the usefulness of prevention programs in addressing alcohol-related issues in the workplace, the linear regression analysis model highlighted that workers who strongly agreed with this view also had lower AUDIT-C scores. This observation may support the hypothesis that prevention initiatives are more likely to be endorsed by workers who are already at lower risk, while higher-risk individuals may be less convinced, less engaged, or more defensive. In practical terms, this suggests that universal and generic prevention messages or tools, although important, might be inadequate for these groups and should be complemented with tailored approaches that reduce stigma, clarify confidentiality limits, and emphasize support rather than punishment.

This study benefits from focusing on a safety-relevant workforce and integrating behavioral measures (AUDIT-C and workplace drinking habits), legal/organizational knowledge, risk perceptions, and accident-related indicators into a single, coherent questionnaire. However, several limitations should be considered. The cross-sectional design prevents causal inferences or conclusions about the direction of associations. Participants were recruited voluntarily during routine occupational health surveillance in small construction companies, representing a convenience sample with possible selection bias, as workers with particular attitudes toward alcohol or health issues may have been more likely to participate. The study was conducted within a specific occupational sector and region, with a predominantly male sample, which may limit generalizability. Additionally, as with most self-administered questionnaires on sensitive behaviors like alcohol use, data may be affected by recall and social desirability biases, potentially underreporting risky behaviors. Some estimates showed wide confidence intervals, especially in the regression analyses (Tables 4 and 5), likely due to limited observations in certain response categories and the variability of perception-based variables. These should be interpreted cautiously, although the overall associations were consistent. Importantly, the questionnaire was completed anonymously and privately to minimize reporting bias and protect privacy.

5. Conclusion

Although the overall AUDIT-C profile did not indicate widespread excessive alcohol consumption, the combined evidence of a significant at-risk subgroup, reported alcohol use during breaks and occasional consumption during work shifts, incomplete health surveillance coverage, and a perception of alcohol-related risk focused mainly on traffic accidents (but not equally on occupational accidents) highlights clear opportunities for strengthening prevention. Improving the consistency and scope of health surveillance, better aligning training with occupational injury prevention (including third-party risks), and adopting supportive, non-stigmatizing engagement strategies can help reduce alcohol-related harm in this high-risk work environment.

With regard to future research, longitudinal studies should clarify the temporal relationships between training, health surveillance, medical examinations, changes in risk perception, and consequent behaviors related to alcohol consumption. At the same time, the evaluation of targeted, context-specific interventions (particularly those addressing alcohol consumption during breaks and third-party safety) will be useful for determining which models, strategies, and policies work best in the construction industry.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Local Ethics Committee for Scientific Research (CERS) (Comitato Etico Territoriale Lazio Area 3, protocol ID 6014/2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: I.I. and I.B. contributed to the design and implementation of the research, I.B., L.F. and A.P. contributed to the analysis of the results, and all the authors contributed to the writing of the manuscript.

References

1. World Health Organization (WHO). Global status report on alcohol and health and treatment of substance use disorders. Geneva: World Health Organization, 2024a. Licence: CC BY-NC-SA 3.0 IGO
2. Rehm J, Gmel GE Sr, Gmel G, et al. The relationship between different dimensions of alcohol use and the burden of disease-an update. *Addiction*. 2017;112: 968-1001. Doi: 10.1111/add.13757
3. World Health Organization (WHO). Global alcohol action plan 2022–2030. Geneva: World Health Organization, 2024b. CC BY-NC-SA 3.0 IGO.
4. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2018;392(10152):1015-1035. Doi: 10.1016 /S0140-6736(18)31310-2.
5. World Health Organization (WHO). Global status report on alcohol and health 2018. Geneva: World Health Organization, 2018.
6. von Wartburg JP. On the importance of the psychotropic effects of alcohol. *Drug Alcohol Depend*. 1990; 25: 135-139. Doi: 10.1016/0376-8716(90)90052-g.
7. Gutiérrez-Fisac JL, Regidor E, Ronda E. Occupational accidents and alcohol consumption in Spain. *Int J Epidemiol.* 1992; 21:1114-1120. Doi: 10.1093/ije/21. 6.1114.
8. Stallones L, Xiang H. Alcohol consumption patterns and work-related injuries among Colorado farm residents. *Am J Prev Med*. 2003;25(1):25-30. Doi: 10.1016 /s0749-3797(03)00096-5.
9. Flannery J, Ajayi SO, Oyegoke AS. Alcohol and substance misuse in the construction industry. *Int J Occup Saf Ergon.* 2021; 27: 472-487. Doi: 10.1080/10803548 .2019.1601376.
10. Bufano P, Poma AM, Frumento S, et al. Meta-analytic study on substance intake and work-related accidents calls for attention to bio-psycho-social factors. *Neurosci Biobehav Rev.* 2023;144:104987. Doi: 10.1016/j.neubiorev .2022.104987.
11. International Labour Organization (ILO). Drug and alcohol abuse prevention programmes in the maritime industry (A manual for planners) (Revised). United Nations International Drug Control Program (UNDCP), International Labour Office. Geneva, 2001.
12. Webb GR, Redman S, Hennrikus DJ, Kelman GR, Gibberd RW, Sanson-Fisher RW. The relationships between high-risk and problem drinking and the occurrence of work injuries and related absences. *J Stud Alcohol.* 1994; 55:434-446. Doi: 10.15288 /jsa.1994.55.434.
13. McNeilly B, Ibrahim JE, Bugeja L, Ozanne-Smith J. The prevalence of work-related deaths associated with alcohol and drugs in Victoria, Australia, 2001-6. *Inj Prev*. 2010;16:423-428. Doi: 10.1136/ip.2010.027052.
14. Ramirez M, Bedford R, Sullivan R, et al. Toxicology testing in fatally injured workers: a review of five years of Iowa FACE cases. *Int J Environ Res Public Health*. 2013; 10: 6154-6168. Doi: 10.3390/ijerph10116154.
15. Institute of Alcohol Studies. *Alcohol in the workplace*. IAS Briefing. London: IAS; 2020.
16. Chikritzhs T, Livingston M. Alcohol and the Risk of Injury. *Nutrients*. 2021;13:2777. Doi: 10.3390/nu13082777.
17. Borrelli I, Santoro PE, Gualano MR, et al. Alcohol Consumption in the Workplace: A Comparison between European Union Countries' Policies. *Int J Environ Res Public Health.* 2022;19:16964. Doi: 10.3390 /ijerph192416964.
18. van Gils Y, Franck E, Dierckx E, van Alphen SPJ, Saunders JB, Dom G. Validation of the AUDIT and AUDIT-C for Hazardous Drinking in Community-Dwelling Older Adults. *Int J Environ Res Public Health.* 2021; 18: 9266. Doi: 10.3390/ijerph18179266. PMID: 34501856; PMCID: PMC8431181
19. Borrelli I, Gualano MR, Rossi MF, et al. Alcohol consumption in healthcare workers and risk of workplace injury: a case-control study. *Discov Sustain.* 2023;4:22. Doi: 10.1007/s43621-023-00137-7
20. Chau N, Dehaene D, Benamghar L, et al. Roles of age, length of service and job in work-related injury: a prospective study of 63,620 person-years in female workers. *Am J Ind Med.* 2014;57:172-83. Doi: 10.1002 /ajim.22259
21. International Labour Organization (ILO). A call for safer and healthier working environments. International Labour Office. Geneva, 2023.
22. Karjalainen K, Blencowe T, Lillsunde P. Substance use and social, health and safety-related factors among fatally injured drivers. *Accid Anal Prev.* 2012;45:731-736. Doi: 10.1016/j.aap.2011.09.044.
23. Kurzthaler I, Wambacher M, Golser K, et al. Alcohol and benzodiazepines in falls: an epidemiological view. *Drug Alcohol Depend*. 2005; 79:225-230. Doi: 10.1016 /j.drugalcdep.2005.01.012.
24. Dong M, Lee YY, Cha JS, Huang G. Drinking and driving: A systematic review of the impacts of alcohol consumption on manual and automated driving performance. *J Safety Res.* 2024;89:1-12. Doi: 10.1016 /j.jsr.2024.01.006.
25. Cherpitel CJ, Bond J, Ye Y, et al. Multi-level analysis of causal attribution of injury to alcohol and modifying effects: Data from two international emergency room projects. *Drug Alcohol Depend.* 2006;82:258–68. Doi: 10.1016 /j.drugalcdep.2005.10.002

26. Foster CA, Dissanaike SD. Prevalence and consequences of positive blood alcohol levels among patients injured at work. *J Emerg Trauma Shock*. 2014; 7:268-73. Doi: 10.4103/0974-2700.142748.
27. Roche AM, Lee NK, Battams S, Fischer JA, Cameron J, McEntee A. Alcohol use among workers in male-dominated industries: A systematic review of risk factors. *Saf Sci.* 2015;78:124–141. Doi: 10.1016/j.ssci.2015.04.007
28. Sawicki, Szóstak M. Impact of Alcohol on Occupational Health and Safety in the Construction Industry at Workplaces with Scaffoldings. *Appl Sci.* 2020;10: 25-30. Doi: 10.3390/app10196690
29. Schulte B, O'Donnell AJ, Kastner S, Schmidt CS, Schäfer I, Reimer J. Alcohol screening and brief intervention in workplace settings and social services: a comparison of literature. *Front Psychiatry*. 2014;5:131. doi: 10.3389/fpsyt.2014.00131.
30. Kilo S, Hofmann B, Eckert E, Göen T, Drexler H. Evaluation of biomarkers assessing regular alcohol consumption in an occupational setting. *Int Arch Occup Environ Health.* 2016;89:1193-1203. Doi: 10.1007/s00420-016-1155-1.
31. Ames GM, Grube JW, Moore RS. Social control and workplace drinking norms: a comparison of two organizational cultures. *J Stud Alcohol.* 2000;61:203-219. doi: 10.15288/jsa.2000.61.203.
32. Elling DL, Almquist YB, Wennberg P, Sundqvist K. Effects of a multi-component alcohol prevention program in the workplace on hazardous alcohol use among employees. *BMC Public Health*. 2023;23:1420. Doi:10.1186/s12889-023-16150-4.

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18002
DOI: 10.23749/mdl.2026.18002

Cognitive Effects of Aluminum Exposure in Cement Factory Workers: A Mini-Mental State Examination (MMSE) Assessment*

Fatma Bozdağ[1,†], Sultan Pinar Çetintepe[1], Mümüne Bozdağ Kiraz[2], Sevgi Suna Karatay Rassmus[3], Mustafa Necmi İlhan[1]

[1]Department of Public Health, Division of Occupational Medicine, Gazi University, Ankara, Türkiye
[2]Department of Neurology, Tarsus State Hospital, Mersin, Türkiye
[3]MSc Student Alumni, University of Bremen, Bremen, Germany

Keywords: Occupational Aluminum Exposure; Cognitive Function; MMSE; Biomarker

Abstract

Background: *Aluminum (Al) is a widely encountered heavy metal with known neurotoxic effects. Occupational exposure, particularly in industrial settings, may impair cognitive functions. This study aimed to evaluate the relationship between Al exposure and cognitive function.* **Methods:** *A retrospective study was conducted at Gazi University Faculty of Medicine Hospital, Occupational Diseases Outpatient Clinic, between December 5, 2024, and January 5, 2025. The exposed group consisted of 20 male cement factory workers with elevated urinary Al expressed as a function of creatinine, and the control group included 40 age-matched males without occupational Al exposure (1:2 matching). Cognitive status was assessed using the Mini Mental State Examination (MMSE). Urinary Al levels were measured by inductively coupled plasma mass spectrometry (ICP-MS). Statistical analyses were performed with SPSS 29.0.* **Results:** *The mean MMSE score was significantly lower in the exposed group compared with controls (24.3±3.7 vs. 28.5±2.3, p<0.001). Subscale scores for orientation, attention/calculation, recall, and language were also lower in exposed workers. All such workers had elevated aluminum (mean 42.3±21.4 mcg/g creatinine). Urinary Al was positively correlated with working duration (r=0.453, p=0.045) and negatively correlated with MMSE (r=-0.486, p=0.030) and orientation scores (r=-0.494, p=0.027).* **Conclusion:** *Workers occupationally exposed to aluminum exhibited significantly lower cognitive performance than non-exposed controls. Higher urinary Al levels were associated with poorer cognitive outcomes, suggesting neurotoxic effects of aluminum and underscoring the importance of preventive strategies and cognitive monitoring in exposed populations.*

1. Introduction

Aluminum (Al) is a metal found in nature combined with other elements such as silicon and oxygen, and obtained from aluminum-containing minerals [1]. Many people are exposed to aluminum through media and routes such as water, fruits and vegetables, processed foods, some medications and vaccines, and food-heating and storage devices (such as aluminum foil) [2]. Occupational exposure occurs during the production, processing, welding, and recycling of aluminum. It is stated that occupational

Received 29.09.2025 – Accepted 08.04.2026
†Corresponding Author: Fatma Bozdag; E-mail: fatmabozdag7@gmail.com
* Part of this work was presented as an oral presentation entitled "Cognitive Effects of Aluminum Exposure in Cement Factory Workers: A Mini-Mental State Examination (MMSE) Assessment" at the 13th International Symposium on Biological Monitoring (ISBM 13), Milan, 2025.

exposure to aluminum results from workers' contact, especially through inhalation [4]. The absorption of aluminum is affected by chemical form, particle size, and the use of dietary chelators (such as citric or lactic acid) [3]. Blood/serum and urine analyses can be performed to monitor aluminum exposure [5].

Studies have shown that aluminum has adverse effects on many systems, including the respiratory, cardiovascular, and gastrointestinal systems [6, 7]. Monographs published by IARC have identified a relationship between aluminum smelting plants and bladder cancer [4]. The nervous system is the most sensitive system to aluminum toxicity [8, 9]. Al exposure is thought to be associated with Alzheimer's, Parkinson's, Multiple Sclerosis, and many other neurological diseases [10-12]. In the meta-analysis conducted by Soleimani et al., it was determined that aluminum exposure increased the risk of Alzheimer's disease by 2.451 times (CI: 0.569-4.334, p=0.011) [13]. When Alzheimer's patients were compared with control patients, aluminum levels were higher in the cerebrospinal fluid, serum, and brain tissue of Alzheimer's patients [14].

It has been reported that aluminum exacerbates oxidative damage due to its pro-oxidant properties and may therefore be a risk factor in ALS [15]. In a review by Killin et al., it was reported that increased aluminum levels in water increase the risk of dementia [16]. Neurotoxic effects of aluminum are explained through various biochemical pathways. As a prooxidant, it can cause oxidative stress and lead to the production of free oxygen molecules. It can also cause mitochondrial dysfunction [17]. As a cholinotoxic agent, it can alter acetylcholinesterase activity [18].

Additionally, considering the effects of aluminum on cognitive functions, workers with occupational exposure were found to have poorer performance in attention, reaction time, and working memory. It has been determined that elevated plasma aluminum levels can predict declines in cognitive performance. In a study of shift workers at an aluminum smelting plant in Norway, it was reported that alertness levels were lower after night shifts than after 3-4 consecutive day shifts [20]. In a cross-sectional study, aluminum levels were higher in cognitively impaired workers than in controls, and gray matter volumes in the left caudate and bilateral hippocampus were lower and positively correlated with aluminum levels [9]. In a study of workers in an aluminum factory, the risk of cognitive impairment increased 2.24-fold when plasma aluminum levels were high compared with low [8].

Based on the literature, our study aimed to evaluate the effects of aluminum exposure on cognitive function. For this purpose, we aimed to determine the relationship between aluminum intoxication and cognitive status assessment in patients referred to us from the cement production industry with suspicion of aluminum intoxication and in the control group.

2. Methods

This study was designed as a retrospective cross-sectional study. The study population consisted of workers who attended the Occupational Diseases Outpatient Clinic of Gazi University Faculty of Medicine between December 5, 2024, and January 5, 2025.

Patients referred by occupational physicians or other physicians with suspected occupational diseases are evaluated at our outpatient clinic. During this period, all clinical and laboratory evaluations were performed as part of routine occupational health assessments, and the data were recorded in institutional medical records. No data were collected specifically for research purposes at that time. After obtaining ethical approval from the Gazi University Rectorate Ethics Committee (No: 2025-1337; date: 29/07/2025), these existing records were retrospectively reviewed and analyzed.

2.1. Occupational Exposure

The cement factory examined in this study produces ready-mixed cement through the stages of raw material preparation, mixing, firing, and grinding. The cement produced in the factory contains limestone and silica; hydraulic hardening occurs through calcium silicates and aluminates. Raw materials, including limestone, silica, and clay-based minerals, contain aluminum-bearing compounds that are essential components of Portland cement chemistry. Workers in the exposed group in this study

are potentially exposed to aluminum-containing cement dust, especially during raw material handling, mixing, kiln operation, and clinker grinding. Occupational exposure occurs mainly through inhalation of cement dust containing aluminum compounds. According to available records, workers were not consistently able to use personal protective masks, resulting in continuous inhalation exposure to cement dust containing aluminum during their shifts.

