ISSN: 2382-6533 (Print)
ISSN: 2424-810X (Online)
Volume 11 · Issue 6
December 2025

Journal of Clinical and Translational Research

Leveraging grouped data techniques in R to enhance binary dose-response analysis

Phase 2 dose-finding literature often reports dose-response information as binary outcomes versus dose. This data is used to understand the potency of the drug, which is characterized by the pharmacologic parameter ED50. Convenient approaches are available to pharmacometricians to analyze this type of data in R with summary-level and individual-level data using generalized linear or nonlinear models.

Summary-level (grouped) data

dose_group	dose_val	time	endpoint	n_subjs	prop_resp
0 mg	0	12	Endpoint A	50	0.08
1 mg	1	12	Endpoint A	50	0.24
5 mg	5	12	Endpoint A	50	0.54
10 mg	10	12	Endpoint A	50	0.84
20 mg	20	12	Endpoint A	50	0.84

```
gnm(formula = prop_resp ~ fx_emax(dose_val),
    weights = n_subjs,
    family  = binomial(link = "logit"),
    ...)
```

*fx_emax is a user-defined function for the sigmoidal Emax relationship

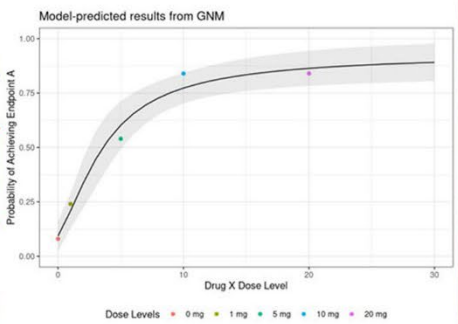
dose_group	dose_val	time	endpoint	n_success	n_fail
0 mg	0	12	Endpoint A	4	46
1 mg	1	12	Endpoint A	12	38
5 mg	5	12	Endpoint A	27	23
10 mg	10	12	Endpoint A	42	8
20 mg	20	12	Endpoint A	42	8

```
gnm(formula = cbind(n_success n_fail) ~ fx_emax(dose_val),
    family  = binomial(link = "logit"),
    ...)
```

Individual-level (ungrouped) data

ID	dose_group	dose_val	time	endpoint	response
1	0 mg	0	12	Endpoint A	1
2	0 mg	0	12	Endpoint A	1
3	0 mg	0	12	Endpoint A	1
4	0 mg	0	12	Endpoint A	1
5	0 mg	0	12	Endpoint A	0

```
gnm(formula = response ~ fx_emax(dose_val),
    family  = binomial(link = "logit"),
    ...)
```



All approaches provide identical model estimates on individual-level or summary-level binary response data, allowing pharmacometricians to perform analyses based on publicly available information from scientific publications or summary basis of approval documents from regulators.

ACCSCIENCE PUBLISHING

ABOUT JCTR

Aims and scope

The Journal of Clinical and Translational Research (JCTR) is an open access, peer-reviewed, multidisciplinary scientific journal that publishes studies with at least an ex vivo, in vivo, or clinical component. The published research is centered on any clearly defined clinical problem, which may comprise a disease or the basis of disease, a form of therapy or intervention, and clinical diagnostics or prognostics. Articles (original research, reviews, technical reports, medical hypotheses, commissioned articles, special issue articles, and editorials) are published continuously online and bimonthly in print. Studies performed in cells only will generally not be accepted unless they contain critical data that are in line with the scope of the journal. Some examples of such studies include molecular pathways that lie at the basis of a disease, novel biotechnological approaches for e.g., the production of drugs, or new techniques that improve clinical diagnostics and prognostics. Articles that combine preclinical and clinical data are given priority. Contributions from academic institutions and industry are welcome.

The research areas that JCTR covers include but are not limited to:

Internal medicine (all branches)
Vascular medicine and phlebology
Oncology
Cardiology
Intensive care medicine
Ophthalmology
Neurology and neurosciences
Anatomy, physiology, and embryology
Pathology
Clinical physics
Epidemiology
Medical devices
Pharmacology
Microbiology
Parasitology
Biomedical spectroscopy and spectrometry
Gastroenterology and hepatology
Surgery and transplantation
Hematology
Nephrology
Dermatology
Endocrinology and metabolism
Anesthesiology
Radiology and nuclear medicine
Clinical chemistry
Genetics and epigenetics
Global health
Nutrition
Immunology
Virology
Biomedical engineering

Key features

- Open access
- Reputable international editorial board
- Easy and fast submissions - no formatting rules (“your paper, your way”)
- No word count or reference restrictions
- Double blind review process to minimize bias
- Rapid online publication of articles upon acceptance
- Outlet for academic institutions and industry

Indexing

The Journal of Clinical and Translational Research is currently indexed by Chemical Abstract Service, Google Scholar, CNKI, and Peking University Library, and is currently working towards being indexed (PubMed, Science Citation Index Expanded, BIOSIS, Scopus, etc.).

Volume 11 • Issue 6 • December 2025
ISSN 2382-6533 (print) ISSN 2424-810X (online)

JOURNAL OF CLINICAL AND TRANSLATIONAL RESEARCH

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

Editorial Board

ACCSCIENCE PUBLISHING

Volume 11 | Issue 6 | December 2025 | Page 1 – 86 |

Journal of Clinical and Translational Research

https://accscience.com/journal/jctr

CONTENTS

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

EDITORIAL

Major issues in translational medicine

Jacek Z. Kubiak[1,2*]

[1]Institute of Genetics and Development of Rennes, UMR 6290 CNRS/University of Rennes, Faculty of Medicine, 35043 Rennes, France

[2]Laboratory of Molecular Oncology and Innovative Therapies, Military Institute of Medicine-National Research Institute(WIM-PIB), Szaserow 128, 04-141 Warszawa, Poland

The goal of modern medical research is to apply fundamental scientific findings and discoveries to develop treatments that help or cure diseases. This has become a core principle, often known as the "basic science model." However, this model is considered imperfect, as many effective treatments were developed before the underlying biological and molecular principles were understood. Recent scientific advances have not always led to real clinical improvements. To address this, the field of translational science has emerged to focus on accelerating the process of moving basic science discoveries into substantial health improvements for patients. This path, however, is not progressing as smoothly as one might expect, and the main problems are briefly discussed in this editorial.

Major problems in translational medicine include scientific challenges, such as the poor predictive power of results obtained with animal models in relation to humans, insufficient funding and other resource limitations, and regulatory hurdles that slow down or block the translation process. Other obstacles include real difficulties in interdisciplinary collaboration, a lack of standardized infrastructure, insufficient training, and the need for better communication between researchers, clinicians, and the public. These issues collectively contribute to the "translational gap," also known as the "Valley of Death," which is the gap between basic research and its potential clinical applications.[1,2] More precisely, it refers to the gap between potentially promising basic scientific discoveries obtained in laboratory practice and their successful application in the form of new treatments, therapies for patients, and eventual market availability. This gap is an important barrier in drug and therapy development, with high failure rates and numerous obstacles in translating models to humans.

The poor predictive power of animal models stems from significant biological differences between species.[3,4] This leads to a high failure rate for drugs in human trials. In both toxicology and efficacy, results from animal studies often fail to translate directly to human pathologies. Many safe compounds prove harmful in humans and *vice versa*. Factors contributing to this include poorly designed studies and the lack of key methodological safeguards, such as randomization or blinded outcome assessments.

Studies have shown that animal models are highly inconsistent at predicting eventual human toxicity. Their success rate may be very low for some toxic responses. Many drugs judged safe in animals are later found to be toxic in humans, and some safe human drugs may be eliminated because they appear dangerous in animals. These kinds of incompatibilities are often difficult to predict. Therefore, even when a drug is highly effective in an animal model, this does not guarantee that it will work in the same way in human trials. For example, many cancer vaccines effective

***Corresponding author:**
Jacek Z. Kubiak
(jacek.kubiak@univ-rennes1.fr)

Citation: Kubiak JZ. Major issues in translational medicine. *J Clin Transl Res*. 2025;11(6):1-3. doi: 10.36922/JCTR025480087

Received: November 27, 2025

Published online: December 9, 2025

in animal models have failed in large-scale human trials.[5] The high failure rate of drugs in human trials is a strong indicator of the poor predictive value of animal models. It is estimated that over 90% of drugs that seem safe and effective in animal models fail to be confirmed in human clinical trials.

The principal human contribution to these problems is poor study design. Many animal studies lack crucial elements found in clinical trials, such as randomization, blinding, and proper statistical analysis. This can lead to an overestimation of results obtained in fundamental animal studies. Other human-born shortcomings in translational research include inadequate funding and supply of resources, a lack of trained staff, complex and lengthy regulatory processes, and bureaucratic delays. Further issues involve cultural differences between basic scientists and clinicians, fragmented infrastructure, poor communication, and the inherent high risk and cost of moving discoveries from the lab to clinical practice.[6] The fundamental scientific challenge of the incomplete understanding of biological mechanisms also significantly impedes progress. Finally, the growing problems of scientific fraud, paper mill publications, and overinterpretation of results are increasingly hampering the efficiency of translational medicine at an increasing pace.

Two principal alternatives may help address the shortcomings outlined above: (i) Human-based research and (ii) novel, innovative approaches. A shift toward human-based research, such as personalized medicine and *in vitro* models, is gaining momentum. Human-based research is a broad category, including two main fields: Human subjects research, which studies living individuals through interventions or interactions, and human-relevant biotechnologies using human cells, tissues, or data to model human biology and predict human responses. While human subjects research is governed by regulations to protect participants, human-based biotechnologies use advanced technologies, such as organoids and computational models to develop more predictive, human-relevant research methods. Currently, there is an urgent need to develop and implement novel approaches to drug toxicity and efficacy studies that provide more reliable predictions for human responses.

Alternative methods in translational medicine include advanced technologies, such as omics (genomics, proteomics, metabolomics) for personalized medicine, novel pre-clinical models, such as three-dimensional organoids and "clinical trials in a dish (CTiD)" for faster drug screening, and different clinical trial designs, such as adaptive and pragmatic trials.[7] CTiD is a pre-clinical drug development method that tests the safety and efficacy of a given therapy on a representative sample of human cells or tissues in a laboratory setting before moving to human clinical trials. This approach, which uses technologies, such as induced pluripotent stem cells and organs-on-a-chip, aims to predict population-level responses to a drug, identify potential responders and non-responders, and reduce the high cost and failure rate of traditional clinical trials. In addition, "reverse translation," which moves findings from patient data back to the laboratory, and drug repurposing are used to accelerate the development of treatments, as clearly demonstrated during the COVID-19 pandemic.

Understanding the nature of challenges associated with translating animal model research into human clinical practice should be common among professionals in translational medicine. The brief description of the key issues included in this editorial can help beginners navigate these challenges and avoid potential errors.

Conflict of interest

Jacek Z. Kubiak is the Editor-in-Chief of this journal. The author declares that he has no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

References

1. Loh JS, Low CY, Low WP, *et al*. Traversing the Valley of Death for nanotechnology-based natural products: Strategies and insights from pharmaceutical stakeholders. *Drug Deliv Transl Res*. 2025;15(11):4126-4140.

   doi: 10.1007/s13346-025-01923-8

2. Failli V, Strittmatter SM, Schwab ME, *et al*. Crossing the valley of death in spinal cord injury: Learning from successful translators. *Neurotrauma Rep*. 2025;6(1):298-310.

   doi: 10.1089/neur.2025.0029

3. Zhu H, Yu J, Luo J, Cai Z, Li L, Zheng Q. Development of a cross-species model to predict clinical outcomes based on efficacy in mouse models of non-alcoholic fatty liver disease. *Clin Res Hepatol Gastroenterol*. 2025;49(9):102702.

   doi: 10.1016/j.clinre.2025.102702

4. Barroca NCB, Della Santa G, Suchecki D, García-Cairasco N, Umeoka EHL. Challenges in the use of animal models and perspectives for a translational view of stress and psychopathologies. *Neurosci Biobehav Rev*. 2022;140:104771.

   doi: 10.1016/j.neubiorev.2022.104771

5. Ostrand-Rosenberg S. Animal models of tumor immunity,

immunotherapy and cancer vaccines. *Curr Opin Immunol*. 2004;16(2):143-50.

doi: 10.1016/j.coi.2004.01.003

6. Ferreira RS Jr., Mantovani CK, Ferreira ASSBS, *et al*. Translational science at the undergraduate level: Awakening talents to overcome the valley of death - case report. *J Venom Anim Toxins Incl Trop Dis*. 2025;31:e20250005.

doi: 10.1590/1678-9199-jvatitd-2025-0005

7. Mir A, Zhu A, Lau R, *et al*. Applications, limitations, and considerations of clinical trials in a dish. *Bioengineering (Basel)*. 2024;11(11):1096.

doi: 10.3390/bioengineering11111096

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

REVIEW ARTICLE

Serum iodine reference ranges and their correlations with urinary iodine and thyroid function: A systematic review and meta-analysis

Lilan Wang*, Zixuan Ru, Shengnan Gao, Na Lv, Kerou Li, and Hong Qiao*

Department of Endocrinology, Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China

Abstract

Background: Serum iodine mainly exists in the form of iodine and iodide ions in thyroid hormones and does not change immediately due to variations in external factors. It is an important indicator reflecting iodine metabolism and actual iodine levels in the body. Given the significant differences in serum iodine among different countries and populations, the standard range of SIC has not yet reached a consensus. To the best of our knowledge, this systematic review and meta-analysis are the first to explore the serum iodine nutritional status of different populations across the country and its correlation with thyroid function. **Objective:** This study aimed to assess reference ranges of serum iodine concentration (SIC) and the potential correlations among SIC, urinary iodine concentration, and thyroid function in various populations globally, providing a reference for the assessment of individual iodine nutritional status. **Methods:** This study was conducted by searching multiple databases, including PubMed, Web of Science, and China National Knowledge Infrastructure, to gather relevant studies on SIC. Two researchers independently screened the literature and extracted and synthesized the data. The data were then analyzed using Stata software to generate pooled estimates of SIC with 95% confidence intervals (CIs) using forest plots. **Results:** A total of 48 eligible studies were included, yielding a mean SIC of 88.68 μg/L (95% CI: 84.70–92.65 μg/L). The range of SIC reported across studies was 23.92–183.50 μg/L. Among these, most studies focused on the correlations between SIC and free thyroxine (FT4) (81.0%) and between SIC and FT3 (72.7%). **Conclusion:** This meta-analysis highlights the importance of SIC as a valuable indicator of individual iodine nutritional status. The findings suggest that SIC is influenced by regional, temporal, and methodological variations, emphasizing the need for standardized testing and reference values. Moreover, the observed associations between SIC and thyroid function markers underscore its clinical relevance in monitoring iodine-related health outcomes. Establishing consistent measurement protocols and population-specific reference ranges will enhance the accuracy of iodine status assessments and support more effective public health interventions. **Relevance for Patients:** This study provides a preliminary synthesis of currently non-standardized SIC reference ranges. Future large-scale, multicenter studies on SIC testing are needed to establish reference values for assessing individual iodine nutritional status, thereby enabling scientifically informed and precision-based iodine supplementation.