2.2. Population

2.2.1. Exposed Group

Twenty male cement factory workers with elevated urinary aluminum levels or aluminum/creatinine ratios were identified from medical records. To ensure that cognitive impairment was specifically attributable to aluminum exposure and not to polymetallic contamination or other confounders, inclusion criteria required that participants have normal levels of blood lead (<8.5 μg/dL), and urinary chromium (<5 μg/L), mercury (<9 μg/L), and manganese (<18.3 μg/L). Additionally, serum vitamin B12 levels had to be within the normal range (197-771 pg/mL). All cement factory workers who agreed to participate and met these criteria were included in the study.

2.2.2. Control Group

Forty age-matched male individuals without aluminum exposure were identified from medical records and met the same heavy metal and vitamin B12 criteria. Exposed – control matching was performed 1:2.

2.2.3. Exclusion Criteria

Participants with elevated blood lead (>8.5 μg/dL), serum chromium (>5 μg/L), blood mercury (>9 μg/L), or blood manganese (>18.3 μg/L) levels, decreased B12 (<197 pg/mL) or history of physician-diagnosed epilepsy, dementia, learning disabilities, and depression; as well as those who declined to participate, were excluded from the study. These thresholds correspond to the reference intervals provided by the accredited clinical laboratory performing the analyses and were used only as screening criteria to exclude participants with markedly elevated levels of other heavy metals.

All clinical evaluations and laboratory measurements had been performed as part of routine clinical practice and were obtained from medical records. Urine and blood samples had been collected at the end of 8-hour work shifts as part of routine occupational exposure assessment. Sample analysis was performed using an ICP-MS device. Urine samples were collected in sterile containers at the end of the shift. Heavy metal levels were analyzed using inductively coupled plasma mass spectrometry (ICP-MS, Agilent 7700X, USA). The argon used in the plasma torch had a purity of over 99.999% and was supplied by HABAS (Kocaeli, Turkey). In the analyses, 65% Suprapur nitric acid (Merck, Darmstadt, Germany) and an aqueous multi-element standard solution (High-Purity Standards, Charleston, USA) were used. Calibration was performed with an aqueous multi-element standard solution. The method was validated by analyzing certified reference materials (Seronorm Trace Elements, Billingstad, Norway). Urine aluminum levels were determined by the ICP-MS method. According to international standards, reference values for aluminum in spot urine are <35 μg/L or ≤14 mcg/g creatinine in spot urine samples [21].

The Mini-Mental State Examination (MMSE) was used to assess participants' cognitive status. Developed by Folstein et al., the MMSE consists of eleven questions and is scored out of 30. It has five subscales: orientation (10 points), registration (3 points), attention and calculation (5 points), recall (3 points), and language and copying (9 points). Orientation assesses an individual's awareness of temporal and spatial context, while immediate memory is assessed by recording and repeating presented words. Attention and calculation measure concentration and working memory using tasks such as serial subtraction or spelling. Recall examines short-term memory by testing delayed recall of previously presented words. Language and copy assess a range of linguistic functions, including naming, repetition, comprehension, reading, and writing,

as well as the ability to accurately reproduce a simple geometric shape. The Turkish translation was made by Güngen et al. [22, 23].

Diagnosis had been established based on clinical findings, laboratory measurements, and Mini Mental State Examination (MMSE) results, as documented in medical records. This process was conducted during a multidisciplinary team meeting involving psychologists, neurologists, and occupational disease specialists.

2.3. Statistical Analysis

Descriptive statistics for the study are presented as numbers, percentages, means, and standard deviations. Normality of continuous variables was assessed using the Shapiro-Wilk test. Chi-square tests, Mann-Whitney U tests, and Spearman's rank correlation were used to assess relationships between groups. Multiple linear regression analysis (enter method) was performed with MMSE score as the dependent variable, and urinary aluminum, aluminum/creatinine ratio, age, education group, and smoking status as independent variables. Statistical analyses were performed using SPSS 29.0. P<0.05 was considered significant.

3. Results

Our study was conducted with 20 patients with aluminum exposure and 40 controls. The mean age of the exposed subjects was 42.99±8.3 years, and the control group was 39.7±8.9 years (p=0.174). The exposed group had longer working experience (p=0.013). Seventeen (85.0%) of the exposed group and 30 (75%) of the control group were married (p=0.375), while 32 (80.0%) of the control group and 11 (55.0%) of the exposed subjects had a high school education or higher (p=0.037). The smoking rate was higher in workers with Al exposure than in those without (p=0.013). However, the number of cigarettes smoked was similar in smokers (p=0.367). In the exposed group, aluminum concentrations ranged from 13.59 to 67.09 μg/L (mean 30.1±14.9 μg/L), while aluminum values were above the normal range in three participants. The corresponding values expressed as a function of creatinine ranged from 20.98 to 100.52 mcg/g (mean 42.3±21.4 mcg/g); thus, all values expressed as a function of creatinine were above the upper reference limit. The corresponding values of the control group were 1.87 ± 0.84 μg/L and 2.05 ± 1.01 μg/g creatinine, respectively, both p<0.001 vs. exposed workers (Table 1).

Table 1. Characteristics of the study group.

	Exposed group		Control		
	Mean±SD/n	**Percentage (%)**	**Mean±SD/n**	**Percentage (%)**	**p**
Age	42.9±8.3		39.7±8.9		0.174*
Working Time	22.4±8.6		16.6±9.8		0.013*
Marital Status					0.375
Single	3	15.0	10	25.0	
Married	17	85.0	30	75.0	
Education Level					**0.037**
Primary and Secondary School	9	45.0	8	20.0	
High School and Above	11	55.0	32	80.0	
Smoking Status					**0.013**
No	3	15.0	19	47.5	
Yes	17	85.0	21	52.5	
Cigarettes (Packs/year)	18.6±9.6		26.5±20.5		0.367*
Urinary Al (μg/L)	30.1±14.9		1.87±0.84		<0.001*
Urinary Al (mcg/g creatinine)	42.3±21.4		2.05±1.01		<0.001*

** Mann Whitney U Test, SD: Standard Deviation.*

The MMSE score was 24.3±3.7 in workers with aluminum exposure, while it was 28.5±2.3 in those without (p<0.001). Furthermore, scores in the orientation, attention, calculation, recall, and language subscales were also lower in the exposed group (Table 2).

There was a positive correlation between the urinary aluminum and total working hours (r=0.453, p=0.045) and a negative correlation with both the orientation score (r=-0.494, p=0.027) and MMSE (r=0.486, p=0.030). However, such correlations were significant only when urinary Al was expressed as a function of creatinine, whereas no correlation was observed between raw urinary Al (unadjusted) and MMSE scores (Table 3). The latter were also sensitive to education, with those with primary and

Table 2. Cognitive assessment of the study group.

	Exposed group		Control		
	Mean±SD	Median (min-max)	Mean±SD	Median (min-max)	p
Orientation	8.8±1.4	9.0(5-10)	10.0±0.0	10.0 (10-10)	<0.001
Registration	3.0±0.0	3.0(3-3)	2.9±0.3	3.0 (1-3)	0.313
Attention and calculation	1.9±2.1	1.0 (0-5)	4.0±1.6	5.0 (0-5)	<0.001
Recall	2.4±0.9	3.0 (0-3)	2.8±0.5	3.0 (1-3)	0.003
Language and copying	8.3±1.0	9.0 (6-9)	8.8±0.5	9.0 (7-9)	0.032
Total score (MMSE)	24.3±3.7	24.0(17-30)	28.5±2.3	30.0 (23-30)	<0.001

Mann Whitney U Test, SD: Standard Deviation.

Table 3. Correlation of factors affecting MMSE scores.

		Aluminum	Aluminum/Creatinine	Working Hours
MMSE	r	0.289	-0.486	-0.197
	p	0.216	**0.030**	0.132
Orientation	r	0.199	-0.494	-0.208
	p	0.400	**0.027**	0.111
Attention and calculation	r	0.271	-0.420	-0.170
	p	0.248	0.065	0.195
Recall	r	0.099	-0.247	-0.236
	p	0.679	0.294	0.069
Language and copying	r	0.206	-0.057	-0.225
	p	0.384	0.813	0.046
Working Hours	r	0.054	0.453	
	P	0.823	**0.045**	1

**Spearman Correlation.*

Table 4. Predictors of MMSE Score in Multivariate Analysis.

	B	SE	β	t	p	95% CI
Aluminum (μg/L)	−0.015	0.021	−0.067	−0.689	0.494	−0.057 – -0.028
Aluminum (μg/g creatinine)	−0.050	0.015	−0.332	−3.260	0.002	−0.081 – −0.019
Age (Years)	0.001	0.030	0.002	0.023	0.981	−0.059 – 0.061
Education Group	2.292	0.317	0.604	7.240	<0.001	1.657 – 2.927
Smoking Status	-0.659	0.544	-0.093	-1.211	0.213	-1.751 – 0.432
Constant	20.582	1.906		10.798	<0.001	16.761 – 24.403

B: unstandardized coefficient, SE: Standard Error, β: standardized coefficient. CI: Confidence Interval.

secondary school degrees showing lower scores than those with high school education or above (21.3±2.3 vs. 26.7±2.7, p<0.001). There were no correlations between age and MMSE, orientation, memory, attention and calculation, recall, and language.

Multiple linear regression analysis was conducted to identify factors associated with MMSE. Urinary Al, when expressed as a function of creatinine, was negatively and independently associated with MMSE; in contrast, the education group showed a strong positive association with the mental score. Age, smoking status, and urinary aluminum concentration alone were not significantly associated with MMSE (Table 4).

4. Discussion

To our knowledge, this is the first study comparing the cognitive status of participants working in the cement industry in Türkiye who have high aluminum levels with that of those not exposed to aluminum. Our study demonstrated cognitive impairment in workers with Al exposure. There are studies in the literature evaluating the occurrence of cognitive impairment due to environmental and occupational exposure to Al. In the meta-analysis, urine aluminum levels in controls averaged 9.61 μg/L (SD = 4.95) and 11.32 μg/g creatinine (SD = 8.13), compared with exposed workers at 87.20 μg/L (SD = 82.92) and 54 μg/g (SD = 38.28), respectively [19]. In this study, cement workers showed markedly higher urinary aluminum and aluminum-to-creatinine ratios than controls, indicating a clear exposure gradient consistent with occupational aluminum exposure. Although environmental air monitoring data were unavailable, biological monitoring demonstrated a clear exposure gradient between cement workers and controls, supporting the conclusion of occupational aluminum exposure. The results of the meta-analysis show decreased performance in processing speed, working memory, attention, and reaction time among workers occupationally exposed to aluminum compared with controls.

In a study by Zhang et al. of aluminum mine workers, the MMSE score was 21.34 among workers with aluminum exposure and 22.95 among those without. The MMSE score was lower in those with aluminum exposure. Furthermore, in an analysis adjusted for age, gender, and education level, the risk of cognitive impairment was reported to be 6.7 times higher in those with aluminum exposure [24]. In the study by Xing et al., the MMSE score was reported as 27.93 ± 1.91 in the Al-exposed group and 28.62 ± 1.25 in the non-exposed group. The score was lower in the exposed group [25]. In the study by Lu et al., the MMSE score in workers exposed to aluminum was reported as 26.13±2.57, which was lower than that of the control group [26]. Additionally, some studies have found that MMSE scores are lower among workers with aluminum exposure than among those without [27, 28]. In our study, the MMSE score was 24.3±3.7 in workers with occupational Al exposure and 28.5±2.3 in non-exposed workers, which is consistent with findings reported in the literature. In addition, aluminum normalized to creatinine was elevated in all exposed participants. This observation suggests that creatinine-adjusted urinary aluminum may represent a more

informative indicator of occupational aluminum exposure when spot urine samples are used, as it helps account for inter-individual variability in urine dilution. Overall, these findings suggest that increased aluminum exposure may be associated with lower cognitive performance.

In evaluating MMSE subscales, Xing et al. found that orientation, recall, attention, calculation, and language scores were similar between the aluminum- and non-aluminum-exposed groups. Visual ability scores were significantly higher in the control group [25]. In the study by Lu et al., scores for orientation, recall, calculation ability, and language skills were lower among workers exposed to aluminum than in the control group [26]. In Yang et al.'s study, recall, calculation ability, and language skill scores were also found to be lower in the group with the highest aluminum level than in the group with the lowest [29]. Similarly, in our study, orientation, recall, attention, calculation, and language scores were lower among workers exposed to aluminum. Furthermore, orientation scores were negatively affected by an increased aluminum-to-creatinine ratio. These results support the negative consequences of aluminum exposure on cognitive functions.

A study by Qui et al. of aluminum factory workers reported lower MMSE scores among workers with higher plasma aluminum levels than among those with lower levels [30]. In a study by Lu et al., which followed workers at an aluminum factory for 2 years, MMSE scores decreased as workers' aluminum levels increased during the follow-up [31]. Additionally, many studies have found a negative correlation between blood plasma aluminum levels and MMSE scores [27-29, 32-34]. When studies evaluating blood aluminum level and MMSE score are evaluated, Pan et al reported an increase of -0.005 times (p=0.001), Abdala et al reported an increase of -0.348 times (p=0.017 CI:-0.632,-0.065), and Zhang et al. reported an increase of -0.630 times (p<0.001) [32, 34, 35]. Contrary to these studies, some studies reported no relationship between aluminum levels and MMSE scores [36-38]. In our study, there was a negative correlation between increased aluminum-to-creatinine ratio and MMSE scores. This supports the notion that increased aluminum exposure will lead to a decline in cognitive function.

Individual factors such as age, education level, and length of service also influence MMSE scores. A study by Pan et al. reported that MMSE scores were negatively affected by age (-0.061, p=0.001) and length of service (-0.058, p=0.001), while education level had no effect [32]. Lu et al. reported that age and education level affect [26]. Abdalla et al. determined that the study period increased the score by -0.044 times (p=0.032, CI:-0.084 to -0.004) [34]. Shang et al. reported that the year was positively correlated in their study [27]. Our study also found that higher education levels served as a protective factor for cognitive function, while working hours had a negative impact. Higher education levels are known to protect cognitive function.

5. Conclusion

In conclusion, this study shows a significant and independent association between aluminum exposure and cognitive performance, as measured by the Mini-Mental State Examination (MMSE), when urinary aluminum levels are expressed as a function of creatinine. Although no correlation was observed when aluminum concentrations in spot urine were analyzed, a significant association was detected when aluminum normalized to creatinine in spot urine was used, highlighting the importance of correcting for urine dilution in biomonitoring studies. Despite the known limitations of the MMSE, the persistence of this link after controlling for key confounding factors suggests that increased aluminum exposure may negatively affect cognitive function beyond the protective effects of cognitive reserve. Consistent with aluminum's known neurotoxic properties, these findings support its potential role as an occupational factor contributing to cognitive impairment and emphasize the importance of considering aluminum exposure among at-risk workers. Additionally, given its low cost, reproducibility, and ease of use, the MMSE could be a practical tool for occupational physicians to monitor cognitive decline among exposed workers.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee

of Gazi University Rectorate Ethics Committee (protocol code: 2205-1337, date of approval: 29/07/2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Acknowledgments: We would like to thank the cement factory employees and the workplace physician who contributed to our study.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: F.B. and S.P.C. contributed to the conception, design, implementation and writing of the manuscript, M.B.K. and S.S.K.R contributed to the analysis of the results, M.N.İ. contributed to the conception and design.

References

1. ATSDR. Aluminum. Available: https://wwwn.cdc.gov/TSP/substances/ToxSubstance.aspx?toxid=34 (Last Accessed on 17 August 2025).
2. Alasfar RH, Isaifan RJ. Aluminum environmental pollution: the silent killer. *Environ Sci Pollut Res.* 2021;28(33):44587-44597. Doi: 10.1007/s11356-021-14700-0.
3. ToxGuide™ for Aluminum CAS 7429-90-5. U.S. Department of Health and Human Services Public Health Service Agency for Toxic Substances and Disease Registry. Available online: https://www.atsdr.cdc.gov/toxguides/toxguide-22.pdf (Last Accessed on 17 August 2025).
4. Chemical Agents and Related Occupations Volume 100 F A Review of Human Carcinogens. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Available at: https://www.ncbi.nlm.nih.gov/books/NBK304416/ (Last Accessed on 17 August 2025).
5. Riihimäki V, Aitio A. Occupational exposure to aluminum and its biomonitoring in perspective. *Crit Rev Toxicol.* 2012;42(10):827-853. Doi: 10.3109/10408444.2012.725027.
6. Igbokwe IO, Igwenagu E, Igbokwe NA. Aluminium toxicosis: a review of toxic actions and effects. *Interdiscip Toxicol.* 2019;12(2):45-70. Doi: 10.2478/intox-2019-0007.
7. Wesdock JC, Arnold IMF. Occupational and environmental health in the aluminum industry: key points for health practitioners. *J Occup Environ Med*. 2014;56:S5-S11. Doi: 10.1097/JOM.0000000000000071.
8. Meng H, Wang S, Guo J, et al. Cognitive impairment of workers in a large-scale aluminium factory in China: a cross-sectional study. *BMJ Open*. 2019;9:e027154. Doi: 10.1136/bmjopen-2018-027154.
9. Meng H, Liu B, Lu X, et al. Alterations of gray matter volume and functional connectivity in patients with cognitive impairment induced by occupational aluminum exposure: a cross-sectional study. *Front Neurol.* 2025;15: 1500924. Doi: 10.3389/fneur.2024.1500924
10. Inan-Eroglu E, Ayaz A. Is aluminum exposure a risk factor for neurological disorders? *J Res Med Sci.* 2018;23:51. Doi: 10.4103/jrms.JRMS_921_17
11. Nabi M, Tabassum N. Role of environmental toxicants on neurodegenerative disorders. *Front Toxicol.* 2022;4:837579. Doi: 10.3389/ftox.2022.837579.
12. Dey M, Singh RK. Neurotoxic effects of aluminium exposure as a potential risk factor for Alzheimer's disease. *Pharmacol Rep.* 2022;74(3):439-450. Doi: 10.1007/s43440-022-00353-4.
13. Soleimani H, Dehghani S, Abolli S, et al. Environmental aluminum exposure and Alzheimer's disease risk: evidence from a systematic review and meta-analysis. *Ecotoxicol Environ Saf.* 2025;302:118759. Doi: 10.1016/j.ecoenv.2025.118759.
14. Virk SA, Eslick GD. Aluminum levels in brain, serum, and cerebrospinal fluid are higher in Alzheimer's disease cases than in controls: a series of meta-analyses. *J Alzheimers Dis*. 2015;47(3):629-638. Doi: 10.3233/JAD-150193.
15. Bocca B, Forte G, Oggiano R, et al. Level of neurotoxic metals in amyotrophic lateral sclerosis: a population-based cross-sectional study. *J Neurol Sci.* 2015;359(1-2):11-17. Doi: 10.1016/j.jns.2015.10.023.
16. Killin LO, Starr JM, Shiue IJ, Russ TC. Environmental risk factors for dementia: a systematic review. *BMC Geriatr.* 2016;16(1):175. Doi: 10.1186/s12877-016-0342-y.
17. Kawahara M, Tanaka K, Kato-Negishi M. Neurotoxicity of aluminum and its link to neurodegenerative diseases. *Metallomics Res.* 2021;1(1):rev-47. Doi: 10.11299/metallomicsresearch.MR202104.
18. Kumar V, Gill KD. Aluminium neurotoxicity: neurobehavioural and oxidative aspects. *Arch Toxicol.* 2009;83(11):965-978. Doi: 10.1007/s00204-009-0455-6.
19. Vlasak T, Dujlovic T, Barth A. Aluminum exposure and cognitive performance: a meta-analysis. *Sci Total Environ*. 2024;906:167453. Doi: 10.1016/j.scitotenv.2023.167453.
20. Pedersen M, Bast-Pettersen R, Matre D, et al. Alertness after night shifts among workers in the aluminium industry. *J Occup Environ Med*. 2025;67(8):e568-e575. Doi: 10.1097/JOM.0000000000003412.
21. Deutsche Forschungsgemeinschaft (DFG). *MAK and BAT Values List 2025: Maximum Concentrations and Biological Reference Values in the Workplace.* Report 61 of the Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area. Düsseldorf: *German Medical Science*; 2025. Doi: 10.34865/mbwl_2025_deu.

22. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12(3):189-198. Doi: 10.1016/0022-3956(75)90026-6.
23. Güngen C, Ertan T, Eker E, Yaşar R, Engin F. Standardize mini mental test'in Türk toplumunda hafif demans tanısında geçerlik ve güvenilirliği. *Türk Psikiyatri Derg*. 2002;13(4): 273-281.
24. Zhang T, He F, Lin S, et al. Does aluminum exposure affect cognitive function? a comparative cross-sectional study. *PLoS One.* 2021;16(2):e0246560. Doi: 10.1371 /journal.pone.0246560.
25. Meng HX, Wang SS, Lu XT, et al. Cognitive status of electrolytic aluminum workers: a cross-sectional study using cognitive screening tests. *Biomed Environ Sci*. 2019;32(11):869-873. Doi: 10.3967/bes2019.110.
26. Lu XT, Liang RF, Jia ZJ, et al. Effect of aluminum exposure on cognitive function in electrolytic workers and its influential factors. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi.* 2013;31(2):13-116.
27. Shang N, Zhang P, Wang S, et al. Aluminum-induced cognitive impairment and PI3K/Akt/mTOR signaling pathway involvement in occupational aluminum workers. *Neurotoxicol Res.* 2020;38(2):344-358. Doi: 10.1007 /s12640-020-00230-z.
28. Polizzi S, Pira E, Ferrara M, et al. Neurotoxic effects of aluminium among foundry workers and Alzheimer's disease. *Neurotoxicology*. 2002;23(6):761-774. Doi: 10.1016 /S0161-813X(02)00097-9.
29. Yang X, Yuan Y, Lu X, et al. The relationship between cognitive impairment and global DNA methylation decrease among aluminum potroom workers. *J Occup Environ Med*. 2015;57(7):713-717. Doi: 10.1097/JOM.0000 000000000474.
30. Qiu HY, Ren P, Li R, et al. Association between H3K4me3/BDNF and the cognitive function of workers occupationally exposed to aluminum. *Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi.* 2016;34(12):900-904. Doi: 10.3760/cma.j.issn.1001-9391.2016.12.005.
31. Lu XT, Xu SM, Zhang YW, et al. Longitudinal study of the effects of occupational aluminium exposure on workers' cognition. *Chemosphere*. 2021;271:129569. Doi: 10.1016/j.chemosphere.2021.129569.
32. Pan B, Zhou Y, Li H, et al. Relationship between occupational aluminium exposure and histone lysine modification through methylation. *J Trace Elem Med Biol.* 2020;61:126551. Doi: 10.1016/j.jtemb.2020.126551.
33. Zhang J, Hao Y, Wang Y, et al. Relationship between the expression of TNFR1-RIP1/RIP3 in peripheral blood and cognitive function in occupational Al-exposed workers: a mediation effect study. *Chemosphere*. 2021;278:130484. Doi: 10.1016/j.chemosphere.2021.130484.
34. Abdalla YS, Mishriky AM, Abdel-Halim AW, et al. Occupational exposure to aluminum and cognitive performance. *Egypt J Occup Med.* 2024;48(2);75-89. Doi: 10.21608/ejom.2024.274936.1330.
35. Zhang Y, Huan J, Gao D, et al. Blood pressure mediated the effects of cognitive function impairment related to aluminum exposure in Chinese aluminum smelting workers. *Neurotoxicology*. 2022;91:269-281. Doi: 10.1016 /j.neuro.2022.05.017.
36. Xu SM, Zhang YW, Ju XF, et al. Cross-sectional study based on occupational aluminium exposure population. *Environ Toxicol Pharmacol*. 2021;83:103581.
37. Wang S, Xue Y, Zhang J, et al. Interaction between aluminum exposure and ApoEε4 gene on cognitive function of in-service workers. *Chemosphere*. 2023;323:138282. Doi: 10.1016/j.chemosphere.2023.138282.
38. Zhao X, He C, Wang S, et al. The association between blood lymphocyte NMDAR, group I mGluRs and cognitive function changes in occupationally aluminum-exposed workers and verification in rats. *J Trace Elem Med Biol.* 2022; 69, 126875. Doi: 10.1016/j.jtemb.2021.126875. mb.2021.126875. Epub 2021 Oct 14. PMID: 34673477.

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18730
DOI: 10.23749/mdl.2026.18730

Self-Reported Non-Auditory Effects of High Sound Pressure Levels Exposure in Academic Musicians in Uruguay

Alice Elizabeth González[1,*], Fernando Tomasina[2], Julián Ortiz Umaña[1], Lady Carolina Ramírez[1], Bettina Tellechea[2], Myriam Adriana Pisani[2]

[1]Department of Environmental Engineering, Faculty of Engineering, Universidad de la República, Uruguay
[2]Academic Unit of Occupational Health, Faculty of Medicine, Universidad de la República, Uruguay

Keywords: Musicians' Health; Sound Exposure Effects; Symphonic Musicians; Extra-Auditory Effects of Noise; Anamnesis Data

Abstract

Background: *While hearing loss is the most emblematic effect of exposure to high sound pressure levels, it is not neither the first nor the only one. This article explores self-reported effects of exposure to high sound pressure levels among a population of 306 academic musicians (singers, choir members and instrumentalists) in Uruguay.* **Methods:** *A special anamnesis form was prepared for use in interviews of each of the participants. The collected information was anonymized and processed. Some interesting results were found, both through direct processing and PCA analysis.* **Results:** *Most participants reported several non-auditory effects. These were classified by sex, age, and role (singer or instrumentalist, voice type, or played instrument).* **Conclusions:** *Among the reported effects, the most prevalent were muscle contractures, fatigue/tiredness, difficulty sleeping, and noise sensitivity, followed by arthralgia and headaches. The most frequent effects in women were muscle contracture (86%), followed by tiredness/fatigue, noise susceptibility, and tinnitus. The most frequent effects in men were muscle contracture (68%), followed by tinnitus, noise susceptibility, and irritability after musical activity. It was possible to establish some "effects profiles" according to instrument and gender. For example, woodwind players experience muscle contracture, tinnitus, post-musical activity irritability, noise susceptibility, and decreased performance, while soprano singers have a higher prevalence of muscle contracture, tiredness/fatigue, noise sensitivity, difficulty sleeping, arthralgia, digestive disorders, headache, tinnitus, and vertigo. They are the second most affected category of musicians, according to their effect profile.*

1. Introduction

Noise is a pervasive workplace risk, evident in construction and manufacturing but also in call centers, schools, hospitals, and the arts, particularly among music professionals [1-3]. As an occupational pollutant, noise threatens health not only through hearing loss but also through a wide range of extra-auditory effects [2]. This has transformed a workplace issue into a public health problem, often referred to as recreational and environmental hearing loss [4].

Research highlights non-auditory impacts, including cognitive impairment (especially in children), sleep disorders, and cardiovascular conditions [5]. Subtle but continuous exposure can lead to discomfort, frustration, and stress-related symptoms, which contribute to cardiovascular disease [6]. According to the CCOHS, non-auditory effects include

*Corresponding Author: Alice Elizabeth González; E-mail: elizabet@fing.edu.uy
Arrived 20.02.2026 – Accepted 20.03.2026

physiological responses (muscle tension, respiratory changes, altered heartbeat) and performance-related issues [7].

Cognitive performance is also at risk; workers exposed to noise exhibit poorer results on mental tasks [8]. While short-term exposure might improve simple activities, it hinders complex ones like professional music performance. Musical intelligence is notably adversely affected [9]. Hypertension can appear at levels as low as 75 dBA—well below those typically considered safe for hearing [10]. Chronic exposure acts as a psychosocial stressor even when it doesn't damage the inner ear [5].

Occupational health services play a strategic role in worker protection [11]. Described extra-auditory effects include endocrine changes (elevated cortisol/catecholamine levels), elevated blood pressure, and altered heart rate. Impacts also extend to the central nervous system (changes in EEG), sleep quality, respiratory rate, and visual function. Effects during pregnancy include risks of premature birth and fetal hearing loss [12-14].

Noise can reduce salivary and gastric secretions, slowing digestion and increasing gastric acid [15, 16]. Studies show positive correlations with tinnitus, oxidative imbalance affecting the immune system, and hippocampal damage, which impacts memory and learning [17-19]. Further reported effects include annoyance, mental health issues, altered blood clotting, increased cholesterol, diabetes, and irritability [20-22]. Self-reported symptoms among workers include otalgia, insomnia, and the need to raise one's voice [23]. Noise also modifies social behavior, potentially reducing helping behavior and increasing aggression [24].

This article addresses extra-auditory effects in academic musicians in Uruguay. The interdisciplinary team "SAVEM" (Auditory, Vocal, and Ergonomic Health in Academic Musicians) comprises experts from Health, Social, Artistic, and Technical fields at the Universidad de la República (UdelaR) [25]. Since 2018, SAVEM has worked with the National Choir, the Montevideo Symphonic Band, the National Symphony Orchestra, and other professional groups. The team's goal is to understand the health consequences of occupational conditions (high noise, vocal strain, postural problems) and provide tools for auditory, vocal, and postural hygiene.

2. Methods

2.1. Study Design and Setting

This research was a cross-sectional study of academic musicians in Uruguay, part of SAVEM at Universidad de la República since 2018. The team used health check-ups, interviews, and environmental measures to identify non-auditory effects of noise exposure. Participants attended interviews at each stage for health info and personal insights.

2.2. Participants

The study population included professional and student musicians from several institutions and ensembles [25]: (i) National Choir of Uruguay; (ii) Montevideo Symphonic Band; (iii) Lyric Singing Department of the Faculty of Arts; (iv) National Symphony Orchestra (approx. 95 musicians); (v) Montevideo Philharmonic Orchestra (approx. 95 musicians). Over the course of the project stages, three datasets were compiled, comprising 163, 86, and 59 individual records, respectively. The participants included a broad spectrum of musical roles, such as singers, string players, woodwind players, brass players, percussionists, keyboardists, and conductors.

2.3. Data Collection Procedures

2.3.1 Individual Interviews and Anamnesis

A structured interview guide, developed by the research team, was used. It included variables related to occupational health as well as music-specific factors (such as instrument played, vocal register, frequency of artistic activity, daily practice hours, and use of amplification). Individual interviews were conducted by an occupational physician and a speech therapist. The appointments were made in advance.

2.3.2 Clinical and Audiological Examination

Participants were instructed regarding precautions prior to audiometric testing.

2.3.3 Environmental Measurements

Sound pressure levels (SPL), vibration, carbon monoxide concentration, temperature, and humidity recorded during rehearsals. Observation sessions were also carried out. Acoustic maps of rehearsal rooms were constructed to determine exposure profiles (minimum $L_{AF,eq}$ levels of 85 dBA were recorded). See Figure 1.

2.4. Ethical Considerations

All participants signed informed consent prior to inclusion. Data was anonymized for analysis. Each participant received individual feedback on results, in accordance with medical confidentiality standards.

The protocol was reviewed under the ethical framework of Udelar, in line with institutional and national occupational health regulations.

2.5. Data Analysis

Quantitative analysis included descriptive statistics (distribution by sex, age, and musical role), prevalence of self-reported extra-auditory effects (e.g., tinnitus, irritability, vertigo, musculoskeletal complaints, sleep disturbances, cardiovascular and cognitive symptoms).

Comparative analyses were conducted to explore associations between effects and demographic variables (gender, age, role). Multivariate analysis was performed, including:

- Principal Component Analysis (PCA) to identify effect clusters (e.g., cognitive vs. psychological profiles);
- Hierarchical cluster analysis to examine associations between reported symptoms and population characteristics.

Results were stratified by dataset and presented in figures to illustrate prevalence patterns.

3. Results and Discussion

3.1. Acoustic Maps

First, the acoustic maps generated in the rehearsal rooms of the National Choir and the Montevideo Symphonic Band are shown, providing a quantitative starting point for the sound exposure of academic musicians in their daily work. As it can be seen in Figure 1, the lowest levels recorded among the musicians' groups are $L_{AF,eq}$ = 85 dBA. These high sound pressure levels lead to the expectation of both auditory and non-auditory effects on musicians.