Keywords: Population variability; Iodine nutritional status; Serum iodine; Medical reference values; Thyroid functions

***Corresponding authors:**
Hong Qiao
(qiaohong@hrbmu.edu.cn)
Lilan Wang
(202401443@hrbmu.edu.cn)

Citation: Wang L, Ru Z, Gao S, Lv N, Li K, Qiao H. Serum iodine reference ranges and their correlations with urinary iodine and thyroid function: A systematic review and meta-analysis. *J Clin Transl Res*. 2025;11(6):4-19.
doi: 10.36922/JCTR025260033

Received: June 29, 2025

Revised: September 9, 2025

Accepted: October 9, 2025

Published online: November 13, 2025

1. Introduction

The significance of scientifically evaluating iodine nutrition status and guiding targeted iodine supplementation for populations cannot be overstated. The demand for individual iodine status assessment is increasing, particularly among individuals with specific physiological conditions, such as children, pregnant women, and patients with thyroid diseases, who require timely and accurate evaluation of iodine status.[1] Indicators used to assess iodine status include dietary iodine intake, thyroid volume, thyroid function, urinary iodine concentration (UIC), and blood iodine concentration; however, all these indicators have certain limitations. For instance, recall bias may arise when investigating dietary intake; changes in thyroid volume require a longer period to manifest and are considered long-term indicators; thyroid function indicators are typically employed for patients with thyroid diseases and can be influenced by various factors.[2,3] Numerous previous studies have examined UIC in diverse populations.[4,5] Nevertheless, UIC reflects only the excretion of iodine from the body and does not represent the level of biologically active iodine utilized by the thyroid. Moreover, UIC is easily affected by factors such as dietary iodine intake, water iodine content, and sweating, resulting in considerable fluctuations and making it difficult to accurately reflect an individual's iodine nutritional status.[6] Serum iodine concentration (SIC) mainly exists in the form of iodine and iodide within thyroid hormones, reflecting the iodine metabolism in the body and the actual iodine level. Its fluctuations are relatively minor and less susceptible to external factors when compared with UIC, making it more stable. SIC can more precisely reflect recent iodine nutritional status and has unique advantages in accurately assessing individual iodine nutritional status. However, relatively few clinical studies on SIC have been conducted both domestically and internationally, and it is gradually emerging as a research focus.[7]

In other nations, renowned laboratories such as Mayo Clinic, Quest Diagnostics, and the World Health Organization (WHO) provide SIC reference ranges of 52–109 μg/L, 40–92 μg/L, and 45–90 μg/L, respectively, measured using inductively coupled plasma-mass spectrometry (ICP-MS). However, these ranges do not differentiate between populations.[8] In China, only a limited number of studies have determined SIC reference ranges for specific regions and populations. A multicenter cross-sectional study covering six provinces reported an SIC reference range of 36.0–79.3 μg/L, measured using ICP-MS. Nevertheless, the scientific community has not yet established a comprehensive SIC reference range.[9]

Given the substantial variations in SIC among countries and populations, there is currently no consensus on standard SIC ranges for diverse populations domestically or internationally. Hence, this systematic review and meta-analysis represent the inaugural effort to explore serum iodine status and its correlation with UIC and thyroid function across different populations, regions, sampling times, and detection methods. It aims to provide reference data for the future establishment of SIC reference ranges in various populations, thereby enabling scientific and precise iodine supplementation to maintain appropriate iodine nutritional levels.

2. Materials and methods

2.1. Inclusion criteria

The studies included in this meta-analysis were selected based on predefined criteria regarding study type, research subjects, and outcome indicators. Eligible studies comprised cross-sectional studies, case–control studies, longitudinal cohort studies, and prospective studies. The research subjects included pregnant women, children, infants, healthy adults, and individuals with abnormal thyroid function. To ensure relevance to the study objectives, eligible studies were required to report quantitative indicators of iodine nutritional status, such as mean values, quartiles, medians, or 95% confidence intervals (CI).

2.2. Exclusion criteria

Studies were excluded according to the following criteria: (i) studies published in languages other than Chinese or English; (ii) duplicate publications; (iii) studies without accessible full text or with incomplete data (e.g., SIC values only reported in figures without numerical data); (iv) case reports, review articles, conference abstracts, and animal studies; (v) studies in which subjects were exposed to environmental factors that could affect SIC (e.g., a high-iodine diet, iodine-based contrast agents, radioactive iodine, povidone-iodine disinfection, or amiodarone use within 3 days before testing, potentially resulting in iodine overload); (vi) studies with apparent statistical errors.

2.3. Search strategies for literature retrieval

A computer-based search was conducted in the China National Knowledge Infrastructure (CNKI), VIP, WanFang, PubMed, Web of Science, and Excerpta Medica dataBASE (Embase) databases. The search period was set from the establishment of each database to November 1, 2024, without country restrictions, although the language was limited to Chinese or English. In addition, other relevant databases were searched, including those maintained by the WHO (www.who.int/en/), the United Nations Children's Fund (UNICEF) (www.unicef.org/), and the International Council for the Control of Iodine Deficiency Disorders (ICCIDD) (www.iccidd.org/). The

search process employed a combination of subject terms and free-text words, with adjustments made according to the characteristics of each database. We also screened the reference lists of included studies ("snowballing") to identify additional eligible publications. The Chinese search terms encompassed: "serum iodine," "serum iodine concentration," "iodine nutrition assessment," "iodine nutritional status," "iodine nutrient level," "iodine nutrition monitoring," "evaluation of iodine nutritional status," "iodine nutrition survey," "iodine nutrition of residents," "iodine nutrition for children," "iodine nutrition for pregnant women," iodine nutrition in patients with thyroid disease, iodine nutrition in patients with thyroid dysfunction, and related terms. The English search terms consisted of: "serum iodine," "serum iodine concentration," "iodine nutrition assessment," "iodine nutritional status," "iodine nutrient level," "iodine nutrition monitoring," "evaluation of iodine nutritional status," "iodine nutrition survey," "iodine nutrition of residents," "iodine nutrition for children," "iodine nutrition for pregnant women," "iodine nutrition in patients with thyroid disease," "iodine nutrition in patients with thyroid dysfunction," and similar terms. For instance, the specific search strategy employed in PubMed is presented in Figure 1.

2.4. Literature screening and data extraction

Two researchers independently screened the literature, extracted the data, and cross-checked the results. Any disagreements were resolved through discussion or consultation with a third party. If necessary, the original study authors were contacted via email to obtain information that was not explicitly stated but considered crucial to the current study. During the literature screening process, the titles of the articles were initially inspected, and those that were obviously irrelevant were excluded from the study. The abstracts and full texts were then reviewed in detail to determine eligibility for inclusion. The extracted data encompassed: (i) basic information of the included studies, such as the first author, year of publication, country and region where the study was conducted, study type, and study duration; (ii) basic characteristics of the study subjects, including the type of population studied (e.g., pregnant women, adults, children, infants), the method employed for SIC detection, and the number of individuals tested; (iii) SIC (mean, quartiles, median, 95% CI), and the correlation between SIC, UIC, and thyroid function; (iv) key elements for bias risk assessment.

#1 Serum iodine [Mesh Terms]
#2 Serum iodine concentration [Title/Abstract]
#3 Iodine nutrition assessment [Title/Abstract]
#4 Iodine nutritional status [Title/Abstract]
#5 Iodine nutrient level [Title/Abstract]
#6 Iodine nutrition monitoring [Title/ Abstract]
#7 Evaluation of iodine nutritional status [Title/Abstract]
#8 Iodine nutrition survey [Title/Abstract]
#9 Iodine nutrition of residents [Title/Abstract]
#10 Iodine nutrition for children [Title/Abstract]
#11 Iodine nutrition for pregnant women [Title/Abstract]
#12 Iodine nutrition in patients with thyroid disease [Title/Abstract]
#13 Iodine nutrition in patients with thyroid dysfunction [Title/Abstract]
#14 #1 OR #2 OR #3OR #4OR #5 OR #6 #7 #8 #9 #10 OR #11 OR #12 OR #13

Figure 1. PubMed search strategy. Image created by the authors.

2.5. Quality assessment of included studies

Two researchers independently evaluated the risk of bias for the studies included in the review and cross-checked the outcomes. In case of any disagreement, a third reviewer was consulted to resolve discrepancies. The criteria recommended by the Agency for Healthcare Research and Quality were employed to assess the risk of bias in the cross-sectional studies. Each item of the criteria was rated as "yes," "no," or "unclear," with a score of 1 assigned for "yes" and 0 for "no." The total possible score was 11, with results ranging from 0 to 3 indicating low quality, 4–7 indicating medium quality, and 8 or above indicating high quality. The Newcastle–Ottawa Scale was utilized to evaluate the quality of the included cohort studies, case–control studies, and prospective studies, with each item awarded up to two stars. The overall quality score was computed as the sum of the individual star ratings. Total scores ranged from 0 to 10; scores 2–3 were considered low quality, 4–6 moderate quality, and 7–10 high quality. The quality assessment results for the studies included in this review were mainly concentrated on scores ranging from 6 to 8 (Table S1).

2.6. Statistical analysis

A systematic review and meta-analysis of the literature were carried out, summarizing SIC levels using mean ± standard deviation, quartiles, median, and 95% CI. Stata software (17.0, StataCorp, USA) was utilized for data cleaning and statistical analysis, and the "meta" command was employed to generate forest plots of the average SIC values and their corresponding 95% CI. Subgroup analyses were further conducted based on detection year, region, method, and population characteristics. The results were presented in forest plots to compare between-group heterogeneity. Heterogeneity was assessed using the I^2 statistic. An I^2 value of 0% indicates no observable heterogeneity, while values of 25%, 50%, and 75% represent low, moderate, and high heterogeneity, respectively. A random-effects model was applied when $I^2 > 50\%$; otherwise, a fixed-effects model was used. Statistical significance was set at $p<0.05$.

3. Results

3.1. The process and results of literature screening

Initially, a total of 2,394 relevant articles were identified for the following sources: PubMed (n = 488), Web of

Science (n = 358), Embase (n = 421), CNKI (n = 164), VIP (n = 466), WanFang Data (n = 472), and other sources (n = 25). NoteExpress (version 2.0, China) was employed to eliminate duplicate articles (n = 276), articles in languages other tn Chinese or English (n = 41), articles not pertinent to the research theme, case reports, systematic reviews, conference abstracts, and animal studies (n = 1915), as well as articles that were inaccessible in full, lacked complete data, or involved high iodine exposure before testing (n = 114). After sequential screening, a total of 48 studies[1-47] were ultimately included, as depicted in Figure 2.

3.2. Basic characteristics of the included studies

A total of 48 eligible studies were included, comprising 33 cross-sectional studies, four case–control studies, three longitudinal cohort studies, four prospective studies, two single-center studies, and two multicenter studies. Twenty studies reported the serum iodine nutritional status of pregnant women; 25 studies reported the serum iodine nutritional status of healthy adults (of which 16 additionally evaluated serum iodine nutritional status across early, middle, and late pregnancy stages, as indicated in Table S2); eight studies reported the serum iodine nutritional status in children or infants; and 11 studies reported the serum iodine nutritional status of patients with thyroid function disorders. The study populations encompassed pregnant women (n = 11,354), healthy adults (n = 13,563), children (n = 4,413), infants (n = 287), and patients with thyroid function disorders (n = 3,975). The quality score range of the studies incorporated in this meta-analysis was 5–9 points (Table S1).