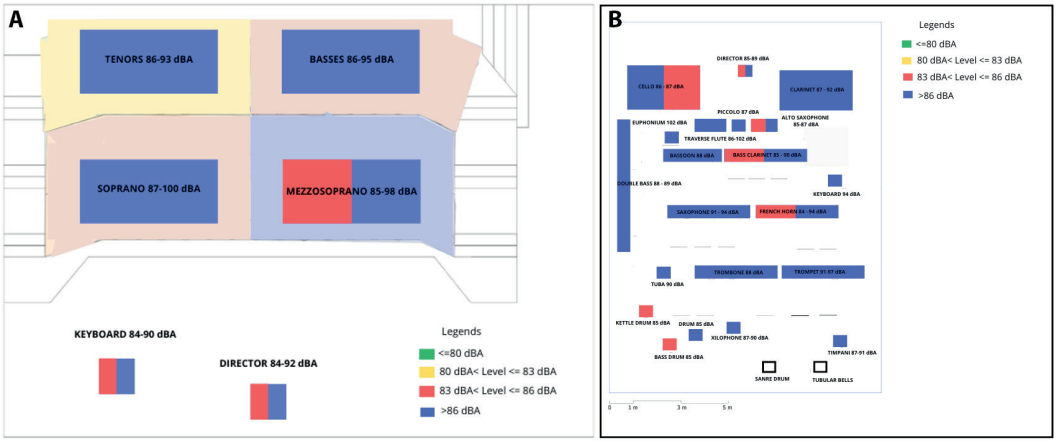

Figure 1. Maps of Sound Pressure Levels: (a) at the National Choir rehearsal; (b) at the Montevideo Symphonic Band rehearsal (redrawn from Tomasina et al. 2025) (26).

It is worth noting that the Montevideo Symphonic Band has incorporated an important preventive measure: a "street" between the brass and woodwind instruments to reduce the latter's sound exposure levels.

3.2. First Data Set Processing: 163 Records

3.2.1 General Information

This data set includes 86 women (53%) and 77 men (47%). 37% of the sample corresponds to individuals under 35 years of age, while the remaining 63% are over 35 years of age. Regarding gender distribution, the sample is fairly balanced.

Regarding the percentage distribution of the participants' roles, singers are 47% of respondents, constituting the largest group, followed by bowed string players (22%) and woodwind players (13%). Brass players represent 10% of the participants. It is worth noting that the roles of keyboard, conductor and plucked strings, and cases without data, represent approximately 5% of the total records.

3.2.2 Reported Effects

This set of data assessed the presence of tinnitus, vertigo, ear pain, and other effects such as hypertension. The effects in the surveyed population showed tinnitus being the most prevalent effect, present in 90 of the respondents (55%), followed by post-musical activity irritability with 64 cases (39%). Only 28 people (17%) reported not experiencing any of the effects assessed. The high prevalence of tinnitus found in our data set is much higher than the 27% reported for classical musicians by Di Stadio et al. (2018) [27].

Regarding the proportion of people experiencing at least one effect, in all roles with 10% or more participants, the presence of some effect exceeds 75%, with woodwind instrument players showing the highest percentage (91%).

3.2.3 Prevalence of Effects by Sex and Age

The prevalence of the reported effects was shown by age group and sex. The results for women revealed a clear trend in the prevalence of tinnitus: its prevalence was lower among those under 35 years of age and increased progressively thereafter, reaching values above 60% in the over-56 age group. In contrast, irritability after musical activity was most prevalent in the 26-35 age group, decreased among those under 25, and declined with increasing age. Regarding vertigo, no clear pattern was identified, and otalgia was not present in participants over 56 years of age. In men, tinnitus was recorded in all age groups, with a prevalence exceeding 50%. Irritability after musical activity exhibited a similar pattern, remaining at levels above 40% in most age groups, except for those over 56 years of age, where it decreased considerably. Vertigo was more prevalent in the extreme age ranges (under 25 years of age and over 56 years of age), while, as in women, ear pain was not present in participants in the oldest age group.

3.2.4 Prevalence of Effects by Role

Regarding the effects by role, of the 40 women with tinnitus, 24 were singers (60%), 9 string instrumentalists (23%), and 6 wind instrumentalists (15%). Of the 50 men with tinnitus, 12 were singers (24%), 12 string instrumentalists (24%), and 18 wind instrumentalists (36%), of whom 13 (26%) were brass instrumentalists. Of the 8 violinists with tinnitus, 6 were bilateral and 2 were left-handed. In addition, 4 viola players had tinnitus: 3 bilateral and 1 right-handed.

Regarding those who experienced irritability after musical activity, of the 33 women, 13 were singers (39%), 10 string instrumentalists (30%), and 10 wind instrumentalists (30%). Of the 31 men who experienced the effect, 3 were singers (10%), 8 (26%) were string musicians (6 of them violinists), and 13 were wind musicians (42%), of whom 9 (29%) were brass musicians.

40 people experienced vertigo, with a higher prevalence in women (27 cases, 68%) than in men. Among the women who experienced vertigo, 19 were singers (70%), and of these, 10 were sopranos (53%). Among the 13 men who experienced vertigo, 5 were wind musicians (38%) and 4 of them were brass musicians (80%). Ear pain occurred primarily in string and brass musicians (35% in each case).

String musicians with ear pain were more men, and brass musicians were women.

Only 5.5% of participants had high blood pressure (9 cases). Of these, 8 (89%) were men, primarily string and brass musicians (43% each). High blood pressure in men was distributed more or less evenly across all age groups from 26 years of age.

3.3. Second Data Set Processing: 84 Records

3.3.1 General Information

In this second data set, 43 women and 41 men participated (51% and 49% respectively). 50% of this population were under 35 years old, while approximately 11% were over 56. The remainder of the sample is made up of participants between 36 and 55 years old.

49% of respondents were singers, constituting the largest group, followed by bowed string (27%), brass (9%), and woodwind (6%) instrumentalists. As in the first data set, roles such as percussion and plucked strings, and records without data, represent less than 10% of the total population.

3.3.2 Reported Effects

18 of the participants (21%) reported needing a conversation to be repeated; the number of cases was equal between men and women. In the case of women, this effect was reported more frequently among singers, while in men it was reported equally among singers and instrumentalists. In both sexes, 4 of the 9 cases (44%) occurred before the age of 35.

On the other hand, 13 of the participants (15%) reported having hearing problems. The effect was more common in younger women and older men. Most cases occurred among singers (69%), and of these, the majority (56%) were sopranos. Furthermore, 12 of the participants (14%) reported having to turn up the television volume. The effect was more evident in men (58%) than in women (42%) and more in singers (75%) than in instrumentalists (25%).

In contrast, 71% of respondents reported no effect.

3.4. Third Data Set Processing: 59 Records

3.4.1 General Information

In this set of data, the sampled population has very few participants under 25 years of age. Two clearly differentiated groups are identified: on the one hand, people over 46 years of age, who represent 54% of the population, with a majority of men; on the other hand, people under 46 years of age, where there is a greater presence of women in the 26-35 age range and men in the 36-45 age range.

Related to the proportion of participants' roles: 29% are singers, followed by woodwind instrumentalists at 27%, bowed string players at 19%, and brass players at 14%. Finally, conductor, percussion, keyboards, and plucked strings make up the remaining percentage.

3.4.2 Prevalence of Effects by Sex, Age and role

The most common effect was muscle contractures, reported by 45 individuals (76%). Of the 24 women who presented contractures, 11 (46%) were singers and 13 (54%) instrumentalists, with a similar prevalence across string and wind instruments (46% in each case). Of the 21 men who reported contractures, 20 (95%) were instrumentalists, primarily wind instruments (65%) and among them, the majority were brass instruments (85%).

Noise susceptibility was the second most reported effect (28 cases, 47%). The effect was reported in an equal number of men and women, but it appeared earlier (in age) in women than in men. Among women, 57% (8) were singers, and of these, 87% were sopranos. Among instrumentalists (43%), the majority (83%) were wind instrumentalists. Among men, the vast majority of cases occurred in instrumentalists (86%). Among them, 7 (58%) were wind instrumentalists, and most of them (71%) played brass instruments.

24 people (41%) reported having trouble sleeping, especially falling asleep. 13 of them (54%) were women and 11 (46%) were men. The women were more or less equally distributed between singers and instrumentalists. All the singers were sopranos, and the instrumentalists were predominantly woodwind.

The majority of the men were instrumentalists (91%), and among them, half were brass players.

Arthralgia was reported by 23 people (39%). The 13 women with arthralgia (56%) were equally distributed between singers and instrumentalists. All men (44%) were instrumentalists, and 50% of them played string instruments.

19 people (32%) reported tiredness or fatigue. Most of the cases (74%) were women. Of these, 8 (57%) were singers, 7 of whom were sopranos (87%). Among instrumentalists (43%), the majority (83%) were wind musicians. The effect was present in all age groups. In men (26% of the total cases), no singer reported the effect. 60% of cases occurred in wind instrumentalists, primarily brass instruments.

Headaches occurred in 17 cases (29%), primarily in women (76%). Among them, the majority (69%) were instrumentalists and 31% were singers (sopranos). In men, the effect occurred in instrumentalists, but not in singers.

15 cases reported a decline in performance (25%). Among men (53%), the effect was more common in instrumentalists, and 88% among them were brass musicians. Among women, the effect occurred in singers (43%) and instrumentalists (57%); among this second group, most of them also played wind instruments (75%).

In 13 cases (22%), digestive disorders, including heartburn and reflux, occurred. Most cases were men (54%), with majority (71%) of wind instrumentalists; among them, 60% played brass instruments. As in women, most cases occurred after the age of 45. In the case of women (46%), the effect was only reported in singers, with 67% sopranos.

The same number of cases (13, or 22%) reported difficulty concentrating. Most of the cases were women (62%), and a greater number of them were instrumentalists (75%), equally distributed between string and wind instruments. Among men, the majority (60%) were brass instrumentalists.

A slightly lower number of cases (12, or 20%) reported memory impairment. In this case, there were more men (58%) than women (42%). Most of the men were instrumentalists (86%), but most of the women (60%) were singers.

The results show that contractures reach values close to 80% in most groups, with the exception of those over 56 years of age. Effects such as difficulty sleeping, noise sensitivity, and headaches tend to decrease with age, while arthralgia, decreased performance, and digestive disorders tend to increase (see Figure 2).

3.5. Effect Profiles by Role and Gender

Given that in some roles the number of recordings was considerable, it was possible to determine profiles of the effects prevalent in these roles. In these profiles, the effects with a prevalence of 40% of more are mentioned in next paragraphs and synthetized in Figure 3.

Cases (a) and (b) show the effect profiles of female and male woodwind instrumentalists. As it can be seen, they are not the same. In the case of female, 89% reported irritability after noise exposure, 86% reported muscle contracture, 71% reported tiredness/fatigue and the same percentage reported noise susceptibility, difficulty sleeping; 56% reported tinnitus, 43% reported headaches, difficulty concentrating, arthralgia and performance decrease. In the case of male, 78% reported muscle contracture, 67% reported tinnitus, 50% irritability after noise exposure, 44% reported noise susceptibility and the same percentage reported performance decrease.

Case (c) presents the profile of extra-auditory effects in sopranos. Approximately 78% of them reported contractures and the same percentage reported fatigue/tiredness. 56% was the prevalence of noise sensitivity difficulty sleeping, arthralgia and digestive disorders. and. Subsequently, 44% reported headache, 40% reported tinnitus and the same percentage reported vertigo.

Case (d) shows the profile of extra-auditory effects in male brass musicians. Approximately 71% reported tinnitus, 67% of them reported contracture and 50% irritability after noise exposure and 44% reported difficulty sleeping.

Cases (e) and (f) show the effect profiles of bowed strings instrumentalists of both genders. As it can be seen, they are not the same. In the case of female, 100% reported muscle contracture and headache, 60% reported irritability after noise exposure, 50% reported difficulty concentration and 47% reported tinnitus. In the case of male, 100% reported muscle

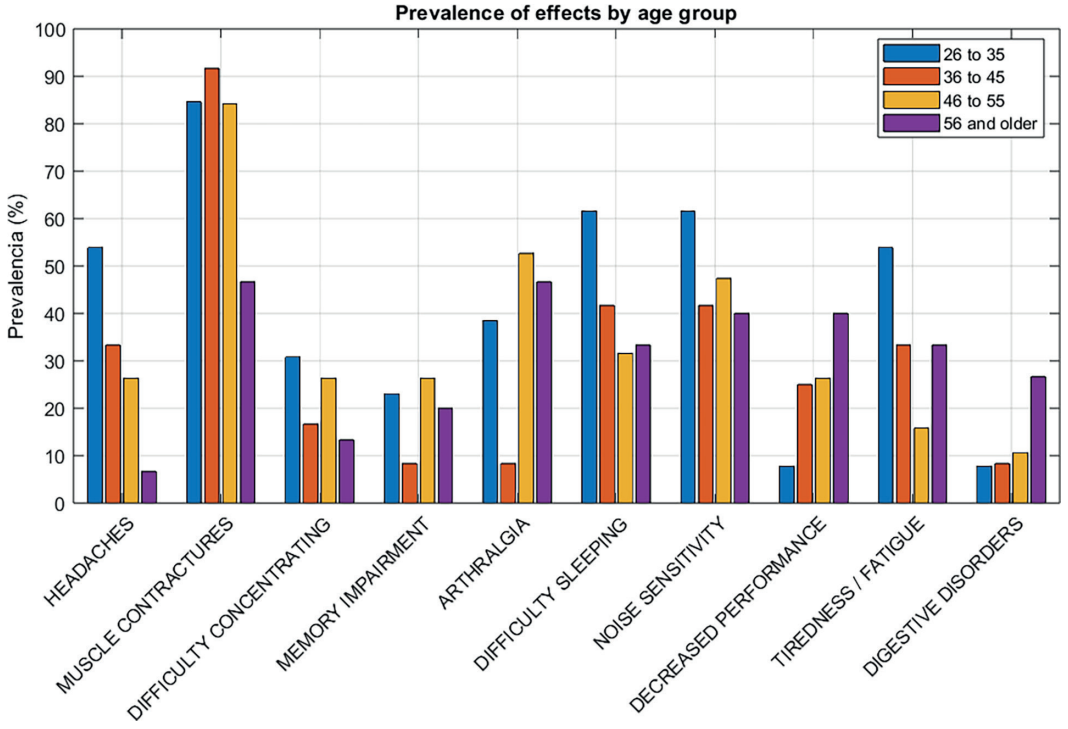

Figure 2. Prevalence of effects by age ranges (third data set).

contracture, 80% reported arthralgia, 60% reported sleeping difficulty, 55% reported tinnitus, 40% reported irritability after noise exposure and the same percentage reported noise susceptibility.

Case (f) refers to violinists of both genders. 100% reported muscle contracture, 50% reported arthralgia and the same percentage reported headache, 40% reported tinnitus and the same percentage reported irritability after noise exposure and difficulty concentration.

3.6. Multivariate Analysis

The objective of this analysis is to identify associations between the extra-auditory effects and some population characteristics, such as gender, age, and role. To this end, it was decided to omit certain effects with a low prevalence in the established age ranges, as well as those with a high prevalence in most age ranges, as they do not provide differentiating information. Consequently, the following effects were selected: vertigo, noise susceptibility, irritability after noise exposure (post-musical irritability), difficulty sleeping, difficulty concentrating, decreased performance, and memory impairment.

The first step in the analysis was to perform a principal components analysis (PCA). This method generates new variables, called principal components, which are linear combinations of the original variables. From PCA, several components were obtained. The top five of them were selected for use in further analyses, as they explain most of the variability in the data set.

Figure 4 shows the first two principal components along with the contributions of the selected variables. It can be seen that the variables "difficulty concentrating", "decreased performance", and "memory disturbance" are oriented toward the lower part of the graph, forming a kind of intellectual or cognitive profile. On the other hand, the variables "difficulty sleeping," "post-musical irritability," and "noise susceptibility" are oriented toward the positive part of the first principal component (PC1), forming a psychological profile associated with noise burden. In this case, gender also plays a significant role. Age and role appear to have little influence on the differentiation of the reported effects.

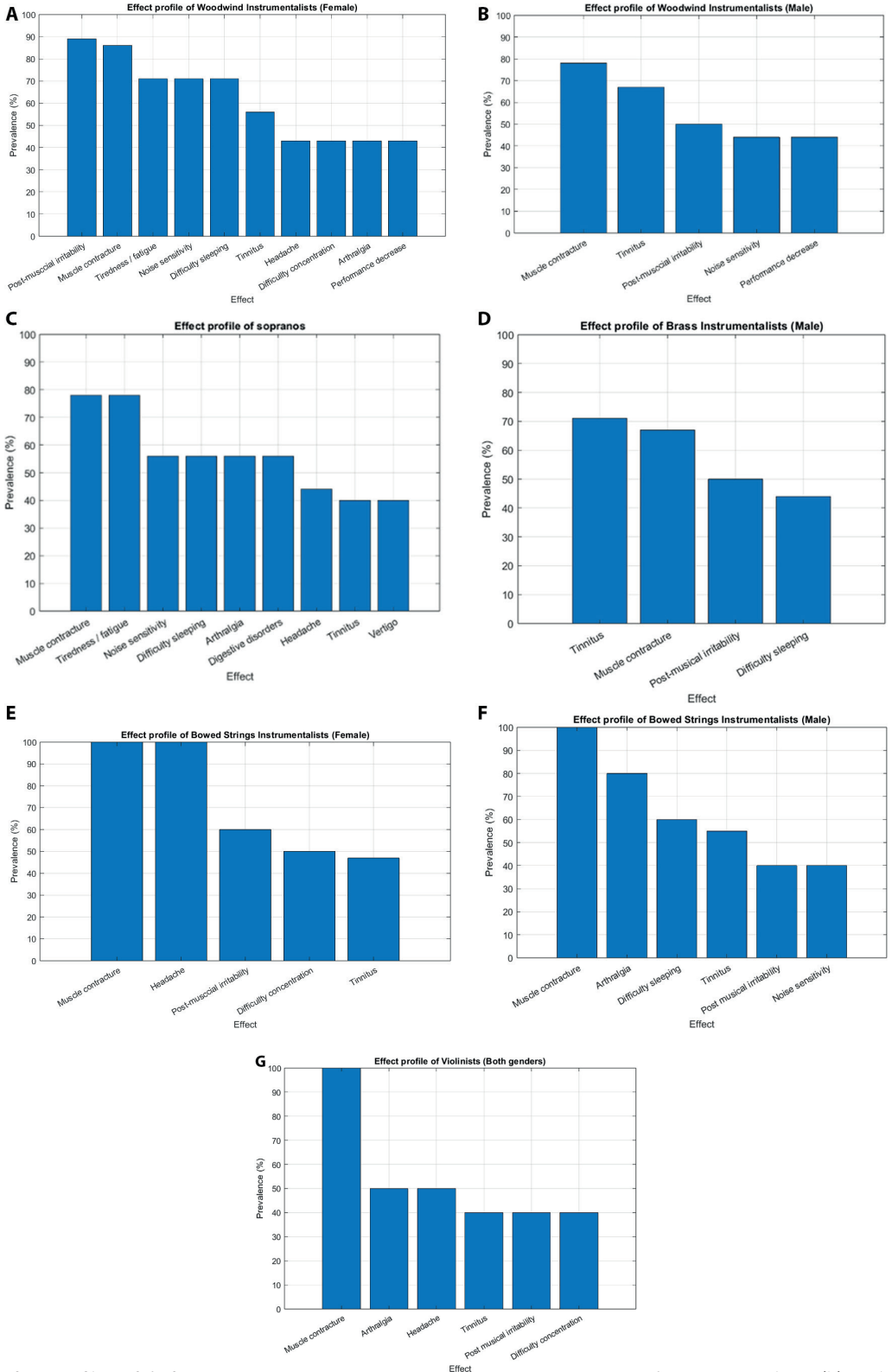

Figure 3. Effect profiles of different instrumentalists, by gender: (a) female woodwind instrumentalists; (b) male woodwind instrumentalists; (c) soprano singers; (d) brass male instrumentalists; (e) female bowed strings instrumentalists; (f) male bowed strings instrumentalists; (g) violinists (both genders).

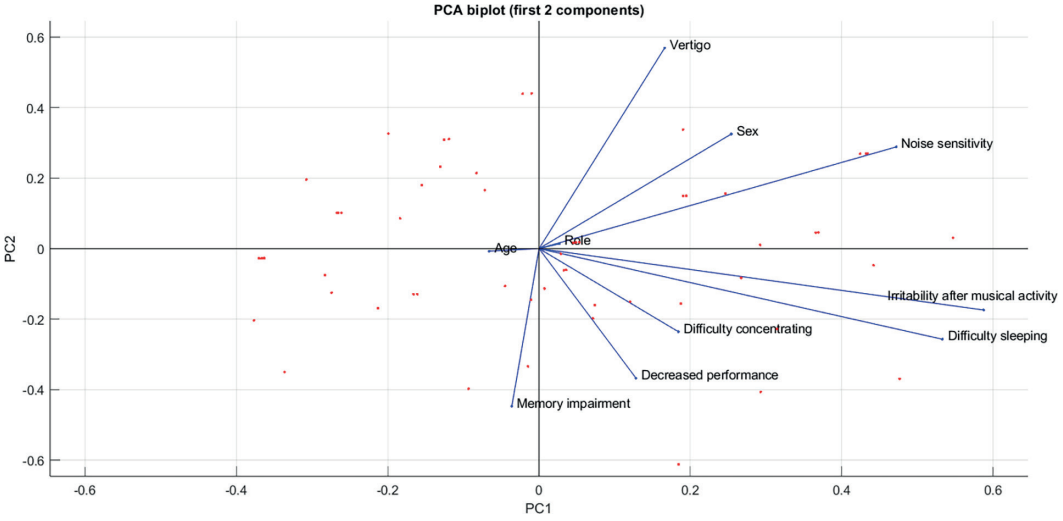

Figure 4. Biplot PCA (first two components).

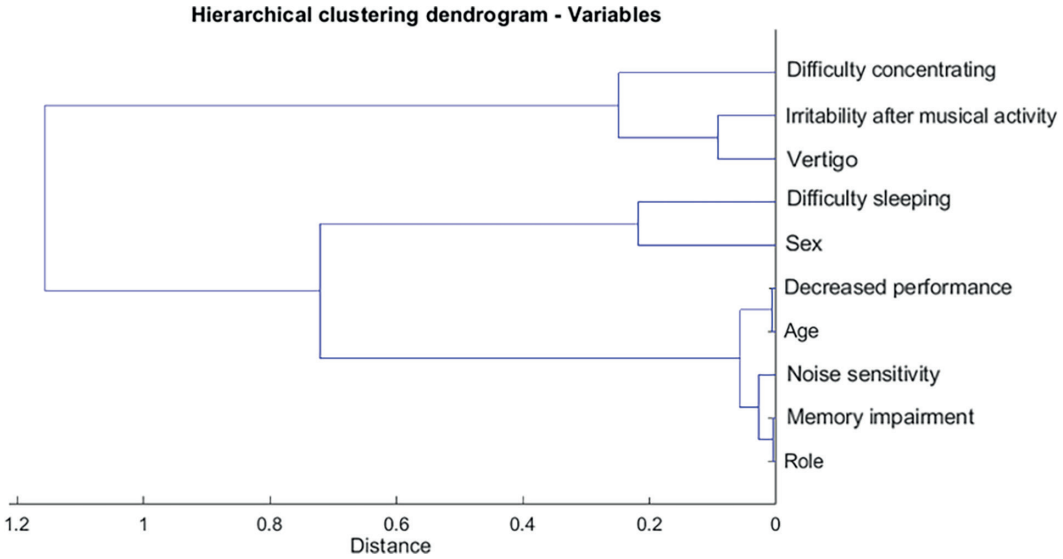

Figure 5. Hierarchical dendrogram (cluster analysis).

To further explore the relationships between the effects and the demographic characteristics of the population, a hierarchical dendrogram was constructed (see Figure 5). The dendrogram shows that certain effects are clearly associated with each other; for example, "memory impairment" and "noise susceptibility" are related to roles (i.e., the instrument played). Similarly, a relationship is identified between age and decreased performance. On the other hand, some variables, such as difficulty concentrating or post-musical activity irritability, appear further separated in the dendrogram, indicating that they are less correlated with the other variables analyzed, as if they were independent of gender, age, or the instrument played.

4. Conclusions

Three data sets about extra-auditory effects of high sound pressure levels exposure of musicians

were analyzed. The distribution by sex and age range was balanced in most cases, and a significant group of those over 46 years of age was observed.

Singers constituted the largest group in all sets (45-50%), followed by bowed string, woodwind, and brass players. Other roles, such as percussion, keyboard, conducting, or plucked strings, generally represented less than 10% of the samples.

Among the reported effects, the most prevalent were contractures, fatigue/tiredness, difficulty sleeping, and noise sensitivity, followed by arthralgia and headaches. Tinnitus was also common, especially among singers and some instrumentalists. Tinnitus increased with age in women and remained high in men; irritability was higher in young women and decreased in men over 56 years of age.

The most frequent effects in women were muscle contracture (86%), followed by tiredness/fatigue (50%), noise susceptibility (50%) and tinnitus (47%).

The most frequent effects in men were muscle contracture (68%), followed by tinnitus (65%), noise susceptibility (45%) and post-musical activity irritability (40%).

Role and gender profiles showed that:

- Soprano singers have a higher prevalence of muscle contracture, tiredness/fatigue, noise sensitivity, difficulty sleeping, arthralgia, digestive disorders, headache, tinnitus and vertigo. They are the second most suffering category of musicians, according to their effect profile.
- Male brass musicians experience tinnitus, muscle contracture, post-musical activity irritability and difficulty sleeping.
- All woodwind instrumentalists experiment muscle contracture, tinnitus, post-musical activity irritability, noise susceptibility and decreased performance.
- Female woodwind musicians also experiment tiredness/fatigue, headache, difficulty sleeping, difficulty concentration and arthralgia. They are the most suffering category of musicians, according to their effect profile.
- Bowed strings instrumentalists of both genders reported muscle contracture, post-musical activity irritability and tinnitus. Men also experienced arthralgia, difficulty sleeping and noise sensitivity, while woman presented headache and difficulty concentration.
- Violinists, regardless of gender, have a high prevalence of muscle contracture, arthralgia, headache, tinnitus, post-musical activity irritability and difficulty concentration.

Funding: This research was funded as Semillero Interdisciplinario in the 2018 call and as Núcleo Interdisciplinario in the 2020 call of the Espacio Interdisciplinario of the Universidad de la República.

Institutional Review Board Statement: Although the research involved anamnesis data, all subsequent analysis were based on anonymized data. For this, the participants were informed and signed an informed consent form that detailed the purpose of the study and the use of the data for research purposes.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to thank all those who participated in the experimental work, both the singers and musicians and those who collaborated on the interviews and logistics.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: AEG, FT, AP, LCR and BT contributed to the conception and design of the research. FT, AP and BT performed the interviews. AEG, JOU and LCR performed the data processing and management. AEG, FT, JOU, LCR, BT and AP contributed to the analysis of the results and to the writing of the manuscript.

References

1. González A. E. Contaminación Sonora y Derechos Humanos. Serie Investigaciones: Derechos humanos en las políticas públicas. N° 2. Available at https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/38860/1/Gon12.pdf
2. World Health Organization. Occupational health: Noise [Internet]. Ginebra: OMS; 2021 [Last Accessed 2025 sep 24]. Available at: https://www.who.int/news-room/fact-sheets/detail/noise
3. International Labour Organization. Safety and health at the heart of the future of work: Building on 100 years of experience [Internet]. Geneva: ILO; 2019 [Last Accessed 2025 Sep 24]. Available at: https://www.ilo.org

4. Orbe M, Betancourt Ó. Epidemiología de los trastornos auditivos. Servicios integrales de salud y seguridad en el trabajo, Herramientas y técnicas, manejo electrónico de la información. 1er edición, Quito Ecuador, 2021. P 199.
5. Basner M, Babisch W, Davis A, et al. Auditory and non-auditory effects of noise on health. *The Lancet*. 2014;383(9925):1325–32. Doi: 10.1016/S0140-6736(13)61613-X
6. Hubspotusercontent-na1.net. [Last Accessed 2025 September 26]. Available at: https://2841087.fs1.hubspotusercontent-na1.net/hubfs/2841087/DKV-%20IC%20Ebook%20/IC%20-%20Ebook%20-%20Ruido%20y%20salud/DKV-Ebook-Observatorio-ruido-2023.pdf
7. CCOHS. Noise. Non-auditory effects (2022). Fact sheet last revised 2022-04-27.
8. Jones, K (2010). Environmental noise and health: a review. ERCD Report 0907, Environmental Research and Consultancy Department, UK Civil Aviation Authority. Published by TSO (The Stationery Office) on behalf of the UK Civil Aviation Authority.
9. Díaz Torrealba, Mariana (2012). Efectos neurológicos y psicológicos producidos por ruido en la población que ocupa el centro de Barquisimeto, PhD Thesis, Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Industriales.
10. Fernandes de Souza TC, Santos Périssé AR, Moura M. Noise exposure and hypertension: investigation of a silent relationship. *BMC Public Health* (2015);15:328 Doi: 10.1186/s12889-015-1671-z
11. Randolph SA. Promoting Hearing Health. *Workplace Health Saf.* 2018;66(8):412. Doi: 10.1177/2165079918782720. Epub 2018 Jun 17. PMID: 29909742.
12. Stansfeld S, Clark C. Health effects of noise exposure in children. *Curr Environ Health Rep.* 2015;2(2):171-8. Doi: 10.1007/s40572-015-0044-1
13. World Health Organization, Regional Office for Europe. Environmental noise guidelines for the European region [Internet]. Copenhagen: WHO; 2018 [Last Accessed 2025 September 24]. Available at: https://www.euro.who.int
14. Tomasina F, Stolovas N. Manual técnico básico de Salud Ocupacional. 2022. Available at: https://proyectoecho.fmed.edu.uy/wp-content/uploads/2022/12/Manual-tecnico-basico-de-Salud-Ocupacional_Version-digital-1.pdf
15. Anticaglia JR, Cohen A. Extra-Auditory. Effects of Noise as a Health Hazard, *Am Ind Hyg Assoc J.* 1970;31(3):277-281. Doi: 10.1080/00028899708506243
16. Lee Y, Lee S, Lee W. Occupational and environmental noise exposure and extra-auditory effects on humans: A systematic literature review. *GeoHealth,* 2023;7:e2023GH000805. Doi: https://doi.org/10.1029/2023GH000805
17. Tekriwal RP, Dharitri M. Extra Auditory Effect of Noise – A study on textile workers of Surat city. *Nat J Physiol Pharm Pharmacol*. 2012;2(1):45-51.
18. Molina SJ, Miceli M, Guelman LR. Noise exposure and oxidative balance in auditory and extra-auditory structures in adult and developing animals. Pharmacological approaches aimed to minimize its effects. *Pharmacol Research*. 2016;109:86-91.
19. Molina SJ, Guelman LR. Noise-induced hippocampal damage: potential mechanisms. *Neural Regen Res.* 2022;17(3):563-564.
20. Mehrotra A, Shukla SP, Shukla AK, et al. A Comprehensive Review of Auditory and Non-Auditory Effects of Noise on Human Health. *Noise Health*. 2024;26:59-69.
21. Themann, Christa L, Masterson, Elizabeth A. Occupational noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden. *J Acoust Soc Am.* 2019;146:3879. Doi: 10.1121/1.5134465
22. De Sio S, Perri R, Durel Tchaptchet F, et al. Extra-auditory effects of noise exposure in school workers and preventive measures: a systematic review. *Eur Rev Med Pharmacol Sci.* 2023;27:1467-1473.
23. Taoussi AA, Yassine AA, Malloum MSM, Assi C, Fotclossou T, Ali YA. Effects of noise exposure among industrial workers in power plants of the National Electricity Company in N'Djamena, Chad. *Egypt J Otolaryngol.* 2022;38:63. Doi: https://doi.org/10.1186/s43163-022-00253-7
24. Stansfeld, Stephen A, Matheson, Mark P. Noise pollution: non-auditory effects on health. *Br Med Bull*. 2003;68:243-257. Doi: 10.1093/bmb/ldg033
25. González AE, Tomasina F, Gianoli Kovar P, et al. (2023). Núcleo Interdisciplinario SAVEM, Informe Final.
26. Tomasina F, González AE, Pisani A, et al. Audiometric Database of Academic Musicians in Uruguay. *Med Lav*. 2025;116(3):16424. Doi: 10.23749/mdl.v116i3.16424
27. Di Stadio A, Dipietro L, Ricci , et al. Hearing Loss, Tinnitus, Hyperacusis, and Diplacusis in Professional Musicians: A Systematic Review. *Int J Environ Res Public Health.* 2018;15:2120. Doi:10.3390/ijerph15102120

Mattioli 1885

La Medicina del Lavoro

Original article

Med. Lav. 2026; 117 (2): 18283
DOI: 10.23749/mdl.2026.18283

Diagnostic Utility of Serum Krebs von den Lungen-6 (KL-6) and Surfactant Protein-D (SP-D) Levels in Hypersensitivity Pneumonitis

Asuman Aslan Kara[1*], Adem Koyuncu[1], Gülden Sari[2] Tuğçe Şahin Özdemirel[3], Berna Akinci Özyürek[3], Cebrail Şimşek[1]

[1]Department of Occupational Diseases, Ankara Atatürk Sanatorium Training And Research Hospital, Ankara, Turkey
[2]Department of Occupational Diseases, Yenimahalle Training And Research Hospital, Ankara, Turkey
3Department of Pulmonology, Yenimahalle Training And Research Hospital, Ankara, Turkey

Keywords: Krebs von den Lungen-6 (KL-6); Surfactant Protein (SP-D); Hypersensitivity Pneumonitis

Abstract

Background: *This study aimed to investigate the diagnostic and prognostic values of serum levels of Krebs von den Lungen-6 (KL-6) and surfactant protein-D (SP-D) in patients with hypersensitivity pneumonitis (HP).* **Methods:** *Serum samples were collected from patients diagnosed with HP and from a healthy control group. KL-6 and SP-D levels were measured using ELISA. HP cases were further compared across fibrotic and non-fibrotic subgroups and between those receiving treatment and those not. The relationships between respiratory function tests (DLCO, FEV_1, FVC) and biomarker levels were examined.* **Results:** *Both KL-6 (median 5.95 ng/ml vs 5.4 ng/ml, p<0.001) and SP-D (median 14.87 ng/ml vs 14.72 ng/ml, p<0.05) levels were significantly higher in HP patients than in the control group. In non-fibrotic HP patients, KL-6 levels were higher than in the fibrotic group (median 6.07 ng/ml vs 5.62 ng/ml, p<0.05), while no significant difference was observed for SP-D (p=0.71). KL-6 levels were significantly higher in untreated cases than in treated cases (median 6.30 ng/ml vs 5.65 ng/ml, p<0.01), whereas the difference in SP-D was not significant (p=0.26).* **Conclusions:** *KL-6 emerges as a sensitive biomarker for the diagnosis of HP, assessment of disease activity, and monitoring of treatment response. SP-D, although reflecting inflammatory processes, seems to have limitations in evaluating fibrotic progression.*

1. Introduction

Hypersensitivity pneumonitis (HP) is an inflammatory and/or fibrotic interstitial lung disease (ILD) that affects the lung parenchyma and small airways. It develops in susceptible individuals following exposure to inhaled antigens [1, 2]. HP is the third most common form of ILDs, accounting for up to 47.3% of all cases [3–8]. More than 50 different occupations and exposures associated with HP have been identified [1, 2]. The main causative agents in the disease's etiology are microorganisms, animal proteins, and organic and inorganic dust particles. HP is classified as fibrotic or non-fibrotic based on imaging and histopathological findings. The 7-year survival rate for fibrotic HP is 40.8%, indicating a poorer prognosis compared to many types of cancer [9–11].