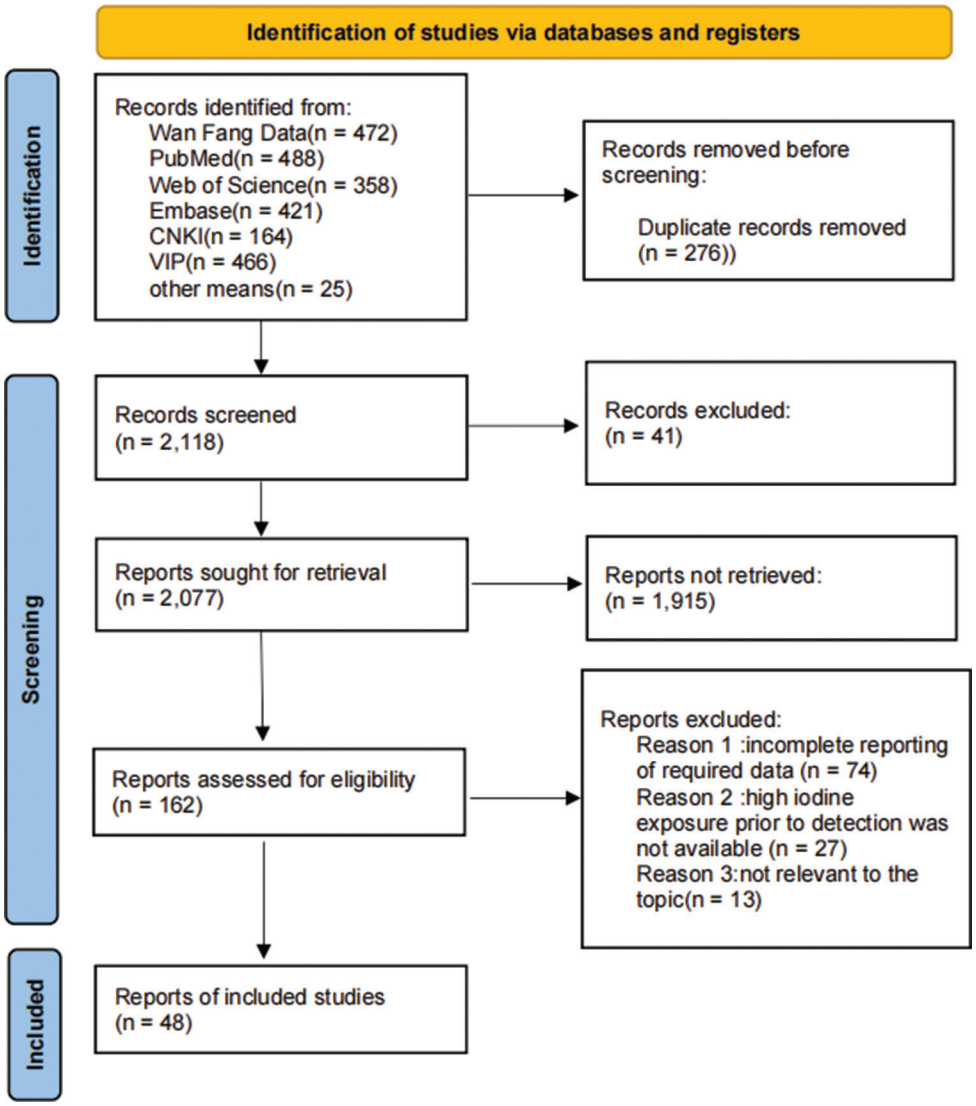


Figure 2. Flowchart illustrating the literature screening process and results. Image created by the authors.

3.3. Results of meta-analysis

3.3.1. Summary of average SIC values and 95% CIs across all populations

A meta-analysis was carried out to evaluate SIC levels across diverse populations, regions, detection times, and detection methods. The results were presented in a forest plot showing the average SIC values and corresponding 95% CI for each population (Figure 3). The analysis revealed heterogeneity among the studies (I^2 = 11.6%, p=0.238); therefore, a fixed-effects model was adopted. The average SIC concentration was 88.68 μg/L (95% CI: 84.70–92.65 μg/L) across all studies.

3.3.2. Subgroup analysis of average SIC values and 95% CIs by detection year

The forest plot in Figure 4 presents the average SIC values and 95% CIs by year of measurement. The results were as follows: 2017–84.02 μg/L (95% CI: 66.98–101.05 μg/L); 2019–85.54 μg/L (95% CI: 75.69–95.39 μg/L); 2020–113.85 μg/L (95% CI: 87.09–140.61 μg/L); 2021–103.31 μg/L (95% CI: 96.37–110.26 μg/L); 2022–68.50 μg/L (95% CI: 48.20–

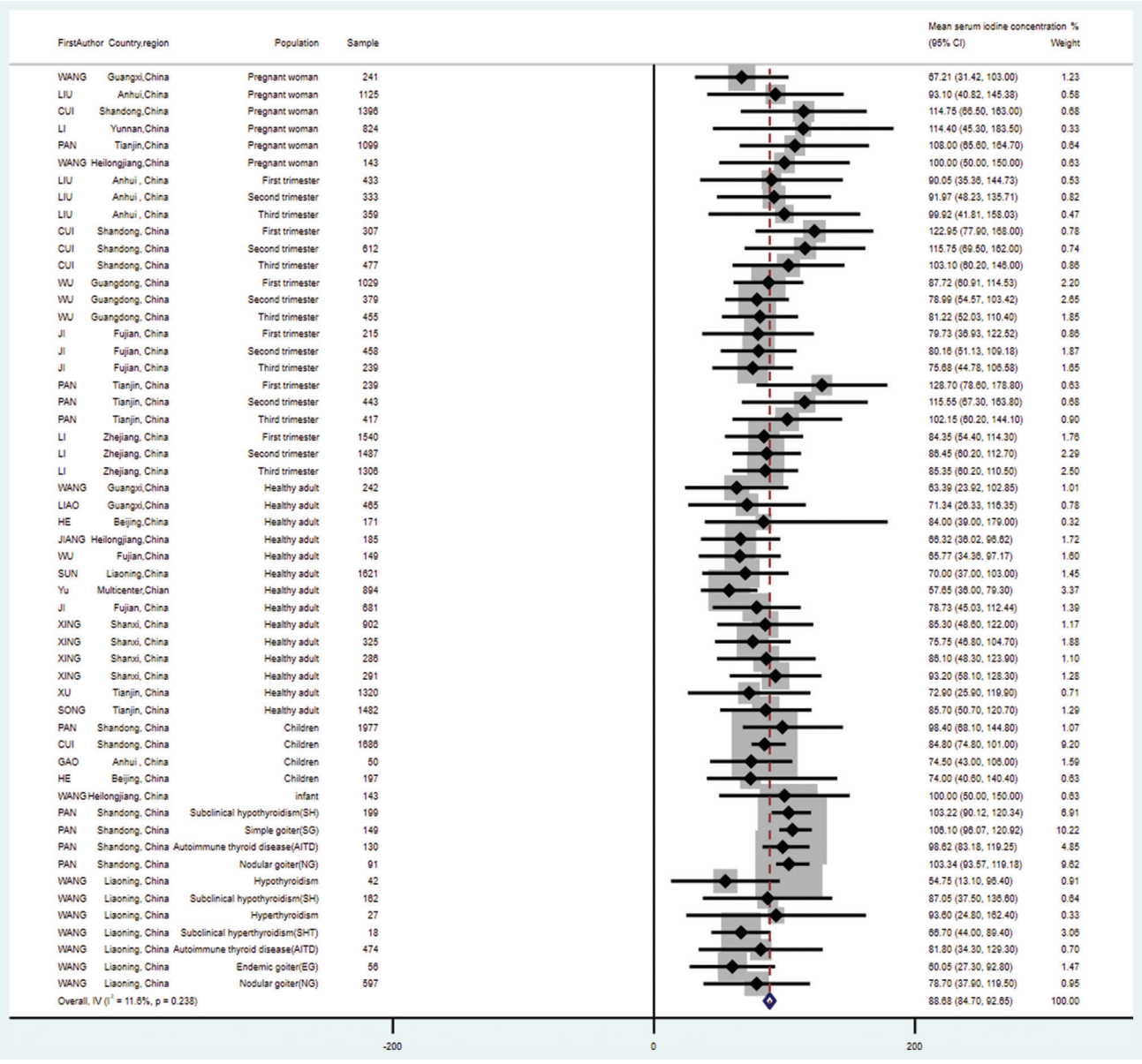


Figure 3. Forest plots showing the average serum iodine (SIC) concentration values and 95% confidence intervals (CIs) for different populations. Image created by the authors.

Notes: The hollow diamonds represent the pooled estimates for each study and the overall SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the squares is proportional to the relative weight of each study.

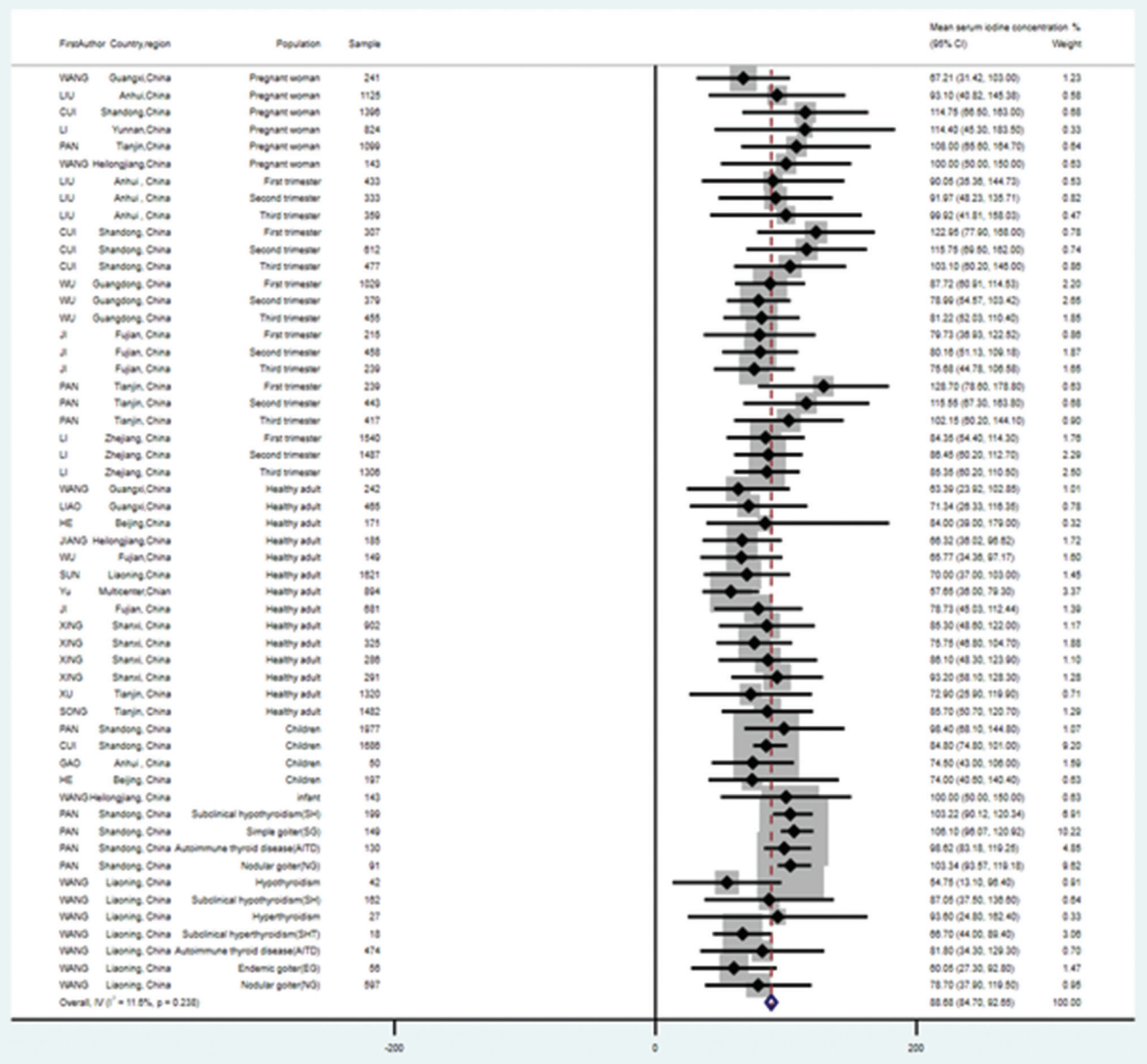


Figure 4. Forest plots showing the average serum iodine concentration (SIC) values and 95% confidence intervals (CIs) for different detection years. Image created by the authors.

Notes: The hollow diamonds represent pooled estimates for each study and the overall SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the squares is proportional to the relative weight of each study.

88.81 μg/L); 2023–82.92 μg/L (95% CI: 71.44–94.40 μg/L); 2024–78.41 μg/L (95% CI: 70.19–86.63 μg/L). There was significant heterogeneity among the subgroups (I^2 = 0.00%, p<0.001; Figure 4).