Clinically, cough, dyspnea, fever, and fatigue are prominent symptoms [12–14]. While some patients present with acute symptoms, others may progress to the fibrotic stage [6]. History, exposure

Received 23.11.2025 – Accepted 17.03.2026
*Corresponging Author: Asuman Aslan Kara; E-mail: drasuaslan@gmail.com

determination, and high-resolution computed tomography (HRCT) are critical for diagnosis [15]. Antigen removal is the basis of treatment for HP. Survival is poorer in cases where exposure cannot be identified and continues [13, 16]. Serum-specific IgG, bronchoalveolar lavage (BAL) lymphocytosis, and lung biopsy are supportive methods in the diagnosis of HP [4]. Due to the high mortality and rapid progression of fibrotic HP, there is a need for biomarkers to identify high-risk patients at an early stage [5, 8].

In this context, KL-6, which reflects type II alveolar epithelial damage, and SP-D, which is involved in the immune response, are prominent biomarkers. KL-6 is a high molecular weight glycoprotein found on the membrane of type II alveolar epithelial cells. It is released during cell damage and has been associated with activity and prognosis in interstitial lung diseases [5, 8, 17, 18]. SP-D is a collectin secreted from type II pneumocytes and involved in the innate immune response [8, 19]. Serum KL-6 and SP-D levels are used in the diagnosis and follow-up of various ILDs, including HP, particularly in Japan. They have been associated with epithelial damage and mortality [8, 13, 20, 21]. However, studies specific to HP are limited [1, 5]. Although KL-6 and SP-D biomarkers have been evaluated in hypersensitivity pneumonia in previous studies, many of these studies have not thoroughly analyzed biomarker levels in conjunction with clinical findings, radiological features, and treatment strategies. This study aimed to evaluate serum KL-6 and SP-D levels in patients diagnosed with HP and to determine the factors affecting these levels.

2. Methods

This cross-sectional study included 42 patients with hypersensitivity pneumonitis (HP) and 42 age- and gender-matched healthy volunteers. The Ethics Committee of Ankara Sanatorium Training and Research Hospital approved the study (2024-BCEK/247), and all participants provided informed consent. The HP diagnosis was established through a multidisciplinary evaluation by an experienced pulmonologist, radiologist, pathologist, and occupational disease specialist at the Interstitial Lung Diseases Council.

The control group comprised individuals in whom HP was ruled out. Their occupations and environmental exposures (e.g., dust, chemicals, animal allergens) were investigated, and they had no connective tissue, autoimmune, or malignancy-related interstitial lung diseases. Patient data included demographics, exposure history, occupation, smoking status, comorbidities, HRCT findings, pulmonary function tests (FEV_1, FVC, FEV_1/FVC, DLCO), 6-minute walk test (6MWT), and lab results. Patients were classified as fibrotic or non-fibrotic HP based on radiological and histopathological features.

Serum KL-6 and SP-D levels were measured in ng/mL using commercial ELISA kits (Shanghai Coon Koon Biotech) according to the manufacturer's instructions. Washing was performed with HUMAN COMBI WASH, and optical density was read with the NEXT LEVEL ALISEI device. All analyses followed the manufacturer's protocols; raw results were multiplied by the dilution factor to obtain final concentrations. Internal quality control results were within the reference ranges. Samples were collected via centrifugation at appropriate conditions, and measurements at 450 nm confirmed ELISA accuracy and reliability. No additional accuracy tests were conducted. The study evaluated the relationships between KL-6 and SP-D levels in fibrotic and non-fibrotic HP groups and controls.

Using GPower 3.1, the minimum sample size required to achieve 90% power with an effect size of 0.6 was calculated as 39 patients with hypersensitivity pneumonia and 39 healthy controls. To account for potential data loss, 42 patients and 42 healthy volunteers were included. Data were analyzed using IBM SPSS Statistics 22.0 software. Categorical variables were presented as counts (%), and numerical variables as mean ± standard deviation. Normality of distribution was assessed using the Kolmogorov-Smirnov test. The chi-square test was used for categorical variables, the Kruskal-Wallis and Mann-Whitney U tests for intergroup comparisons, and the Spearman correlation coefficient for correlation analyses. ROC analysis was performed for the diagnostic accuracy of biomarkers. A p-value < 0.05 was considered statistically significant.

3. Results

The study included 42 healthy individuals in the control group and 42 individuals with HP in the study group. Smoking history was detected more frequently in the HP group (64.3% vs. 50%), but the difference was not statistically significant (p=0.27). Symptoms were detected in 83.3% of the HP group, while all individuals in the control group were asymptomatic. The most common symptoms were cough (61.9%), shortness of breath (69.0%), sputum (33.3%), fever (19.0%), and weight loss (16.7%). The presence of comorbidities was similar in both groups (control 42.9%, HP 45.2%). When exposure types were examined in the HP group, animal protein exposure was the most common (54.7%). This was followed by unknown exposure (26.1%), metal exposure (7.1%), fungal-mold exposure (7.1%), and plant protein exposure (4.7%).

Laboratory tests revealed significantly elevated WBC and CRP levels in the HP group (p<0.001 for both). There was no difference in LDH values (p=0.40). In respiratory function tests, FEV1, FVC, and FEV1/FVC were similar, while DLCO/SB was significantly lower in the HP group (82.1±24.0 vs 98.1±9.9; p<0.01).

The 6-Minute Walk Test (6-MWT) distance was similar in both groups (p=0.21), but the HP group had lower start and end oxygen saturations (O_2 sat) (p<0.01 and p<0.001, respectively), a lower start heart rate, and a significantly higher end heart rate (p<0.01 for both).

High-resolution computed tomography (HRCT) findings revealed that ground-glass opacity was the most frequently observed finding in HP patients (41.7%), followed by centrilobular nodules, linear reticular pattern, mosaic attenuation, honeycomb pattern, traction bronchiectasis, and fibrosis. 66.7% of HP patients were recorded as non-fibrotic and 33.3% as fibrotic type. In terms of treatment, 42.8% of HP patients received no treatment, while 57.1% received treatment. Among the treated patients, 83.3% (n=20) received steroid treatment, and 16.6% (n=4) received a combination steroid and antifibrotic treatment (Table 1).

When HP patients were classified as fibrotic (n=14) and non-fibrotic (n=28), the age distribution was similar between groups (p=0.36). A significant difference was observed by gender: the proportion of women in the non-fibrotic group was 39.3%, whereas in the fibrotic group it was only 7.1% (p < 0.05). Smoking status and presence of comorbidities did not differ between groups (p=0.30 and p=0.74, respectively). There was no significant difference in the presence of symptoms between the two groups (non-fibrotic 85.7%, fibrotic 78.6%; p=0.66). The two groups were similar in terms of exposure types; animal protein exposure was the most common (non-fibrotic 57.1%, fibrotic 50.0%; p=0.74).

Laboratory findings showed that WBC values were slightly higher in the fibrotic group (Table 2), but the difference was borderline significant (p=0.05); there was no significant difference between the groups in CRP and LDH values.

In respiratory function tests, FEV1, FVC, and FEV1/FVC were similar; however, DLCO/SB was significantly lower in the fibrotic group (69.2±19.1 vs. 88.5±23.9; p<0.05). There were no differences between the groups in 6-Minute Walk Test distance and oxygen saturation values. In HRCT findings, centrilobular nodules were more common in the non-fibrotic group (92.9% vs 57.1%; p<0.05), while honeycombing was more common in the fibrotic group (85.7% vs 7.1%; p<0.001). Traction bronchiectasis (85.7% vs 17.9%; p<0.001) and fibrosis (100% vs 0; p<0.001) were predominant in the fibrotic group. Ground-glass opacity was similar in both groups.

When examining the diagnostic methods for cases diagnosed with hypersensitivity pneumonitis (HP), it was determined that 61.9% of patients (n=26) were diagnosed based on clinical and radiological findings, 30.9% (n=13) were confirmed histopathologically, and 7.1% (n=3) were diagnosed based on bronchoalveolar lavage (BAL) findings. Regarding treatment status, approximately half of both groups had received treatment (non-fibrotic 57.1%, fibrotic 57.1%). (Table 2)

In terms of KL-6 and SP-D biomarkers, both KL-6 (median 5.95 ng/ml vs 5.4 ng/ml, p<0.001) and SP-D (median 14.87 ng/ml vs 14.72 ng/ml, p<0.05) levels were significantly higher in HP patients than in the control group. In non-fibrotic HP patients, KL-6 levels were higher than in the fibrotic

Table 1. Sociodemographic Characteristics of Patients

			Without Hypersensitivity Pneumonitis (Control Group) n=42	With Hypersensitivity Pneumonitis (Study Group) n=42	
			n(%)	n(%)	*p* value
Age (decade)	30-39		6(14.3)	5(11.9)	0.97[a]
	40-49		10(23.8)	11(26.2)	
	50-59		12(28.6)	12(28.6)	
	≥60		14(33.3)	14(33.3)	
Gender	Female		11(26.2)	12(28.6)	>0.99[a]
	Male		31(73.8)	30(71.4)	
Smoking	No		21(50.0)	15(35.7)	0.27[a]
	Yes		21(50.0)	27(64.3)	
		Former	0	17(40.5)	
		Current	21(100)	10(23.8)	
Symptoms	No		42(100)	7(16.7)	**<0.001**
	Yes[1]		0	35(83.3)	
			Cough	26(61.9)	
			Sputum	14(33.3)	
			Dyspnea	29(69.0)	
			Fever	8(19.0)	
			Weight loss	7(16.7)	
Comorbidities	No		24(57.1)	23(54.8)	>0.99[a]
	Yes[2]		18(42.9)	19(45.2)	
		DM	12(28.6)	13(31.0)	
		HT	11(26.2)	11(26.2)	
		other	0	6(14.3)	
Exposure Type			Animal protein	23(54.7)	
			Unknown	11(26.1)	
			Metal	3(7.1)	
			Fungus-mold	3(7.1)	
			Plant protein	2(4.7)	
Laboratory[3]	WBC		6(5-8)	7(5-13)	**<0.001**[b]
	CRP		2(1-5)	4(1-81)	**<0.001**[b]
	LDH		174.5(139-210)	174.5(26-436)	0.40[b]
Pulmonary function test[4]	FEV1		83.0±6.7	85.2±21.8	0.44[b]
	FVC		84.5±8.2	88.4±22.7	0.21[b]
	FEV1/FVC		79.4±3.3	78.6±5.2	0.37[b]
	DLCO/SB		98.1±9.9	82.1±24.0	**0.001**[b]

		Without Hypersensitivity Pneumonitis (Control Group) n=42	With Hypersensitivity Pneumonitis (Study Group) n=42	
		n(%)	n(%)	*p* value
6 MWT	Not completed	0	2(4.8)	0.49[a]
	Completed	42(100)	40(95.2)	
	Distance (meters)	480(400-600)	480(400-720)	0.21[b]
	Start sat O_2	97(92-98)	95(82-99)	**0.002**[b]
	End sat O_2	96(90-102)	93(65-99)	**<0.001**[b]
	Start heart rate	96(83-99)	91.5(70-119)	**0.005**[b]
	End heart rate	97(83-108)	105(71-142)	**0.001**[b]
High-Resolution Computed Tomography findings (HRCT) [5]		Ground glass	35(41.7)	
		Centrilobular nodule	34(40.5)	
		Linear reticulation	29(34.5)	
		Mosaic attenuation	22(26.2)	
		Traction bronchiectasis	17(20.2)	
		Honeycombing	14(16.7)	
		Fibrosis	14(16.7)	
HP type	Non-fibrotic		28(66.6)	
	Fibrotic		14(33.3)	
Treatment status	No		18(42.8)	
	Yes		24(57.1)	

[1]A patient has more than one symptom; [2]A patient has more than one comorbidity; [3]median(min-max); [4] mean±standard deviation; [5]A patient has more than one radiology finding; [a] χ^2; [b] Mann-Whitney U test; DM: Diabetes Mellitus; HT: Hypertension; 6 MWT: 6-minute walk test

Table 2. Comparison of hypersensitivity pneumonitis subgroups

		Non-fibrotic n=28	Fibrotic n=14	
		n(%)	n(%)	*p* value
Age (decade)	30-39	4(14.3)	1(7.1)	0.36[a]
	40-49	6(21.4)	5(35.7)	
	50-59	10(35.7)	2(14.3)	
	≥60	8(28.6)	6(42.9)	
Gender	Female	11(39.3)	1(7.1)	**0.03**[a]
	Male	17(60.7)	13(92.9)	
Smoking	No	12(42.9)	3(21.4)	0.30[a]
	Yes	16(57.1)	11(78.6)	

Table 2 (*Continued*)

		Non-fibrotic n=28	Fibrotic n=14	
		n(%)	n(%)	*p* value
Symptoms	No	4(14.3)	3(21.4)	0.66[a]
	Yes [1]	24(85.7)	11(78.6)	
Comorbidities	No (n=23)	16(57.1)	7(50.0)	0.74[a]
	Yes [2] (n=19)	12(42.9)	7(50.0)	
Exposure types	Animal protein (n=23)	16(57.1)	7(50.0)	0.74[a]
	Unknown (n=11)	8(28.6)	3(21.4)	
	Metal (n=3)	1(3.6)	2(14.3)	
	Fungus-mold (n=3)	2(7.1)	1(7.1)	
	Plant protein (n=2)	1(3.6)	1(7.1)	
Laboratory[3]	WBC	7(5-13)	8(5-11)	0.05[b]
	CRP	3.5(1-28)	4.5(3-81)	0.07[b]
	LDH	171.5(99-436)	187.5(26-260)	0.28[b]
Pulmonary function test[4]	FEV1	87.5±22.1	80.5±21.1	0.29[b]
	FVC	92.3±23.5	80.5±19.6	0.13[b]
	FEV1/FVC	77.8±5.1	80.2±5.2	0.06[b]
	DLCO/SB	88.5±23.9	69.2±19.1	**0.01**[b]
6 MWT [3]	Not completed (n=2)	2(7.1)	0	0.54[a]
	Completed (n=40)	26(92.9)	14(100)	
	Distance (meters)	480(400-720)	480(400-640)	0.80[b]
	Start sat O_2	95(89-99)	94.5(82-98)	0.33[b]
	End sat O_2	94(80-98)	91(65-99)	0.16[b]
	Start heart rate	92(70-111)	90(76-119)	0.33[b]
	End heart rate	105(71-140)	105(85-142)	0.95[b]
High-Resolution Computed Tomography findings (HRCT) [5]	Ground-glass (n=35)	24(85.7)	11(78.6)	0.66[a]
	Centrilobular nodule (n=34)	26(92.9)	8(57.1)	**0.01**[a]
	Linear reticulation (n=29)	17(60.7)	12(85.7)	0.15[a]
	Mosaic attenuation (n=22)	17(60.7)	12(85.7)	0.23[a]
	Honeycomb (n=14)	2(7.1)	12(85.7)	**<0.001**[a]
	Traction bronchiectasis (n=17)	5(17.9)	12(85.7)	**<0.001**[a]
	Fibrosis (n=14)	0	14(100)	**<0.001**[a]
Treatment status	No (n=18)	12(42.9)	6(42.9)	>0.05[a]
	Yes (n=24)	16(57.1)	8(57.1)	

[1] A patient has more than one symptom.
[2] A patient has more than one comorbidity.
[3] median(min-max) [4] mean±standart deviation.
[5] A patient has more than one radiology finding.
[a] Chi sqare test.
[b] Mann-Whitney U test.