3.3.3. Subgroup analysis of average SIC values and 95% CIs in different detection areas

ArcGIS 10.8 mapping software (Esri, USA) was used to generate a weighted map of SIC detection levels for different populations across provinces in China (Figure 5). The studies involved the following provinces: Beijing (n = 8), Tianjin (n = 5), Shanghai (n = 1), Hebei (n = 1), Shanxi (n = 2), Liaoning (n = 2), Jilin (n = 1), Heilongjiang (n = 7), Jiangsu (n = 1), Zhejiang (n = 1), Anhui (n = 4), Fujian (n = 2), Shandong (n = 6), Henan (n = 1), Hubei (n = 2), Guangdong (n = 6), Hainan (n = 1), Guizhou (n = 1), Yunnan (n = 2), Guangxi Zhuang autonomous region (n = 3), Xinjiang Uygur autonomous region (n = 2). We found that Beijing, Heilongjiang, and Shandong had the most SIC detection studies, while no reports were found from Chongqing,

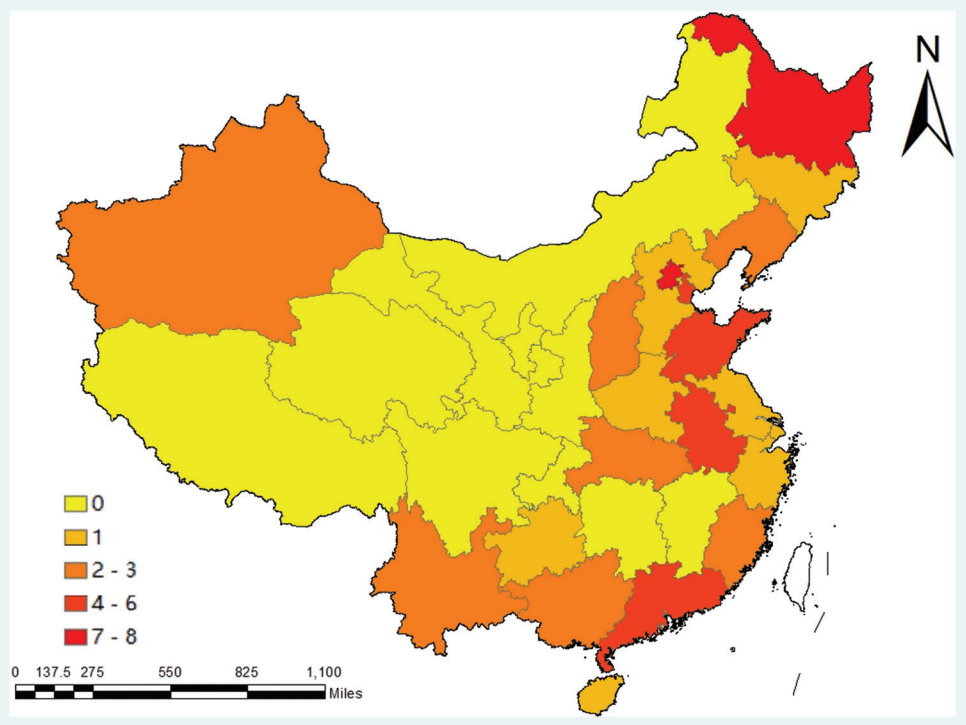


Figure 5. Serum iodine concentration detection levels for different populations in each province across China. Image created by the authors.

Hunan, Jiangxi, Sichuan, Shaanxi, Gansu, Qinghai, Inner Mongolia, Tibet, Ningxia, Hong Kong, and Macao.

The forest plots present the average values and 95% CIs of SIC across different provinces through subgroup analysis. We found that the average SIC in Guangxi Zhuang autonomous region was 66.99 μg/L (95% CI: 44.15–89.84 μg/L); in Anhui province, 85.84 μg/L (95% CI: 65.95–105.74 μg/L); in Shandong province, 100.24 μg/L (95% CI: 94.31–106.16 μg/L); in Tianjin municipality, 99.56 μg/L (95% CI: 81.52–117.60 μg/L); in Heilongjiang province, 80.56 μg/L (95% CI: 57.58–103.59 μg/L); in Guangdong province, 82.47 μg/L (95% CI: 67.12–97.83 μg/L); in Fujian province, 75.72 μg/L (95% CI: 61.09–90.34 μg/L); in Zhejiang province, 85.47 μg/L (95% CI: 69.94–100.99 μg/L); in Beijing municipality, 77.37 μg/L (95% CI: 36.74–116.00 μg/L); in Liaoning province, 69.65 μg/L (95% CI: 56.78–82.53 μg/L); and in Shanxi province, 84.02 μg/L (95% CI: 66.98–101.05 μg/L). The SIC varied among provinces (p<0.001; Figure 6).

3.3.4. Subgroup analysis of average SIC values and 95% CIs in different population groups

The forest plots in Figure 7 present the average SIC values and 95% CIs for different population groups. We observed that the average SIC for pregnant women was 94.02 μg/L (95% CI: 74.38–113.66 μg/L); for healthy adults, 72.88 μg/L (95% CI: 63.78–81.97 μg/L); for children, 84.11 μg/L (95% CI: 72.87–95.35 μg/L); for subclinical hypothyroid, 101.84 μg/L (95% CI: 87.39–116.30 μg/L); for autoimmune thyroid disease, 96.50 μg/L (95% CI: 79.64–113.36 μg/L); and for nodular goiter, 101.13 μg/L (95% CI: 88.91–113.35 μg/L). The SIC varied among different groups (p=0.007; Figure 7).

3.3.5. Average SIC values and 95% CIs during different pregnancy periods: Subgroup analysis results

The forest plots in Figure 8 present the average SIC values and 95% CIs for different trimesters of pregnancy in subgroups. We found that the average SIC for pregnant women in the first trimester was 93.88 μg/L (95% CI: 78.59–109.17 μg/L); in the second trimester, 88.04 μg/L (95% CI: 74.83–101.24 μg/L); in the third trimester, 88.99 μg/L (95% CI: 73.13–100.84 μg/L). Heterogeneity between groups (I^2 = 0.0%, p=0.784) was observed in the forest plots. Subgroup analysis of SIC averages and 95% CI for early, middle, and late pregnancies showed no statistically significant heterogeneity (I^2 = 0.0%, p=0.784) (Figure 8 and Table S3).

3.3.6. Subgroup analysis of average SIC values and 95% CIs by different detection methods

The forest plots in Figure 9 present the average SIC values and 95% CIs obtained using different detection methods.

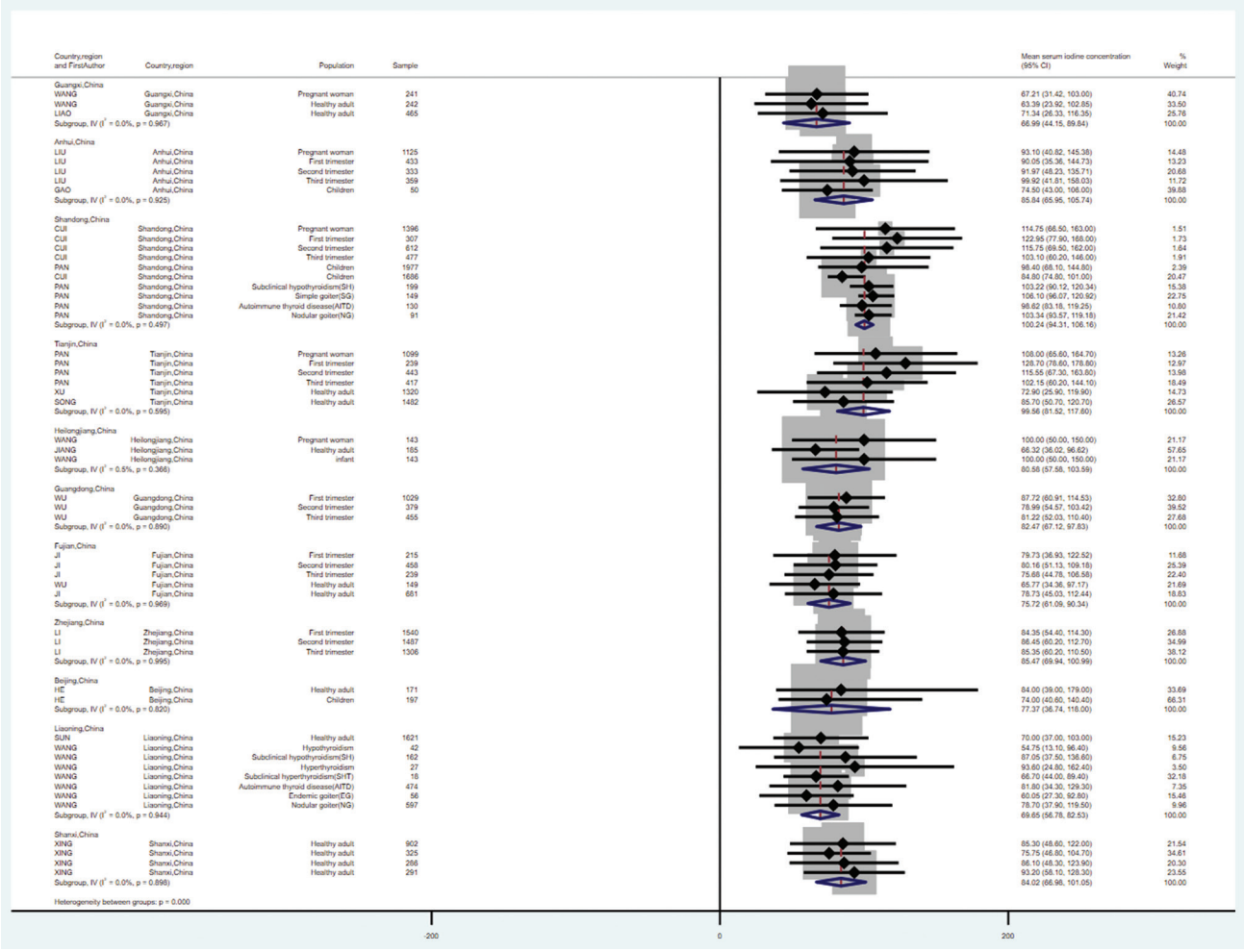


Figure 6. Forest plots showing the average serum iodine concentration (SIC) values and 95% confidence intervals (CIs) for studies from different detection regions. Image created by the authors.
Notes: The hollow diamonds represent the pooled estimates for each study and the overall SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the squares is proportional to the relative weight of each study.

We found that the average SIC detected by arsenic cerium catalytic spectrophotometry was 79.02 μg/L (95% CI: 70.85–87.18 μg/L), while that detected by ICP-MS was 91.67 μg/L (95% CI: 87.12–96.22 μg/L). There was heterogeneity between groups (I^2 = 0.0%, p=0.008; Figure 9).

3.3.7. Summary of the mean and 95% CIs of serum non-protein-bound iodine

According to the available statistics, only five studies have reported the nutritional status of serum non-protein-bound iodine (SnbI) in different populations. Figure 10 presents a forest plot summarizing the means and 95% CI values of SnbI for different populations, revealing an overall average SnbI value of 58.20 μg/L (95% CI: 54.28–62.11 μg/L; I^2=0.0%, p=0.512).

3.3.8. Relationship among SIC, UIC, and thyroid function

This meta-analysis also summarizes the correlations between SIC, UIC, and thyroid function. The range of SIC values reported was 23.92–183.50 μg/L. Within this range, we found the following: Studies reporting a positive correlation between SIC and UIC (n = 8), negative correlation (n = 0), and no correlation (n = 3); studies reporting a positive correlation between SIC and free thyroxine (FT) 3 (n = 8), negative correlation (n = 2), no correlation (n = 3); studies reporting a positive correlation between SIC and FT4 (n = 17), negative correlation (n = 1), no correlation (n = 3); studies reporting a positive correlation between SIC and total triiodothyronine (TT) 3 (n = 8), negative correlation (n = 0), and no correlation (n = 3); studies reporting a positive correlation between SIC and TT4 (n = 4), negative correlation (n = 0), and no correlation (n = 1); studies

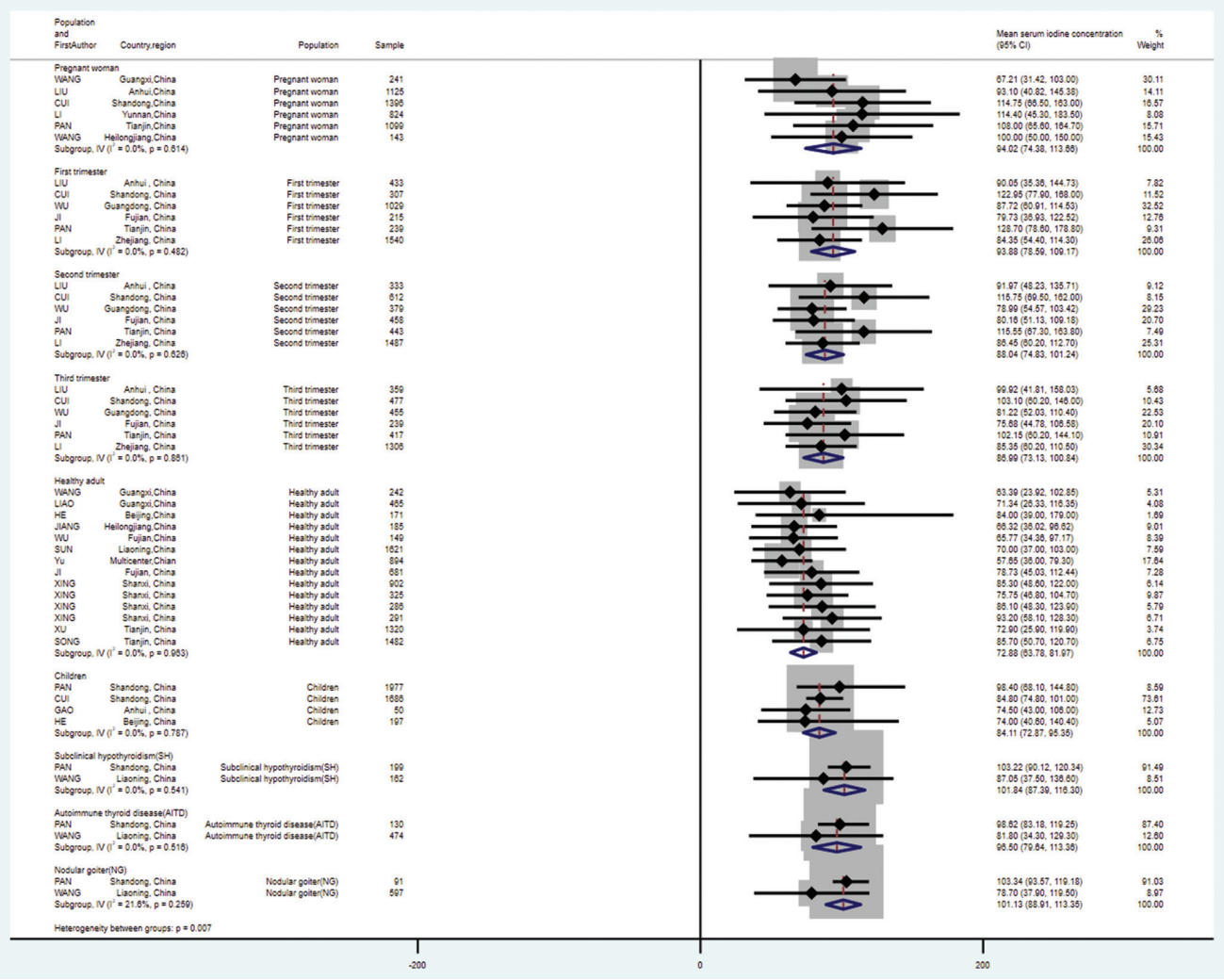


Figure 7. Forest plots showing the average serum iodine concentration (SIC) values and 95% confidence intervals (CIs) for different population groups. Image created by the authors.

Notes: The hollow diamonds represent the pooled estimate for each study and the overall estimate of SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the squares is proportional to the relative weight of each study.

reporting a positive correlation between SIC and thyroid-stimulating hormone (TSH; n = 5), negative correlation (n = 8), and no correlation (n = 11); studies reporting a positive correlation between SIC and thyroglobulin antibody (TgAb; n = 0), negative correlation (n = 1), and no correlation (n = 7); and studies reporting a positive correlation between SIC and thyroglobulin (Tg; n = 2), negative correlation (n = 0), and no correlation (n = 4); and studies reporting a positive correlation between SIC and thyroid peroxidase antibody (TPOAb; n = 0), negative correlation (n = 1), and no correlation (n = 7) (Table S4 and Figure 11).

4. Discussion

It is of great significance to scientifically assess iodine nutritional status and provide precise guidance for iodine supplementation in the population. There is a growing demand for individualized iodine nutritional status assessment, particularly among individuals with specific physiological needs, such as children, pregnant women, and patients with thyroid diseases, who require accurate, personalized evaluation of their iodine levels. Previous studies have established reference ranges for UIC. According to the standards set by WHO/UNICEF/ICCIDD, UIC values are classified as follows: Deficient iodine intake (<100 μg/L), adequate iodine intake (100–200 μg/L), more than adequate iodine intake (200–300 μg/L), and excessive iodine intake (>300 μg/L). However, the medical reference range of SIC remains undefined.[8] This systematic review and meta-analysis is the first to conduct a comprehensive analysis of SIC

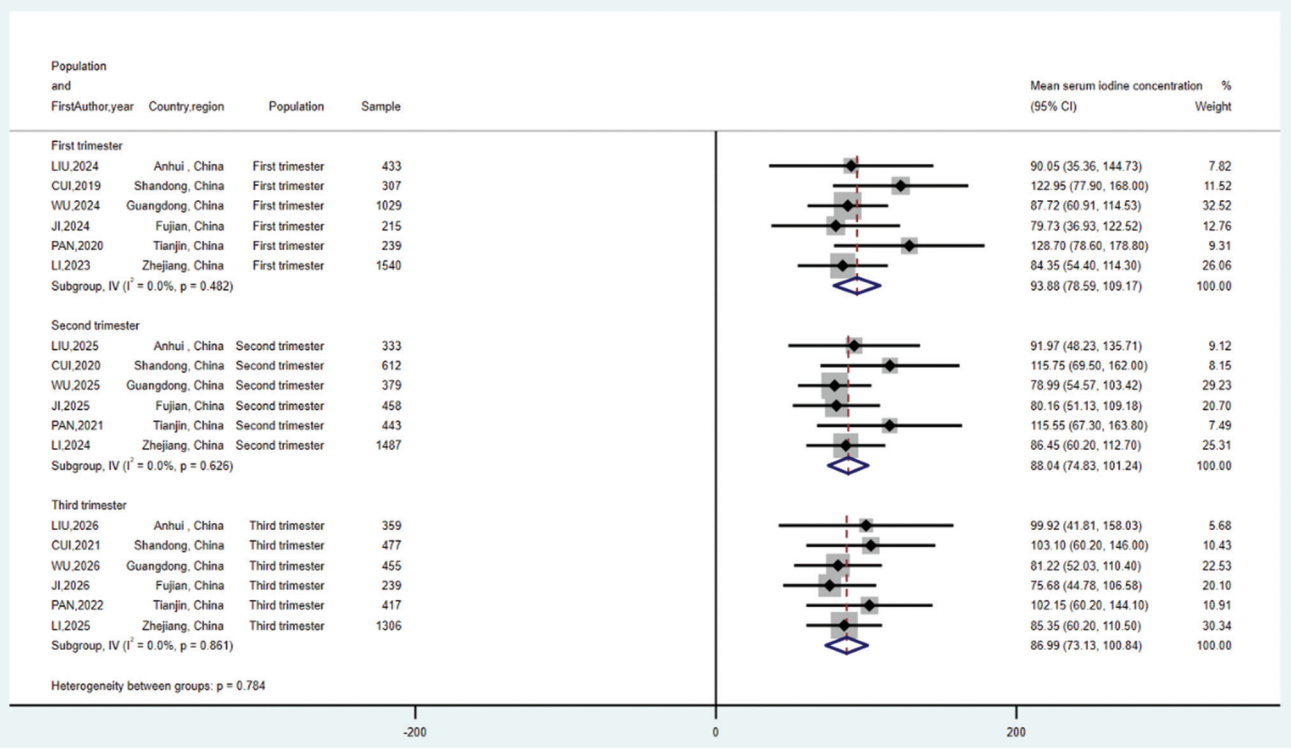


Figure 8. Average serum iodine concentration values and 95% confidence intervals for different gestational periods in pregnant women. Image created by the authors.

across different populations, regions, detection times, and detection methods. The range of SIC values reported in the included studies was 23.92–183.50 μg/L, and the average value and 95% CI of SIC in different populations were presented using forest plots (Figure 3). The meta-analysis found that the mean SIC value was 88.68 μg/L (95% CI: 84.70–92.65 μg/L). The reference ranges provided by the WHO, Mayo Clinic, and Quest Diagnostics, three internationally renowned laboratories, were 45–90 μg/L, 52–109μg/L, and 40–92μg/L, respectively, with validation rates of 80.6%, 80.6%, and 93.5%. A multicenter cross-sectional study in China preliminarily determined the reference range of SIC for healthy adults in China to be 36.0–79.3 μg/L.[6,7] Our pooled mean and its 95% CI are consistent with that study's reference range. However, due to limitations in study quality and quantity, a uniform standard reference range for SIC in medicine cannot be determined at present. In the future, both domestically and internationally, as well as across all provinces in China, a large number of multicenter, high-quality studies are needed to provide evidence for establishing a standard medical reference range for SIC.

This systematic review also performed a subgroup analysis of the average values and 95% CI of SIC for various detection years. The results indicated that the mean SIC and 95% CI of SIC were as follows: 2017, 84.02 μg/L (95% CI: 66.98–101.05 μg/L); 2019, 85.54 μg/L (95% CI: 75.69–95.39 μg/L); 2020, 113.85 μg/L (95% CI: 87.09–140.61 μg/L); 2021, 103.31 μg/L (95% CI: 96.37–110.26 μg/L); 2022, 68.50 μg/L (95% CI: 48.20–88.81 μg/L); 2023, 82.92 μg/L (95% CI: 71.44–94.40 μg/L); and 2024 78.41 μg/L (95% CI: 70.19–86.63 μg/L), with significant heterogeneity among the groups (I^2 = 0.0%, p<0.001). We found that the SIC levels varied across years. The SIC in 2020 was higher than in other years, which might be attributed to the fact that the population consisted entirely of pregnant women. During pregnancy, the body's metabolism and iodine consumption accelerate, while dietary intake also increases. The relatively higher SIC in 2021 may be related to the inclusion of patients with thyroid diseases, as iodine is closely associated with thyroid function. The SIC in 2022 was the lowest among the study years, possibly because the included populations were all healthy adults whose serum iodine levels were within the recommended range.

There was significant heterogeneity among the groups (I^2 = 0.0%, p<0.001). By employing ArcGIS mapping software to present a weighted map of the SIC survey results for each province throughout the country, we observed more studies from coastal and northern regions, while relatively fewer reports are available from the central and northwest regions of China. Among the reported studies, the largest amount of research was carried out in

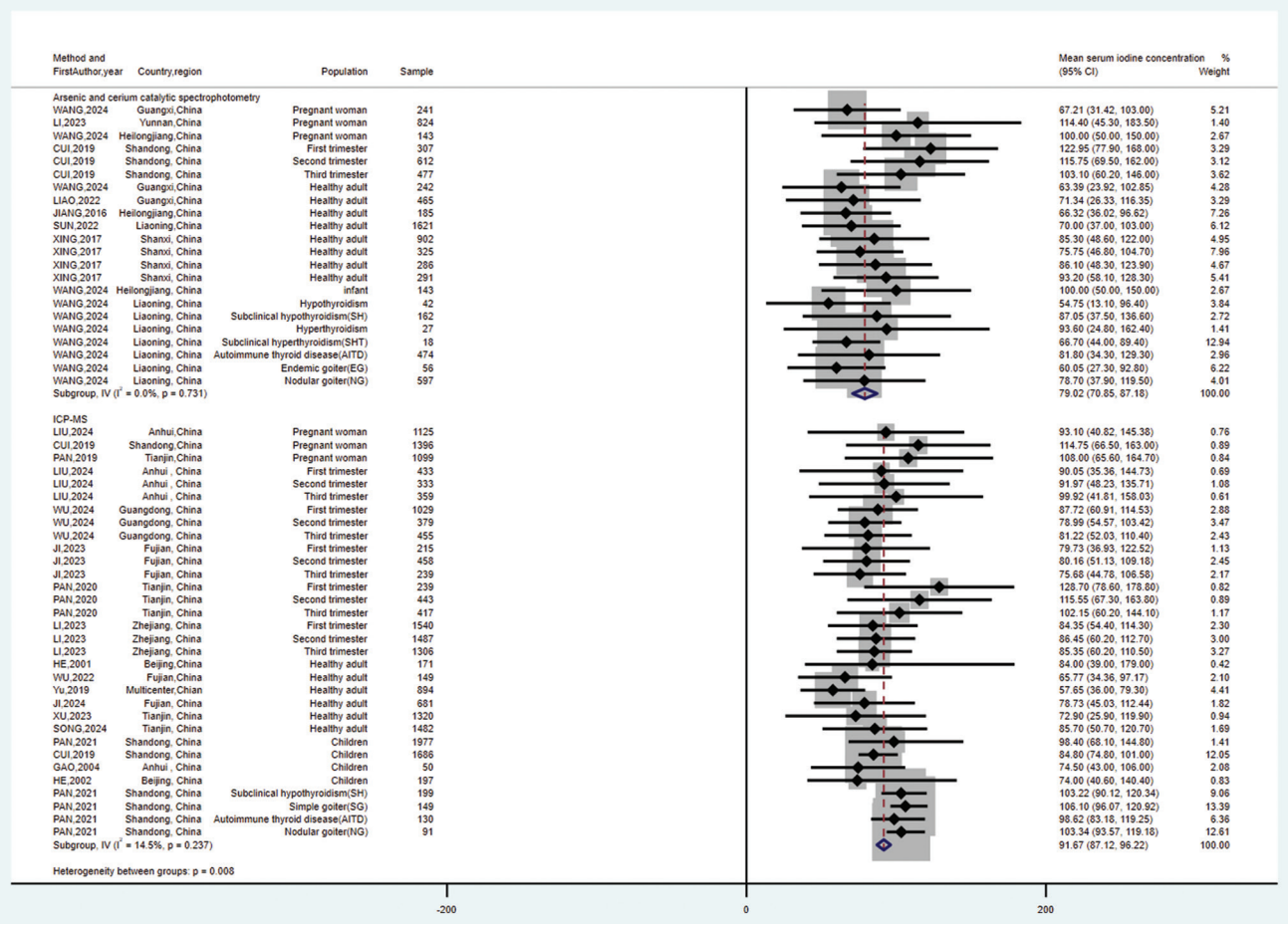


Figure 9. Forest plots depicting the average serum iodine concentration (SIC) values and 95% confidence intervals (CIs) for different detection methods. Image created by the authors.

Notes: The hollow diamonds represent the pooled estimates for each study and the overall SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the squares is proportional to the relative weight of each study.