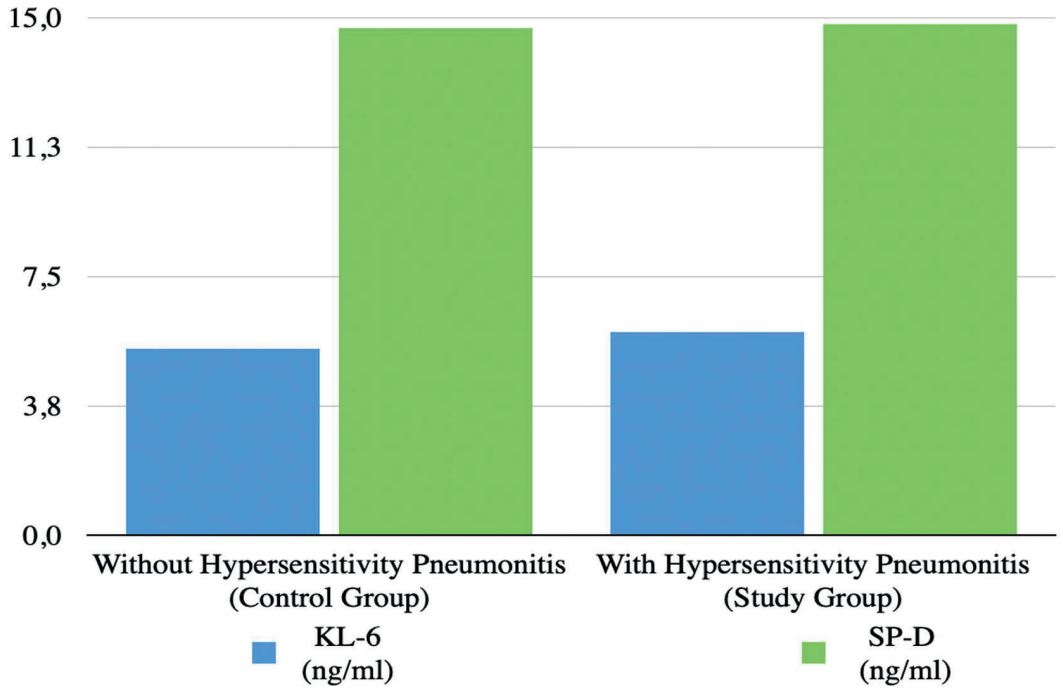

Figure 1. Median values of serum KL-6 and SP-D levels in the hypersensitivity pneumonitis (study) group and the non-hypersensitivity pneumonitis (control) group.

group (median 6.07 ng/ml vs 5.62 ng/ml, p<0.05), and SP-D values did not differ significantly between the groups (p=0.71).

Serum KL-6 and SP-D levels in the hypersensitivity pneumonitis group and the control group are shown in Figure 1.

KL-6 values were significantly higher in untreated patients than in treated patients (median 6.30 ng/ml vs 5.65 ng/ml, p<0.01), whereas there was no significant difference in SP-D levels (p=0.26). (Table 3)

The median KL-6 value was 5.70 (4.05–11.60) in the comorbidities group and 5.50 (3.05–37.30) in the non-comorbidities group. The median SP-D value was 14.95 (5.0–17.80) in the group with comorbidities and 14.65 (7.10–94.95) in the group without comorbidities. KL-6 and SP-D values did not differ significantly between the groups based on the presence of comorbidities (p = 0.17 and p = 0.33, respectively). When examined by exposure type, KL-6 and SP-D values also did not differ (p = 0.41 and p = 0.06, respectively). (Table 4)

When KL-6 and SP-D values were evaluated based on HRCT findings in HP patients, it was found that KL-6 values were significantly higher in individuals without linear reticulation compared to those with linear reticulation (median 6.85 ng/ml vs 5.65 ng/ml; p<0.01), while no difference was observed in SP-D values (p=0.24). Additionally, KL-6 values were significantly higher in patients without fibrosis than in those with fibrosis (6.07 ng/ml vs 5.62 ng/ml; p<0.05), whereas SP-D levels were not statistically different (p=0.07).

Honeycomb and traction bronchiectasis were found to be associated with significantly higher SP-D levels compared to those without (p<0.05 for both), while no significant association was observed for KL-6 levels. There was no significant relationship between the markers and ground-glass opacity, centrilobular nodules, or mosaic attenuation (p>0.05 for all). (Table 5)

Spearman correlation analysis revealed a weak positive correlation between KL-6 and SP-D values

Table 3. Comparison of KL-6 and SP-D values between patients with and without hypersensitivity pneumonitis.

	KL-6* (ng/ml)	p value**	SP-D* (ng/ml)	p value**
Without Hypersensitivity Pneumonitis (Control Group) n=42	5.4(3.05-6.20)	**<0.001**	14.72(5.0-16.80)	**0.03**
With Hypersensitivity Pneumonitis (Study Group) n=42	5.95(4.15-37.30)		14.87(11.8-94.5)	
Non-fibrotic (n=28)	6.07(4.15-37.3)	**0.04**	15.07(14.0-94.95)	0.71
Fibrotic (n=14)	5.62(4.75-6.35)		14.45(11.8-17.05)	
No treatment (n=18)	6.30(5.35-37.3)	**0.005**	14.95(13.35-94.95)	0.26
Treated (n=24)	5.65(4.15-7.30)		14.82(11.8-19.75)	

**median(min-max); **Mann Whitney-U test; KL-6: Krebs von den Lungen-6; SP-D : Surfactant Protein-D.*

Table 4. Association of KL-6 and SP-D markers with comorbidities and exposure.

		KL-6[1] (ng/ml)	p value	SP-D[1] (ng/ml)	p value
Comorbidities	No (n=47)	5.50(3.05-37.30)	0.17 [2]	14.65(7.10-94.95)	0.33 [2]
	Yes (n=37)	5.70(4.05-11.60)		14.95(5.0-17.80)	
Exposure type	Animal protein (n=23)	6.05(4.15-37.30)	0.41[3]	15.10(14-94.95)	0.06 [3]
	Unknown (n=11)	5.90(5.20-8.55)		14.75(14.15-19.75)	
	Metal* (n=3)	6.0(5.65-6.25)		14.15(13.36-14.25)	
	Fungus-mold (n=3)	5.5(4.75-5.50)		14.5(11.80-15.45)	
	Plant protein (n=2)	8.2(4.8-11.6)		16.95(16.70-17.20)	

*[1] median(min-max); [2] Mann Whitney-U test; [3] Kruskal Wallis test; * Occupations of the cases with metal exposure were metal processing (n=1) and hard metal cutting (n=2). Identified exposure agents were metalworking fluids (boron oil, n=1) and cobalt (n=2).*

Table 5. Comparison of HRCT findings with KL-6 and SP-D values.

HRCT Findings[1] (n=42)		KL-6[2] (ng/ml)	p value[3]	SP-D[2] (ng/ml)	p value[3]
Ground-glass	No (n=7)	5.65(5.35-6.35)	0.32	14.6(14.15-16.80)	0.91
	Yes (n=35)	6.05(4.15-37.3)		14.9(11.8-94.95)	
Centrilobular nodule	No (n=8)	5.82(4.80-8.55)	0.91	14.42(13.35-16.70)	0.14
	Yes (n=34)	5.95(4.15-37.3)		14.92(11.8-94.95)	
Linear reticulation	No (n=13)	6.85(4.15-37.3)	**0.009**	15.05(14.05-94.95)	0.24
	Yes (n=29)	5.65(4.75-7.95)		14.85(11.8-17.05)	
Mosaic attenuation	No (n=20)	5.65(4.15-37.30)	0.14	14.75(14.15-94.95)	0.27
	Yes (n=22)	6.20(4.75-9.45)		14.92(11.8-17.8)	
Honeycombing	No (n=28)	6.07(4.15-37.3)	0.07	15.07(14.0-94.95)	**0.02**
	Yes (n=14)	5.62(4.75-7.30)		14.45(11.8-16.7)	
Traction bronchiectasis	No (n=25)	6.15(4.15-37.3)	0.06	15.10(14.0-94.95)	**0.02**
	Yes (n=17)	5.65(4.75-7.30)		14.6(11.8-16.8)	
Fibrosis	No (n=28)	6.07(4.15-37.30)	**0.04**	15.07(14.0-94.95)	0.07
	Yes (n=14)	5.62(4.75-6.35)		14.45(11.8-17.05)	

[1] A patient may have more than one HRCT finding; [2] median (min-max);[3] Mann-Whitney-U test.

in HP patients, but this correlationwas not statistically significant (rho=0.212; p=0.178).

Correlation analysis was also conducted between KL-6 and SP-D levels and respiratory function parameters, showing no statistically significant relationships with DLCO-SB, FEV_1, or FVC.

3.1. ROC Analysis

The diagnostic performance of the KL-6 and SP-D biomarkers for hypersensitivity pneumonitis was evaluated using ROC analyses. KL-6 was shown to provide moderat diagnostic value for the diagnosis of HP (AUC=0.76; SE=0.05; 95% CI 0.66–0.86; p<0.001). The cut-off value was >5.42 ng/ml, with a sensitivity of 78% and specificity of 42%.

SP-D had more limited but significant diagnostic performance(AUC=0.63; SE=0.06; 95% CI 0.51–0.75; p=0.03). The cut-off value was >14.57 ng/ml, with a sensitivity of 64% and a specificity of 52% (Figures 2 and 3).

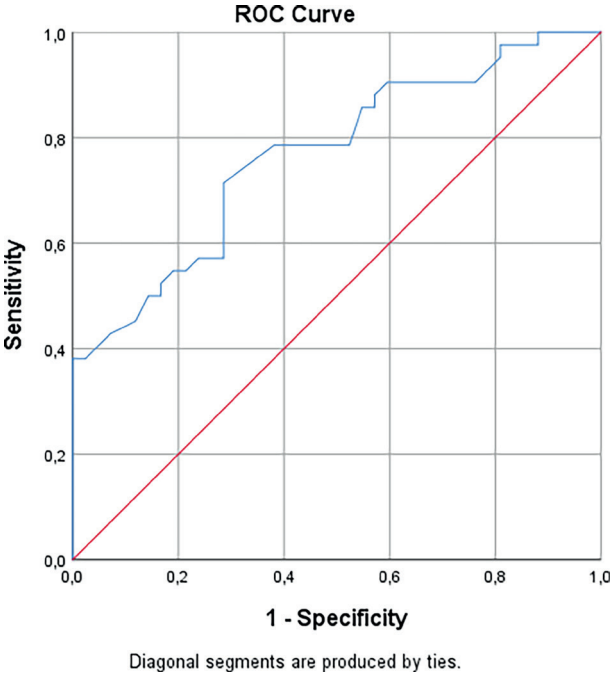

Figure 2. KL-6 marker ROC curve (AUC=0.76; SE=0.05; 95% CI 0.66–0.86; p<0.001).

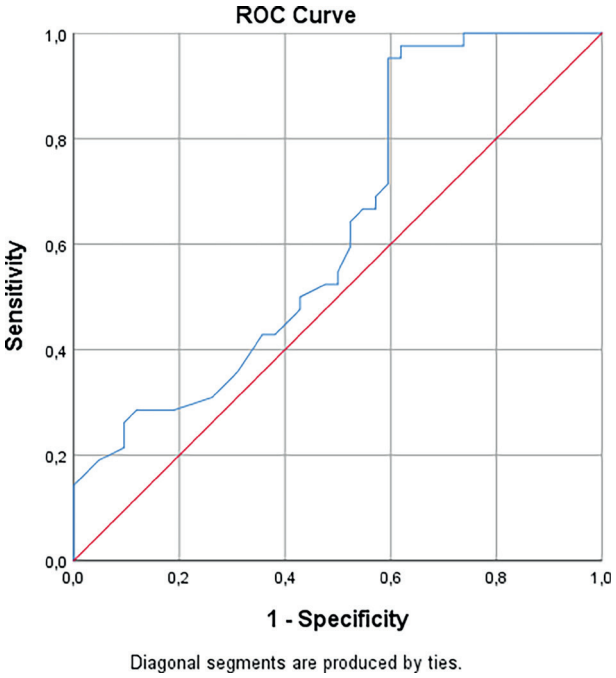

Figure 3. SP-D marker ROC curve (AUC=0.63; SE=0.06; 95% CI 0.51–0.75; p=0.03).

4. Discussion

Hypersensitivity pneumonitis (HP) is an interstitial lung disease that requires early diagnosis and regular follow-up. Its fibrotic form, in particular, can clinically mimic idiopathic pulmonary fibrosis (IPF) and carries a high mortality risk. In this context, KL-6 and SP-D are gaining importance as potential biomarkers for assessing disease activity and course. In our study, significant differences in serum KL-6 and SP-D levels were observed between patients with HP and healthy controls. Important findings related to subtypes and treatment status were also noted. The highest levels of both biomarkers were observed in the non-fibrotic HP group, whereas lower levels were observed in patients receiving systemic steroid treatment, suggesting that KL-6 and SP-D may be associated not only with disease presence but also with inflammatory activity and treatment response. Therefore, it was emphasized that these biomarkers have the potential to serve as helpful indicators of disease course. ROC analyses indicate that KL-6

may show supportive performance in distinguishing between the HP and control groups, whereas SP-D has more limited discriminative value. However, due to the study design, variability in biomarker levels, and sample size, these results should be interpreted with caution.

Although some significant correlations between KL-6 and SP-D and respiratory function parameters have been reported in previous studies [22], these relationships were limited to specific parameters or were weak [23, 24]. Other studies found no significant correlations [25, 26]. In our study, no statistically significant relationships were observed between serum KL-6 and SP-D levels and DLCO, FEV_1, or FVC, suggesting that measurements of biomarkers at a single time point may not reveal a clear linear relationship, which might be more evident when monitoring changes over time or assessing specific patient subgroups. Additionally, factors like population heterogeneity, small sample size, the timing of measurements, treatment status, and disease stage may also contribute to the inability to statistically demonstrate these relationships.

The literature reports that carbon monoxide diffusion capacity (DLCO) is significantly lower in patients with fibrotic HP than in those with non-fibrotic HP [27]. Consistent with these findings, our study also found DLCO values to be lower in the fibrotic HP group than in the non-fibrotic HP group. Furthermore, KL-6 levels were significantly higher in HP patients without a linear reticular pattern or fibrosis, and SP-D levels were significantly higher in those without honeycomb and traction bronchiectasis. Conversely, although increases in both biomarkers were observed in cases with ground-glass opacities, centrilobular nodules, and mosaic attenuation, these differences were not statistically significant. In cases where linear reticulation and fibrotic patterns dominate, low KL-6 levels indicate loss of type II alveolar epithelial cells and decreased cellular activity; whereas in advanced fibrotic stages accompanied by honeycomb and traction bronchiectasis, decreased SP-D levels can be explained by destruction of alveolar architecture and a reduction in surfactant-producing cells. The literature reports that BAL KL-6 levels in HP patients are associated with HRCT findings and lymphocytosis [3], and that elevated serum KL-6 levels correlate positively with reticular patterns and honeycombing [28]. These data suggest that KL-6 may serve as a potential biomarker for monitoring disease activity and prognosis in both HP and IPF. In fibrotic HP, structural HRCT findings have been linked to poor prognosis [13], while inflammatory findings are associated with a better prognosis [29]. Our current study also indicates that KL-6 and SP-D levels may be associated with distinct pathological patterns observed on HRCT and that these biomarkers may vary by disease phenotype. Overall, these findings suggest that both biomarkers are associated with pathophysiological processes reflecting different radiological stages of HP.

In our study, serum KL-6 levels were significantly higher in non-fibrotic HP patients than in fibrotic HP patients, whereas no significant difference was observed in SP-D levels. The interaction between inflammation and fibrosis plays a crucial role in the pathogenesis of HP. The inflammatory process, which is predominant in the early stages, leads to alveolar damage and activation of type II pneumocytes, resulting in increased release of biomarkers such as KL-6 and SP-D. Prolonged or recurrent inflammation triggers fibroblast activation and matrix accumulation through interactions with epithelial-structural cells, leading to fibrotic transformation. Therefore, prolonged disease duration or inadequate inflammation control may cause initially elevated KL-6 and SP-D levels to exhibit different patterns over time as fibrosis develops [30].