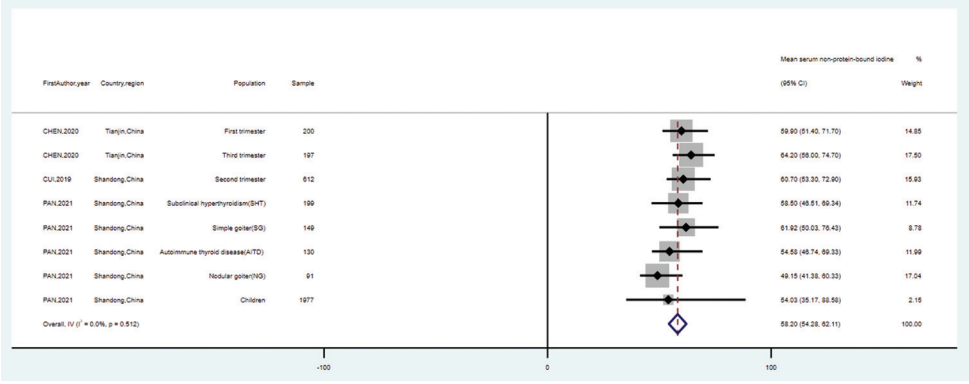
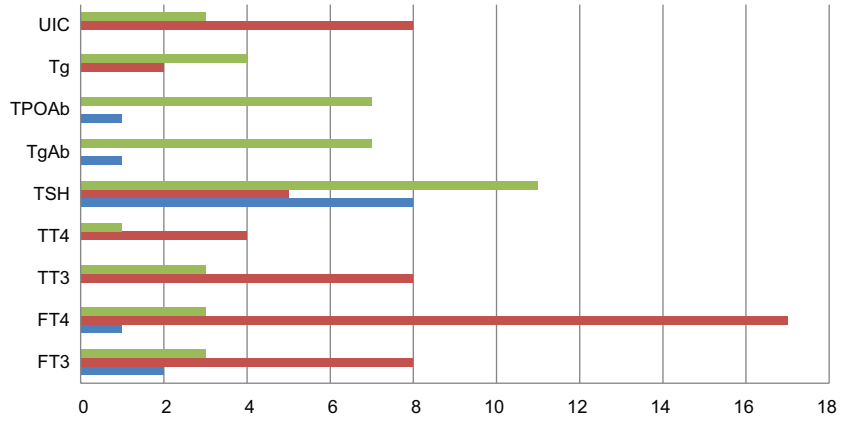


Figure 10. Means and 95% confidence intervals (CIs) of non-protein-bound serum iodine concentration (SIC) for different populations. Image created by the authors.

Notes: The hollow diamonds represent the pooled estimates for each study and the overall SIC concentration, while the solid diamonds represent point estimates (horizontal lines indicate 95% CIs). The size of the square is proportional to the relative weight of each study.

Beijing, Heilongjiang province, and Shandong province. In contrast, no eligible studies were identified from Chongqing, Hunan province, Jiangxi province, Sichuan province, Shaanxi province, Gansu province, Qinghai province, Inner Mongolia autonomous region, Tibet autonomous region, Ningxia Hui autonomous region,



	FT3	FT4	TT3	TT4	TSH	TgAb	TPOAb	Tg	UIC
Uncorrelated	3	3	3	1	11	7	7	4	3
Positive correlation	8	17	8	4	5	0	0	2	8
Negative	2	1	0	0	8	1	1	0	0

Figure 11. Summary chart of the relationship between serum iodine concentration (SIC), urinary iodine concentration (UIC), and thyroid function. Image created by the authors.
Abbreviations: FT: Free thyroxine; Tg: Thyroglobulin; TgAb: Thyroglobulin antibody; TPOAb: Thyroid peroxidase antibody; TSH: Thyroid-stimulating hormone; TT: Total triiodothyronine.

Hong Kong, or Macao special administrative region. Based on the reported studies, we observed that Tianjin and Shandong had higher SIC levels compared with other provinces, and most of the studies included thyroid disease patients and pregnant women. Nevertheless, these levels were all within the internationally recommended range of SIC. China implemented a universal salt iodization policy in 1996, which gradually resolved the long-standing issue of iodine nutritional deficiency and iodine deficiency disease in the country, making significant contributions to the prevention and control of iodine deficiency disorders worldwide. By 2010, 28 provinces had successfully eliminated iodine deficiency disease as a public health problem.[5] This study revealed that there was a statistically significant disparity in SIC levels among different regions; however, all were considered iodine-sufficient areas. Hence, it is essential to carry out nationwide SIC testing in the future to facilitate the establishment of a national reference range and support individualized iodine supplementation.

This systematic review conducted a subgroup analysis of the average SIC values and 95% CI values for different detection populations. The results showed that the average SIC value and 95% CI for pregnant women were 94.02 μg/L (95% CI: 74.38–113.66 μg/L); for the first trimester, 93.88 μg/L (95% CI: 78.59–109.17 μg/L); for the second trimester, 88.04 μg/L (95% CI: 74.83–101.24 μg/L); for the third trimester, 88.99 μg/L (95% CI: 73.13–100.84 μg/L); for healthy adults, 72.88 μg/L (95% CI: 63.78–81.97 μg/L); for children, 84.11 μg/L (95% CI: 72.87–95.35 μg/L); for subclinical hypothyroid, 101.84 μg/L (95% CI: 87.39–116.30 μg/L); for autoimmune thyroid disease, 96.50 μg/L (95% CI: 79.64–113.36 μg/L); and for nodular goiter, 101.13 μg/L (95% CI: 88.91–113.35 μg/L), with significant heterogeneity among the groups (I^2 = 0.0%, p=0.007). We determined that the SIC level of pregnant women was higher than that of healthy adults and children. In addition, when subgroup analysis was performed based on different stages of pregnancy, we observed that the SIC level was relatively higher in early pregnancy; however, there was no statistically significant difference among the groups (p=0.784). Nevertheless, most previous studies have found a statistically significant variance in SIC levels across different pregnancy periods (H = 26.20, p<0.05), with the median SIC in early pregnancy being significantly lower than that in middle and late pregnancy (z = −4.285, −2.880, p<0.05).[10,11] At present, due to the challenges associated with sample collection from infants, assessing iodine nutritional status in infants remains difficult. Thus, no optimal standard for evaluating the nutritional status of infants using SIC has been established, and future research should focus on this population. Iodine is an indispensable element for the synthesis of thyroid hormones, and both excessive and deficient iodine can lead to thyroid diseases of varying degrees. We observed that the SIC level of patients with thyroid diseases was generally elevated. Jin *et al.*[39] investigated the relationship between SIC and thyroid diseases and concluded that SIC >100 μg/L was associated with thyroid disorders. By drawing receiver

operating characteristic curves to verify the consistency between thyroid function and serum iodine, it was found that SIC has significant diagnostic value for thyroid dysfunction and higher accuracy than UIC. Previous studies have demonstrated that thyroid diseases can be triggered by either insufficient or excessive iodine intake. However, current research on thyroid diseases remains relatively scarce, and no studies have comprehensively examined the SIC levels of patients with various types of thyroid diseases. Future research focusing on this aspect could facilitate individualized management for thyroid disease patients.[48]

This systematic review performed a subgroup analysis of the average SIC values and 95% CI values for different detection methods. The results indicated that the average SIC value and 95% CI of arsenic–cerium catalytic spectrophotometry were 79.02 μg/L (95% CI: 70.85–87.18 μg/L), while those of ICP-MS were 91.67 μg/L (95% CI: 87.12– 96.22 μg/L), with significant heterogeneity among the groups (I^2 = 0.0%, p=0.008). We observed that the average SIC value and 95% CI determined via ICP-MS were higher than those determined by arsenic–cerium catalytic spectrophotometry. It is hypothesized that this difference might be influenced by factors such as sample size, detection time, personnel, and reagent types. At present, there are numerous methods for determining SIC, including chromatographic methods (gas chromatography, ion chromatography, high-performance liquid chromatography), the neutron activation method, the alkali ash method, and the graphite furnace atomic absorption method. Nevertheless, the instruments employed in these methods are costly, the conditions for sample decomposition are stringent, and the technical requirements for operators are high, making it challenging to control the precision of results and hindering their widespread application in China.[29,34] Arsenic–cerium catalytic spectrophotometry, a classic method for trace iodine detection recommended by the health industry, offers the advantages of low instrument cost, simple operation, high detection sensitivity, and good accuracy. It is well-suited for large-scale testing of numerous samples in basic laboratories. ICP-MS, a widely utilized micro-iodine detection and analysis technique in the current analysis field, requires no sample digestion and is characterized by rapid analysis, high sensitivity, and excellent precision and accuracy. The iodine standard curve range set by this method is 0–300 μg/L, which can meet the detection needs for low-, normal-, and high-iodine serum levels and fulfill the requirements for individualized iodine nutritional evaluation in clinical practice.[45] Future research trends will likely continue to focus on the detection and analysis of SIC using arsenic–cerium catalytic spectrophotometry or ICP-MS.

This systematic review also summarized the average values and 95% CI of SIC for different detection sample types. The results revealed that five articles reported the nutritional status of SnbI in different groups. We found that the average SnbI value and 95% CI were equivalent to the overall average SnbI value and 95% CI of 58.20 μg/L (95% CI: 54.28–62.11 μg/L) (I^2 = 0.0%, p=0.512). There are extremely few studies on SnbI, protein-bound serum iodine, plasma iodine, and whole blood iodine; thus, more extensive research in this field is warranted in the future. The SnbI levels during the various trimesters of pregnancy were 60.3 (51.0–67.8) μg/L, 54.0 (47.0–62.6) μg/L, and 54.4 (45.7–62.4) μg/L, respectively, showing a decline as pregnancy progressed (p<0.001). The serum total iodine concentration and the SnbI concentration were negatively correlated. SnbI was positively correlated with TSH in children (p<0.05) and with serum Tg concentration (p<0.05). SnbI was negatively correlated with TSH in pregnant women (p<0.05) and positively correlated with serum Tg concentration in both pregnant women and children. Nevertheless, the number of such studies was too limited for further discussion in this meta-analysis. Finally, this review summarized the correlation between SIC and thyroid function, as well as the level of UIC. The range of SIC values examined was 23.92–183.50μg/L. Within this range, the number of studies investigating the correlation between SIC level and FT4 or FT3 was the largest, with the proportions of positive correlation being 81.0% and 72.7%, respectively. This can be explained by the fact that iodine serves as a substrate for the synthesis of thyroid hormones and plays a key role in regulating and maintaining normal thyroid function. The correlations between SIC level and UIC, TT4, TT3, and Tg were either positive or not significant, with no negative correlations reported. The correlations between SIC and TPOAb or TgAb were mostly not significant (87.5% for both), indicating that the association between serum iodine and thyroid antibodies is weak. The correlation between SIC and TSH was inconsistent, with proportions of positive, negative, and no correlation being 20.8%, 33.3%, and 45.8%, respectively, making it difficult to determine whether SIC is a directional marker of thyroid function. Future studies with larger sample sizes should be conducted to explore this further.

The main advantage of this study is that it conducted subgroup analyses for different populations, regions, detection years, and methods, which facilitated the summarization and analysis of reference value ranges for SIC across various levels and populations. However, due to the influence of research methods, study populations, sample sizes, and other factors on the different subgroups, the sources of heterogeneity could not be determined through meta-analysis. The use of different measurement

instruments, standards, and personnel in SIC assessment may lead to discrepancies in study results, and there is a lack of survey data on other possible influencing factors. Therefore, in the future, it is necessary to redefine the reported iodine nutritional status using unified standards, and more studies are needed to establish a comprehensive medical reference range for SIC.

In summary, this systematic review encompassed all studies reporting SIC mean values and 95% CI summary results: 88.68 μg/L (95% CI: 84.70–92.65μg/L) (I^2 = 11.6%; p=0.238), and the mean value and 95% CI of all SnbI were 58.63μg/L (95% CI: 54.74–62.52 μg/L) (I^2 = 20.1%, p=0.284). The meta-analysis results indicated that the mean value and 95% CI of SIC differed across different detection years, regions, and populations (p<0.001) and also varied according to detection methods (p=0.008). Furthermore, the included studies in this systematic review reported SIC levels ranging from 23.92 to 183.50 μg/L. Within this range, the largest number of studies examined the relationship between SIC levels and FT4 and FT3, with proportions of positive correlation being 81.0% and 72.7%, respectively. The correlations between SIC levels and UIC, TT4, TT3, and Tg were all positive or not significant, with no studies reporting negative correlations. SIC was mostly unrelated to TPOAb and TgAb, while its correlation with TSH was inconsistent, with proportions of positive, negative, and no correlation being 20.8%, 33.3%, and 45.8%, respectively. Limitations of this study include: (i) Due to factors such as research methodology, study population source, and sample size, the inclusion of non-randomized or heterogeneous studies may have introduced selection bias; (ii) differences in measurement instruments, standards, and personnel for serum iodine assessment may have led to variations in study results; (iii) there was a lack of survey data on other potential influencing factors.

In summary, this systematic review preliminarily highlighted the current absence of a complete and uniform medical reference range for SIC. However, due to the limited quality and quantity of the included studies, these findings still require validation through high-quality, large-scale epidemiological studies. Future studies should be conducted across multiple centers with large sample sizes to provide reference data for individualized iodine nutrition assessment, thereby supporting scientific and precise iodine supplementation.

5. Conclusion

The pooled analysis showed that the mean SIC value and 95% CI were 88.68 μg/L (95% CI: 84.70–92.65 μg/L). The average value and 95% CI of SIC vary depending on the testing year, region, population and testing method. The range of SIC included in this meta-analysis was 23.92–183.50 μg/L. Among them, most studies investigated the correlation between SIC and FT4 or FT3, of which 81.0% and 72.7% demonstrated positive correlations.

Acknowledgments

None.

Funding

This research was supported by the Non-communicable Chronic Diseases-National Science and Technology Major Project (grant numbers: 2024ZD0537900; 2024ZD0537907), the National Natural Science Foundation of China (grant number: 82373698), and the Key Research and Development Program (Innovation Base) of Heilongjiang Province (grant number: GY2024JD0040).

Conflict of interest

The authors declare that they have no competing of interests.

Author contributions

Conceptualization: Hong Qiao
Formal analysis: Lilan Wang
Methodology: Zixuan Ru
Visualization: Na Lv
Writing–original draft: Lilan Wang
Writing–review & editing: Hong Qiao, Shengnan Gao

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Roche J, Michel R. Thyroid hormones and iodine metabolism. *Annu Rev Biochem*. 1954;23:481-500.
doi: 10.1146/annurev.bi.23.070154.002405
2. Liu Z, Lin Y, Wu J, *et al*. Is the urinary iodine/creatinine ratio applicable to assess short term individual iodine status in Chinese adults? Comparison of iodine estimates from 24-H urine and timed-spot urine samples in different periods of the day. *Nutr Metab* (*Lond*). 2022;19(1):27.
doi: 10.1186/s12986-022-00656-6
3. Guo W, Tan L, Dong S, *et al*. New Reference Values for thyroid volume and a comprehensive assessment for influencing factors in Chinese adults with iodine sufficiency.

Eur Thyroid J. 2021;10(6):447-454.

doi: 10.1159/000513494

4. Li Y, Shan Z, Teng W, Thyroid Disorders, Iodine Status and Diabetes Epidemiological Survey Group. The iodine status and prevalence of thyroid disorders among women of childbearing age in China: National cross-sectional study. *Endocr Pract*. 2021;27(10):1028-1033.

doi: 10.1016/j.eprac.2021.03.017

5. Shan Z, Chen L, Lian X, *et al*. Iodine status and prevalence of thyroid disorders after introduction of mandatory universal salt iodization for 16 years in China: A cross-sectional study in 10 cities. *Thyroid*. 2016;26(8):1125-1130.

doi: 10.1089/thy.2015.0613

6. Pearce EN, Caldwell KL. Urinary iodine, thyroid function, and thyroglobulin as biomarkers of iodine status. *Am J Clin Nutr*. 2016;104:898S-901S.

doi: 10.3945/ajcn.115.110395

7. Allain P, Berre S, Krari N, *et al*. Use of plasma iodine assay fordiagnosing thyroid disorders. *J Clin Pathol*. 1993;46(5):453-455.

doi: 10.1136/jcp.46.5.453

8. WHO/UNICEF/ICCIDD. WHO/UNICEF/ICCIDD. *Assessment of Iodine Deficiency Disorders and Monitoring their Elimination. A Guide For Programme Managers*. 3rd ed. Geneva: WHO; 2007.

9. Yu S, Wang D, Cheng X, *et al*. Establishing reference intervals for urine and serum iodine levels: A nationwide multicenter study of a euthyroid Chinese population. *Clin Chim Acta*. 2020;502:34-40.

doi: 10.1016/j.cca.2019.11.038

10. Wang F, Liao M, Lu H, Luo L. Establishment of SIC reference range for women in guangxi coastal areas. *China J Endemic Dis Control*. 2024;39:274-276.

11. Liu Y, Li W, Yu C, Xu X, Xu S, Tian C. Study on the level of SIC in pregnant women in Anhui Province and reference range. *Chin J Endemic Dis Control*. 2024;39:5-8.

12. Shi M, Li X, Liu Y. Study on the level of SIC in pregnant women in Harbin and the specific reference value range of thyroid hormones. *Chin J Endemic Dis Control*. 2023;38:181-184.

13. Nie J. *Feasibility Study on the Evaluation of Iodine Nutritional Status in Pregnant Women Using Saliva Iodine*. Xinjiang: Xinjiang Medical University; 2023.

14. Li R, Chen W, Liu Y, *et al*. The impact of preconceptional hysterosalpingography with oil-based contrast on maternal and neonatal iodine status. *Reprod Sci*. 2021;28:2887-2894.

doi: 10.1007/s43032-021-00640-0

15. Shi L. *Research on the Correlation between Iodine Nutrition and Thyroid Diseases in the Population with Higher Water Iodine Levels in Xiangyun County*. China: Dali University; 2021.

doi: 10.27811/,dcnki.Gdixy.2021.000147

16. Chen Y. *Investigation on Iodine Nutrition in Pregnant Women and Its Relationship with the Growth and Development of Infants and Young Children*. China: Tianjin Medical University; 2020.

doi: 10.27366/,dcnki.Gtyku.2020.000093

17. Cui T. *Study on the Relationship Between SIC in Children and Pregnant Women and their Iodine Nutrition and Thyroid Function*. China: Tianjin Medical University; 2019.

18. Li S, Peng Li, Huang W, *et al*. Relationship between pregnancy reaction and iodine nutrition and pituitary-thyroid function the first trimester. *Chin J Public Health*. 2003;1:41-42.

19. Huang W, Li S, Wang J, Wei R, Luo R, Wang X. Iodine nutrition and pituitary-thyroid function in early pregnant women. *Chin J Public Health*. 2002;11:14-15.

20. Yang X, Cheng L, Li M, Wu W, Tang Y. A survey study on iodine nutrition status in pregnancy. *Guangxi Med J*. 2000;3:642-644.

21. Yu S, Yin Y, Cheng Q, *et al*. Validation of a simple inductively coupled plasma mass spectrometry method for detecting urine and serum iodine and evaluation of iodine status of pregnant women in Beijing. *Scand J Clin Lab Invest*. 2018;78:501-507.

doi: 10.1080/00365513.2018.1512150

22. Li J, Wu H, Ye F, *et al*. Yunnan province Xiong county and town industry in Jianshui adult and iodine nutritional status analysis. *Chin Local Epidemiol Magaz*. 2023;10:803-807.

doi: 10.3760/cma.J.c.n231583-20220829-00299

23. Ji S, Wu X, Wu J, Chen D, Chen Z. Serum iodine concentration and its associations with thyroid function and dietary iodine in pregnant women in the southeast coast of China: A cross-sectional study. *Front Endocrinol* (*Lausanne*). 2023;14:1289572.

doi: 10.3389/fendo.2023.1289572

24. Pan Z, Cui T, Chen W, *et al*. Serum iodine concentration in pregnant women and its association with urinary iodine concentration and thyroid function. *Clin Endocrinol* (*Oxf*). 2019;90:711-718.

doi: 10.1111/cen.13945

25. Wang S, Bu Y, Shao Q, Cai Y, Sun D, Fan L. A cohort study on the effects of maternal high serum iodine status during pregnancy on infants in terms of iodine status and intellectual, motor, and physical development. *Biol Trace Elem Res*. 2024;202:133-144.

doi: 10.1007/s12011-023-03677-1

26. Fu M, Ren Z, Gao Y, Zhang H, Guo W, Zhang W. Study of iodine transport and thyroid hormone levels in the human

placenta under different iodine nutritional status. *Br J Nutr*. 2024;131:1488-1496.

doi: 10.1017/s0007114524000084

27. Kazi TG, Kandhro GA, Sirajuddin, *et al*. Evaluation of iodine, iron, and selenium in biological samples of thyroid mother and their newly born babies. *Early Hum Dev*. 2010;86:649-655.

doi: 10.1016/j.earlhumdev.2010.07.010

28. Liao M, Ning R, Luo L, Lu H, Wang F. Preliminary exploration of the reference range of SIC in adults with normal thyroid function in Guangxi, China, 2021. *Appl Prevent Med*. 2022;28:515-517.

doi: 10.3969/j.issn.1673-758X.2022.06.004

29. Wang W, Lin D, Zhang X, Zheng S. A study on the correlation between urine and serum levels of iodine and BPA and the incidence of goiter. *Chin J Endemic Dis Control*. 2016;31:377-379.

30. Wang J, Sun L, Kan Z, *et al*. Serum iodine levels and influencing factors of adults with different thyroid health conditions. *Chin J Endemic Dis*. 2023;42(6):502-506.

doi: 10.3760/cma.j.cn231583-20220719-00264

31. He J, Xu X, Zhou H, Cui K, Fan J, Fu C. Serum iodine levels in 368 healthy individuals in urban areas of Beijing. *Trace Elements Health Res*. 2001;(3):55-57.

32. Yang L, Yao Z, Wang Z, *et al*. Comparative study on iodine content in serum, plasma and whole blood. *Chin J Endemic Dis*. 2023;42(6):502-506.

doi: 10.3760/cma.j.cn231583-20220517-00172

33. Bai J, Liu H, Xiong C, Zhu K, Ma Q, Liu X. The value of using inductively coupled plasma mass spectrometry-based detection technology to screen thyroid cancer elemental biomarkers. *Chin J Med Equipment*. 2024;21:29-35.

doi: 10.3969/j.issn.1672-8270.2024.08.006

34. Jiang P, Lu Z, Jin X. *et al*. Initial exploration of reference range of serum iodine concentration in adults with normal thyroid function. *Chin J Endemic Dis*. 2016;35:786-789.

doi: 10.3760/cma.j.issn.2095-4255.2016.11.002

35. Wu X, Chen Z, Wu J, Wang M. Establishment of medical reference range of thyroid function in adults with normal SIC in Fujian Province. *Chin J Endemic Dis*. 2022;41:186-188.

doi: 10.3760/cma.j.cn231583-20210416-00126

36. Sun L, Wang J, Kan Z, Yang Y, Su M, Liu C. A medical reference value and correlation between SIC Content in normal thyroid function adults in liaoning province and thyroid function indicators. *Chin J Endemic Dis*. 2022;41:440-443.

doi: 10.3760/cma.j.cn231583-20210611-00207

37. Jin X, Jiang P, Liu L, *et al*. The application of serum iodine in assessing individual iodine status. *Clin Endocrinol* (*Oxf*). 2017;87:807-814.

doi: 10.1111/cen.13421

38. Fan L, Bu Y, Chen S, *et al*. Iodine nutritional status and its asso95%CIations with thyroid function of pregnant women and neonatal TSH. *Front Endocrinol* (*Lausanne*). 2024;15:1394306.

doi: 10.3389/fendo.2024.1394306

39. Xu T, Guo W, Ren Z, Wei H, Tan L, Zhang W. Study on the relationship between serum iodine and thyroid dysfunctions: A cross-sectional study. *Biol Trace Elem Res*. 2023;201:3613-3625.

doi: 10.1007/s12011-022-03459-1

40. Song Q, Xu T, Wang Y, *et al*. Exploring the correlation between varied serum iodine nutritional levels and anti-thyroglobulin antibodies. *Biol Trace Elem Res*. 2024;203(3):1362–1374.

doi: 10.1007/s12011-024-04275-5

41. Li X, Tu P, Gu S, *et al*. Serum iodine as a potential individual iodine status biomarker: A cohort study of mild iodine deficient pregnant women in China. *Nutrients*. 2023;15:3555.

doi: 10.3390/nu15163555

42. Pan Z. *Research on the Relationship between the Contents of 17 Trace Elements in Children's Serum and Thyroid Function and the Risk of Thyroid Diseases*. China: Tianjin Medical University; 2021.

doi: 10.27366/dcnki.Gtyku.2021.001429

43. Cui T, Wang W, Chen W, *et al*. Serum iodine is correlated with iodine intake and thyroid function in school-age children from a sufficient-to-excessive iodine intake area. *J Nutr*. 2019;149:1012-1018.

doi: 10.1093/jn/nxy325

44. Gao S, Diao X, Liu D. Investigation of UIC and SIC content in children. *J Jining Med Coll*. 2024;27:58.

doi: 10.3969/j.issn.1000-9760.2004.01.027

45. Xu X, Zhou H, He J, *et al*. Investigation of UIC and SIC content in urban children in Beijing. *Microelements Health Res*. 1999;(1):63-64.

46. Shu G. *Saliva Iodine Applied to Children's Iodine Nutrition Evaluation Research*. China: Tianjin Medical University, 2020.

doi: 10.27366/,dcnki.Gtyku.2020.000617

47. Zheng Y. *Research on the Correlation between Iodine Nutritional Status and the Risk of Papillary Thyroid Carcinoma*. China: Jilin University; 2023.

doi: 10.27162/,dcnki.Gjlin.2023.004058

48. Geng J, Wu M, Kong X, Yuan J. Study on the correlation between SIC level and thyroid tumors. *Chin J Modern Gen Surg*. 2017;20:770-772.

doi: 10.3969/j.issn.1009-9905.2017.10.005

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

Sex-based differences in streptozotocin-induced type 2 diabetes rat models

Ugljesa Malicevic[1,2,3], Jacob Smith[1], Devendra K. Agrawal[1], and Vikrant Rai[1]*

[1]Department of Translational Research, College of Osteopathic Medicine of the Pacific, Western University of Health Sciences, Pomona, California, United States of America

[2]Centre for Biomedical Research, Faculty of Medicine, University of Banja Luka, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

[3]Department of Pathophysiology, Pharmacology, Toxicology, and Clinical Pharmacology, Faculty of Medicine, University of Banja Luka, Banja Luka, Republic of Srpska, Bosnia and Herzegovina

Abstract

Background: Diabetic foot ulcers (DFUs) are a severe complication of diabetes mellitus, leading to chronic infections, amputations, and increased mortality despite existing treatments. More effective therapeutics are urgently needed, and animal models provide a critical platform for investigating the molecular mechanisms underlying DFUs. Streptozotocin (STZ)—administered at a low dose to induce type 2 diabetes (T2D) and at a high dose to induce type 1 diabetes—is commonly used in mice and rats in DFU research. **Aim:** The objective of this study is to highlight the importance of including male and female rats in DFU research. **Methods:** Both male and female Sprague–Dawley rats, 6–8 weeks old, were fed a high-fat diet for 9 weeks. STZ (25 mg/kg intraperitoneally) was administered weekly from the 6th week to induce T2D. **Results:** Female rats required a higher dose of STZ compared to male rats. The induction of T2D correlated positively with weight gain, which was greater in males than in females. **Conclusion:** The findings suggest that in addition to gender and weight, other factors may influence the induction of T2D in rats. Most studies in the literature do not use both sexes in DFU research. The distinct responses to STZ and weight gain observed emphasize the need to include both sexes and employ a more detailed approach in preclinical studies to enhance the understanding of DFU wound healing and translate the findings into potential treatments. **Relevance for patients:** The multifactorial effect on diabetes development, which differs in males and females, suggests the need to consider etiological, physiological, and demographic factors, such as body weight, gender, age, and body mass index, in the prevention and treatment of diabetes. This will also help in planning the individualized treatment for DFUs.