In the literature, KL-6 has been linked to the prognosis of fibrotic hypersensitivity pneumonitis (HP), and its serial measurement is valuable for early identification of patients at risk of progression [5]. KL-6 is a cell membrane protein secreted by type II pneumocytes and is elevated in interstitial lung diseases (ILDs) with significant inflammation. It can serve as a prognostic and diagnostic marker to distinguish between non-fibrotic and fibrotic HP. It has been reported to correlate with disease activity in both acute and chronic HP, and in non-fibrotic HP, early alveolitis and elevated serum levels may indicate mild alveolar damage [31, 32]. Similarly, high serum KL-6 levels have been observed in cases diagnosed in both domestic and occupational

contexts. It has been emphasized that alveolitis may be present in the early stages of non-fibrotic HP[3]. KL-6 differs from SP-D structurally; it's a cell membrane protein, and increased serum levels indicate membrane damage and enzyme activation. If SP-D increases without a rise in KL-6, it may suggest mild alveolar damage [33]. Our study supports KL-6 reflecting disease activity and being higher in non-fibrotic HP than in fibrotic cases, whereas SP-D is less sensitive for this distinction.

In our study, we found that serum KL-6 levels were significantly lower in HP patients receiving corticosteroid treatment compared to those not receiving treatment. Although a decrease in SP-D levels was also observed, this difference was not statistically significant. Increased KL-6 production in plasma is considered a sensitive indicator of alveolitis, and serum KL-6 levels in interstitial lung diseases (ILD) have prognostic and diagnostic value, reflecting both disease activity and severity [34]. KL-6, produced during the regeneration of type II alveolar epithelial cells, supports epithelial repair in response to alveolar damage, and high plasma levels are considered an indicator of increased airway epithelial permeability [35]. Therefore, targeting KL-6 therapeutically may help achieve optimal effect before fibrosis develops when corticosteroids are administered in the early stages of the disease. Consequently, the detection of elevated plasma KL-6 levels in early-stage HP patients with normal spirometry is clinically significant [36].

In the literature, higher KL-6 levels were observed in the non-fibrotic HP group compared to fibrotic cases, suggesting a link with pneumocyte renewal in response to alveolar damage [37]. Previous studies have shown that HP exacerbations are associated with increases in KL-6 and SP-D levels, while corticosteroid treatment and avoidance of the antigen decrease these biomarker levels [20]. Our study demonstrates that treatment lowers KL-6 and SP-D levels, with a more significant and stronger effect on KL-6. Fibrotic HP has been reported to have a 7-year survival rate estimated at 40.8% in recent studies and to exhibit worse outcomes than many cancers [10, 11]. Increased mortality has highlighted the need for accurate and timely diagnosis and exposure assessment to improve disease outcomes [8]. Research indicates that serial measurements of KL-6 can predict survival, which is associated with the poor prognosis and high mortality of fibrotic HP [38]. Additionally, acute exacerbations are considered negative prognostic indicators, and simply avoiding the antigen is not always enough [13, 39]. These findings support the clinical value of KL-6 and SP-D in HP diagnosis, disease phenotype differentiation, and monitoring disease activity. They also emphasize the importance of biomarker panels for early diagnosis and prognostic prediction.

One of the strengths of our study is that serum KL-6 and SP-D levels in patients with HP were evaluated in detail across fibrotic and non-fibrotic subgroups. Furthermore, the inclusion of a healthy control group contributed to a clearer understanding of the diagnostic value of these biomarkers. The study's scope was broadened and its clinical relevance increased by examining KL-6 and SP-D in Hypersensitivity Pneumonitis not only biomarker levels but also relationships with pulmonary function tests, HRCT findings, exposure types, and treatment response. The fact that our findings are largely consistent with the current literature also supports the value of KL-6 and SP-D as both diagnostic and prognostic biomarkers in clinical practice for HP.

Our study has limitations, including a small sample size reducing statistical power, especially in subgroup analyses. Unlike previous studies reporting KL-6 and SP-D in U/mL, ours used ng/mL, limiting direct comparison, though findings align with literature [40, 41]. Biomarkers measured at a single time point prevented assessment of disease activity changes. Variations in treatment, exposure, and measurement units hindered interpretation. Non-standardized corticosteroid doses and treatment durations may explain the lack of significant SP-D changes. Being cross-sectional without longitudinal follow-up, conclusions on KL-6 are based on literature. As a single-center study, generalizability may be limited.

5. Conclusion

In conclusion, serum levels of KL-6 and SP-D were higher in HP patients than healthy controls. KL-6 was elevated in non-fibrotic HP cases compared to fibrotic ones, indicating it reflects disease activity. SP-D, though sensitive for monitoring,

showed less difference than KL-6. The decline in serum KL-6 with steroid treatment suggests it helps monitor alveolar damage and treatment response. Our study supports the diagnostic and prognostic value of KL-6 and SP-D in HP, potentially linked to diagnosis and disease phenotype. Lower levels in treated patients indicate these markers vary with treatment and may reflect decreased inflammation.

Funding: This research was funded by *Tuberculosis Prevention Foundation, grant number 3.*

Institutional Review Board Statement: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration (as revised in 2013) and its later amendments or comparable ethical standards. The study was approved by the Ethics Committee of Ankara Sanatorium Training and Research Hospital (2024-BCEK/247).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patients to publish this paper.

Declaration of Interest: The authors declare no conflict of interest.

Author Contribution Statement: A.AK. contributed to the conception, design, implementation, execution, interpretation, analysis, and writing of the manuscript, A.K. contributed to the interpretation, analysis, and writing of the manuscript, G.S. contributed to the interpretation, analysis, and writing of the manuscript, T.ŞÖ. contributed to the implementation and execution of the manuscript, B.AÖ. contributed to the implementation and execution of the manuscript, C.Ş. contributed to the analysis of the results.

References

1. Raghu G, Remy-Jardin M, Ryerson CJ, et al. Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*. 2020;202(3):e36-e69. doi:10.1164/rccm.202005-2032ST
2. Morell F, Villar A, Ojanguren I, Muñoz X, Cruz MJ. Hypersensitivity Pneumonitis: Challenges in Diagnosis and Management, Avoiding Surgical Lung Biopsy. *Semin Respir Crit Care Med*. 2016;37(3):395-405. doi:10.1055/s-0036-1580692
3. Lanzarone N, Gentili F, Alonzi V, et al. Bronchoalveolar lavage and serum KL-6 concentrations in chronic hypersensitivity pneumonitis: correlations with radiological and immunological features. *Intern Emerg Med*. 2020;15(7):1247-1254. doi:10.1007/s11739-020-02281-8
4. Vasakova M, Morell F, Walsh S, Leslie K, Raghu G. Hypersensitivity Pneumonitis: Perspectives in Diagnosis and Management. *Am J Respir Crit Care Med*. 2017;196(6):680-689. doi:10.1164/rccm.201611-2201PP
5. Hanzawa S, Tateishi T, Ishizuka M, et al. Changes in serum KL-6 levels during short-term strict antigen avoidance are associated with the prognosis of patients with fibrotic hypersensitivity pneumonitis caused by avian antigens. *Respir Investig*. 2020;58(6):457-464. doi:10.1016/j.resinv.2020.05.007
6. Cottin V, Hirani NA, Hotchkin DL, et al. Presentation, diagnosis and clinical course of the spectrum of progressive-fibrosing interstitial lung diseases. *Eur Respir Rev*. 2018;27(150):180076. doi:10.1183/16000617.0076-2018
7. Hilberg O, Hoffmann-Vold AM, Smith V, et al. Epidemiology of interstitial lung diseases and their progressive-fibrosing behaviour in six European countries. *ERJ Open Res*. 2022;8(1):00597-2021. doi:10.1183/23120541.00597-2021
8. Pereira JO, Fernandes V, Alfaro TM, Freitas S, Cordeiro CR. Diagnosis of Fibrotic Hypersensitivity Pneumonitis: Is There a Role for Biomarkers?. *Life (Basel)*. 2023;13(2):565. doi:10.3390/life13020565
9. He J, Zhang J, Ren X. Krebs von den lungen-6 as a clinical marker for hypersensitivity pneumonitis: A meta-analysis and bioinformatics analysis. *Front Immunol*. 2022;13:1041098. doi:10.3389/fimmu.2022.1041098
10. Nasser M, Larrieu S, Boussel L, et al. Estimates of epidemiology, mortality and disease burden associated with progressive fibrosing interstitial lung disease in France (the PROGRESS study). *Respir Res*. 2021;22(1):162. doi:10.1186/s12931-021-01749-1
11. Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971-2011: a population-based study. *Lancet*. 2015;385(9974):1206-1218. doi:10.1016/S0140-6736(14)61396-9
12. Walters GI, Mokhlis JM, Moore VC, et al. Characteristics of hypersensitivity pneumonitis diagnosed by interstitial and occupational lung disease multidisciplinary team consensus. *Respir Med*. 2019;155:19-25. doi:10.1016/j.rmed.2019.06.026
13. Barnes H, Troy L, Lee CT, Sperling A, Strek M, Glaspole I. Hypersensitivity pneumonitis: Current concepts in pathogenesis, diagnosis, and treatment. *Allergy*. 2022;77(2):442-453. doi:10.1111/all.15017

14. Pereira CAC, Soares MR, Boaventura R, et al. Squawks in interstitial lung disease prevalence and causes in a cohort of one thousand patients. *Medicine (Baltimore)*. 2019; 98(29):e16419. doi:10.1097/MD.0000000000016419
15. Churg A. Hypersensitivity pneumonitis: new concepts and classifications. *Mod Pathol*. 2022;35(Suppl 1): 15-27. doi:10.1038/s41379-021-00866-y
16. Fernández Pérez ER, Travis WD, Lynch DA, et al. Diagnosis and Evaluation of Hypersensitivity Pneumonitis: CHEST Guideline and Expert Panel Report. *Chest*. 2021;160(2):e97-e156. doi:10.1016/j.chest.2021.03.066
17. d'Alessandro M, Bergantini L, Cameli P, et al. Serum KL-6 levels in pulmonary Langerhans' cell histiocytosis. *Eur J Clin Invest*. 2020;50(8):e13242. doi:10.1111 /eci.13242
18. Ohnishi H, Miyamoto S, Kawase S, Kubota T, Yokoyama A. Seasonal variation of serum KL-6 concentrations is greater in patients with hypersensitivity pneumonitis. *BMC Pulm Med*. 2014;14:129. doi:10.1186 /1471-2466-14-129
19. Dyhr JJ, Linderoth IR, Hansen DN, et al. Confined placental mosaicism: placental size and function evaluated on magnetic resonance imaging. *Ultrasound Obstet Gynecol*. 2023;62(1):130-136. doi:10.1002/uog.26174
20. Okamoto T, Fujii M, Furusawa H, Tsuchiya K, Miyazaki Y, Inase N. The usefulness of KL-6 and SP-D for the diagnosis and management of chronic hypersensitivity pneumonitis. *Respir Med*. 2015;109(12): 1576-1581. doi:10.1016/j.rmed.2015.10.005
21. Nukui Y, Yamana T, Masuo M, et al. Serum CXCL9 and CCL17 as biomarkers of declining pulmonary function in chronic bird-related hypersensitivity pneumonitis. *PLoS One*. 2019;14(8):e0220462. doi:10.1371 /journal.pone.0220462
22. Rai M, Parthasarathi A, Beeraka NM, et al. Circulatory Serum Krebs von Den Lungen-6 and Surfactant Protein-D Concentrations Predict Interstitial Lung Disease Progression and Mortality. *Cells*. 2023;12(9):1281. doi:10.3390/cells12091281
23. Chung C, Kim J, Cho HS, Kim HC. Baseline serum Krebs von den Lungen-6 as a biomarker for the disease progression in idiopathic pulmonary fibrosis. *Sci Rep*. 2022;12(1):8564. doi:10.1038/s41598-022-12399-8
24. Letellier A, Rolland-Debord C, Luque-Paz D, et al. Prognostic value of serum Krebs von den Lungen-6 (KL-6) levels in COVID-19 pneumonia. *Respir Med Res*. 2023;84:101054. doi:10.1016/j.resmer.2023.101054
25. Cambier M, Henket M, Frix AN, et al. Increased KL-6 levels in moderate to severe COVID-19 infection. *PLoS One*. 2022;17(11):e0273107. doi:10.1371/journal.pone .0273107
26. Bonella F, Volpe A, Caramaschi P, et al. Surfactant protein D and KL-6 serum levels in systemic sclerosis: correlation with lung and systemic involvement. *Sarcoidosis Vasc Diffuse Lung Dis*. 2011;28(1):27-33.
27. Yazawa S, Suzuki Y, Tanaka Y, et al. 3D-CT-derived lung volumes and mortality risk in patients with fibrotic hypersensitivity pneumonitis. *Allergol Int*. 2025; 74(1):78-85. doi:10.1016/j.alit.2024.07.002
28. Zhu C, Zhao YB, Kong LF, Li ZH, Kang J. *Zhonghua Jie He He Hu Xi Za Zhi*. 2016;39(2):93-97. doi:10.3760 /cma.j.issn.1001-0939.2016.02.004
29. Chung JH, Montner SM, Adegunsoye A, et al. CT findings associated with survival in chronic hypersensitivity pneumonitis. *Eur Radiol*. 2017;27(12):5127-5135. doi:10.1007/s00330-017-4936-3
30. Zhao X, Chen J, Sun H, Zhang Y, Zou D. New insights into fibrosis from the ECM degradation perspective: the macrophage-MMP-ECM interaction. *Cell Biosci*. 2022;12(1):117. doi:10.1186/s13578-022-00856-w
31. d'Alessandro M, Bergantini L, Cameli P, et al. Krebs von den Lungen-6 as a biomarker for disease severity assessment in interstitial lung disease: a comprehensive review. *Biomark Med*. 2020;14(8):665-674. doi:10.2217 /bmm-2019-0545
32. Sánchez-Díez S, Munoz X, Ojanguren I, et al. YKL-40 and KL-6 Levels in Serum and Sputum of Patients Diagnosed With Hypersensitivity Pneumonitis. *J Allergy Clin Immunol Pract*. 2022;10(9):2414-2423. doi:10.1016/j.jaip.2022.06.031
33. Waseda Y, Yasui M, Kurokawa K, et al. Surfactant protein D: a useful marker for differentiation of drug-induced pneumonia and bacterial pneumonia. *Pneumonia (Nathan)*. 2021;13(1):11. doi:10.1186 /s41479-021-00087-6
34. Jiang Y, Luo Q, Han Q, et al. Sequential changes of serum KL-6 predict the progression of interstitial lung disease. *J Thorac Dis*. 2018;10(8):4705-4714. doi:10.21037 /jtd.2018.07.76
35. Miyazaki Y, Tsutsui T, Inase N. Treatment and monitoring of hypersensitivity pneumonitis. *Expert Rev Clin Immunol*. 2016;12(9):953-962. doi:10.1080/17446 66X.2016.1182426
36. Ji Y, Bourke SJ, Spears M, et al. Krebs von den Lungen-6 (KL-6) is a pathophysiological biomarker of early-stage acute hypersensitivity pneumonitis among pigeon fanciers. *Clin Exp Allergy*. 2020;50(12):1391-1399. doi:10.1111/cea.13744
37. Mostafa AI, Salem AE, Ahmed HAM, Bayoumi AI, Halim RMA, Samie RMA. Role of Krebs von den Lungen-6 (KL-6) in Assessing Hypersensitivity Pneumonitis. *Tuberc Respir Dis (Seoul)*. 2021;84(3):200-208. doi:10.4046/trd.2020.0122
38. Wakamatsu K, Nagata N, Kumazoe H, et al. Prognostic value of serial serum KL-6 measurements in patients with idiopathic pulmonary fibrosis. *Respir Investig*. 2017;55(1):16-23. doi:10.1016/j.resinv.2016 .09.003
39. Kang J, Kim YJ, Choe J, Chae EJ, Song JW. Acute exacerbation of fibrotic hypersensitivity pneumonitis: incidence

and outcomes. *Respir Res*. 2021;22(1):152. Published 2021 May 20. doi:10.1186/s12931-021-01748-2
40. Ejima M, Okamoto T, Suzuki T, Miyazaki Y. Role of serum surfactant protein-D as a prognostic predictor in fibrotic hypersensitivity pneumonitis. *Respir Investig*. 2022;60(3):369-378. doi:10.1016/j.resinv.2021.12.003
41. Yokoyama A, Kohno N, Hamada H, et al. Circulating KL-6 predicts the outcome of rapidly progressive idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 1998;158(5 Pt 1):1680-1684. doi:10.1164/ajrccm.158.5.9803115