Keywords: Diabetic foot ulcer; Animal model; Type 2 diabetes; Streptozotocin; Gender

***Corresponding author:**
Vikrant Rai
(vrai@westernu.edu)

Citation: Malicevic U, Smith J, Agrawal DK, Rai V. Sex-based differences in streptozotocin-induced type 2 diabetes rat models. *J Clin Transl Res.* 2025;11(6):20-28. doi: 10.36922/JCTR025170020

Received: April 26, 2025

Revised: August 16, 2025

Accepted: October 9, 2025

Published online: October 28, 2025

1. Introduction

Diabetic foot ulcers (DFUs) represent one of the most severe complications of diabetes mellitus, often resulting in chronic infections, amputations, and significantly

increased mortality rates.[1] DFUs are the leading cause of hospitalization among all diabetes complications,[2] placing a substantial burden on healthcare systems due to their complex treatments and heightened risk of life-threatening outcomes. DFUs affect approximately 19 million people worldwide, including around 1.6 million in the United States, significantly burdening healthcare costs, with annual estimated treatment expenses ranging from USD 9 to USD 13 billion.[3] Despite current treatment, which includes wound debridement, antibiotics, growth factors, hyperbaric oxygen therapy, negative pressure wound therapy, topical oxygen, and skin grafts, the amputation rate of the lower extremity ranges from 3.34% to 42.83%.[4] Furthermore, the recurrence rate in DFU after treatment is 40% within 1 year, 60% in 3 years, and 68% within 6 years.[5] This suggests the need for improved treatment strategies. Moreover, chronic wounds, such as DFUs, present a major challenge for clinicians, complicating the understanding of their pathophysiology, hindering effective interventions, and highlighting the urgent need for targeted research and innovation in this field. Animal models are useful for studying the underlying molecular mechanisms and for investigating the effects of drugs and small compounds, which are tested for targeted therapy, in improving DFU healing.[6-8] Investigating underlying molecular mechanisms, therapeutic targets, and potential drugs in animal models, particularly in rodents and pigs, raises interest in translating these insights to enhance human patient care. While these models have offered valuable insights into the mechanisms and therapeutic approaches for wound healing, it is important to note that animal healing processes differ significantly from those in humans, especially given the chronic nature of human wounds that are often absent in animal models.[9]

Animal models used in preclinical studies of DFU healing, including mice, rats, pigs, guinea pigs, dogs, rabbits, zebrafish, or sheep, have their advantages and limitations.[8,10] The widely used Zucker diabetic Sprague–Dawley (ZDSD) rat model, along with the ob/ob and db/db mouse models, carries genetic defects in the leptin signaling pathway, leading to obesity and insulin resistance (IR).[8] However, these mutations do not fully represent the typical pathophysiology of human type 2 diabetes (T2D), making these models less ideal for capturing the complexity of T2D development in humans. To mimic diabetic ulcers in humans, inducing diabetes in mice or rats is a common practice in DFU research. Streptozotocin (STZ) is commonly used to induce diabetes in rodent models,[11] and research suggests that a high dose of STZ (50 mg/kg body weight) induces type 1 diabetes (T1D) while a low dose (25 mg/kg body weight) induces T2D.[8] Female rodents tend to be more resistant to low-dose STZ-induced T2D compared to males.[7] Males tend to develop severe metabolic disturbances, including IR, hyperglycemia, and obesity, while females often experience a delayed onset of hyperglycemia and milder IR. This difference is particularly evident in younger male rats, where hyperglycemia occurs earlier and more frequently than in females, likely due to the protective effects of estrogen.[12] This hormone has been shown to enhance insulin sensitivity, reduce inflammation, and offer protection against metabolic syndrome,[13] which explains why female ZDSD rats tend to display milder diabetes symptoms compared to males. Along with the role of estrogen in mice, the role of estrogen in insulin sensitivity in male and female rats has also been discussed,[14-16] suggesting its role in rats. Male rats also tend to gain more weight and accumulate more visceral fat,[17] a major contributor to IR development. Since obesity is a key driver of T2D in ZDSD rats, males typically display more extreme obesity-induced metabolic impairments than females. This suggests that induction of diabetes in rodents may be affected by age, hormones, metabolism, and weight.

T2D in rodents is induced with low-dose STZ, inducing β-cell dysfunction after the rodents are fed a high-fat diet (HFD) for at least 8 weeks, while T1D can be induced without HFD.[18,19] After STZ injection, severe T1D develops in nearly 50% of mice around 3 weeks after injection,[20] while the rate of diabetes induction in rats is 76% and 89% for T1D and T2D, respectively.[21] A study by Zhang *et al.*[22] reported that repeated injections of low-dose STZ are a better way to induce diabetes for developing a stable animal model of T2D. Consumption of HFD is associated with weight gain and adiposity, and it may depend on the rate of consumption and compositional changes.[23] Most studies in DFU research using rat models have either not mentioned the sex of the rats being used, used only female rats, induced T1D, or induced T2D in male rats after a single STZ injection (dose varying between 30–35 mg/kg). In the case of more than one STZ injection, the studies mentioned multiple injections but not the exact numbers or timing. Furthermore, the association or correlation between weight gain after HFD, the number of STZ injections, and the development of T2D in male versus female rats has not been mentioned.[24-31]

In this observational study, we used Sprague–Dawley (SD) rats to induce T2D for DFU healing research. We observed that after 8 weeks of HFD, male and female rats did not develop diabetes at the same rate, and female rats required more STZ injections over a longer period to develop T2D. Furthermore, the weight gain in the rats also differed between males and females. The findings of this study underscore the importance of including both sexes

in DFU healing research and provide valuable insight into sex-specific responses to STZ during diabetes induction.

2. Materials and methods

2.1. Animal model

SD rats (n = 42; 21 males and 21 females), aged 6–8 weeks and weighing approximately 175 g, were obtained from Charles River Laboratories (United States of America). The animals were housed in the Animal Resource Facility at the Western University of Health Sciences, Pomona, California, under standardized conditions, including a constant temperature of 22°C and a 12-h light/dark cycle, following the facility's standard operating procedures. The research involving these rats was approved by the Institutional Animal Care and Use Committee (IACUC) at Western University of Health Sciences with protocol approval # R24IACUC013. The rats were randomly divided into two groups: a control non-diabetic group (14 rats) and a diabetic group (28 rats). The control group was fed a normal diet (20% protein, 70% carbohydrate, 10% fat; Research Diet Inc., United States of America) and water *ad libitum*. The diabetic group rats received an HFD (35% carbohydrate, 20% protein, 45% fat; 5.7 kcal/g total; Research Diet Inc., United States of America), and water was available *ad libitum*. The rats were weighed at three time points: at the start of the experiment, after 6 weeks of feeding (before diabetes induction), and at the end of the experiment (week nine, following the development of diabetes). The blood glucose levels in the rat tail vein were also measured initially, before inducing diabetes, 2 weeks after the first STZ injection, and at the end of the experiment. All of the diabetic rats were utilized for further study.

We performed power analysis considering the primary outcome of animal weight and the secondary outcome of blood glucose levels and chose the sample size that would suffice for all experiments. Based on the power analysis using G*power software (Version 3.0.10, University of Kiel, Germany) with a value of 0.05, the sample size required to have at least 95% power to detect a significant change is 10 in each group. Based on our previous experience with mortality, we used a minimum of 14 rats in each group.

2.2. Induction of T2D

After 6 weeks of HFD, both male and female rats in the diabetic group were injected intraperitoneally (IP) with a low dose of STZ (25 mg/kg, dissolved in 0.1 M sodium citrate buffer, pH 4.4; Sigma-Aldrich, United States of America) to induce T2D. Our previous experience using IACUC protocol R22IACUC020 showed that SD rats did not develop T2D after a single STZ injection; a second injection was required 2 weeks after the first injection (unpublished own data). Based on this experience, we administered a second STZ injection in the current study to induce T2D in SD rats (n = 28, 14 males, and 14 females). The second dose was administered 1 week after the first STZ injection. Rats in the control group were injected with vehicle citrate buffer (0.25 mL/kg) on the same day. Hyperglycemia developed consistently in diabetic rats around 2-week post-induction. The blood glucose levels were monitored in both groups using tail vein blood samples and measured with an AlphaTrak glucometer (Zoetis Inc., United States of America). Rats with a blood glucose level exceeding 250 mg/dL 2 weeks after the first injection in males and after the third injection in females were considered diabetic.

2.3. Statistical analysis

Data were presented as mean ± standard deviation. The statistical significance was analyzed using a Student's t-test to compare two groups. A $p<0.05$ was considered statistically significant.

3. Results

3.1. HFD-induced greater weight gain in males than in females but did not induce diabetes

The initial weight for male and female rats was 195.06 ± 13.48 g and 152.21 ± 25.21 g, respectively, while the random blood glucose levels were 118.9 ± 2.29 mg/dL and 116.48 ± 5.15 mg/dL (Figure 1A). After 6 weeks of HFD, male SD rats gained more weight compared to female rats. The average weight was 303.93 ± 27.68 g in females and 492.28 ± 35.12 g in males, while the average blood glucose was 144.57 ± 16.33 mg/dL in females and 137.93 ± 19.66 mg/dL in males. The weight gain in female and male rats was statistically significant ($p<0.0001$ for both weight and blood glucose in female rats and $p<0.0001$ for weight and $p<0.01$ for blood glucose in males, respectively) (Figure 1B). Though the change in weight and blood glucose was significant, neither male nor female rats developed diabetes (Figure 1).

3.2. Males developed diabetes after two STZ injections but not females

All males developed diabetes after two injections of STZ (25 mg/kg), while the majority of the females did not develop diabetes (3 out of 14 females developed diabetes) (Figure 2A). The 11 females were injected with a third injection 1 week after the second STZ injection. The female rats developed diabetes after the third injection (Figure 2B). It was also evident from the results that male rats weighed more than females (Figure 2A and B), and

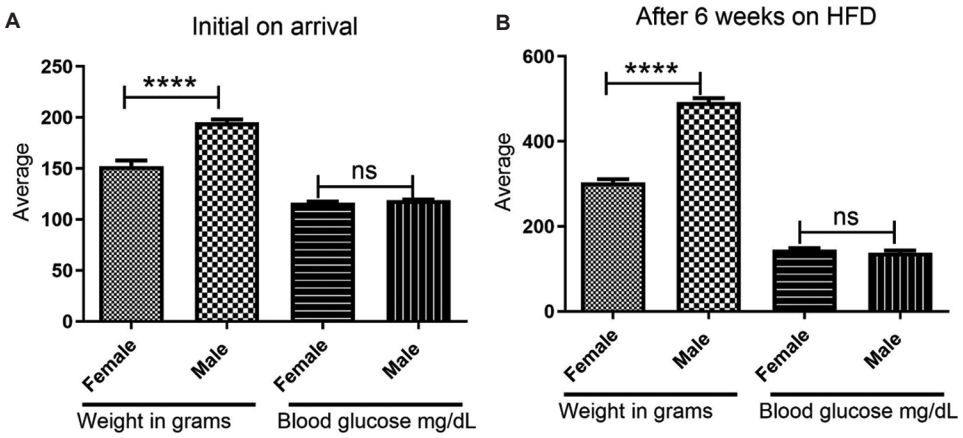


Figure 1. Weight and blood glucose at the start of the experiment and after 6 weeks of a high-fat diet (HFD). The data are presented as mean ± standard deviation, with $p<0.05$ indicating statistical significance. (A) Weight and blood glucose at the start of the experiment. (B) Weight and blood glucose after 6 weeks of HFD.
Notes: ****$p<0.0001$.
Abbreviation: ns: non-significant.

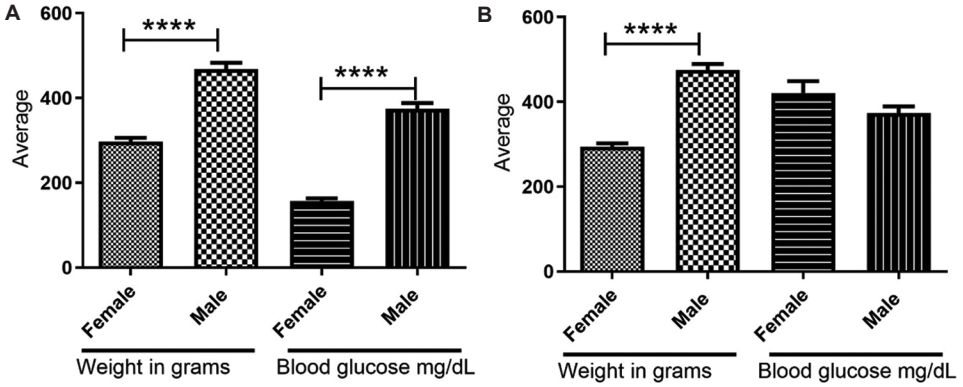


Figure 2. Streptozotocin (STZ)-induced type 2 diabetes in male and female rats. The data are presented as mean ± standard deviation, with $p<0.05$ indicating statistical significance. (A) Weight and blood glucose after two injections of STZ. (B) Weight and blood glucose before inducing ulcers (two STZ injections in males and three STZ injections in females).
Note: ****$p<0.0001$.

after the third injection, blood glucose levels were higher in females compared to males, although not statistically significant. The average blood glucose levels in males were 374.34 ± 48.6 mg/dL, and the average weight was 468.04 ± 51.85 g. The average blood glucose levels in females were 348 ± 51.17 mg/dL in those developing diabetes (n = 3) and 156.63 ± 21.92 mg/dL in those not developing diabetes (n = 11). The average weight of female rats was 296 ± 28.96 g.

3.3. Blood glucose increased after STZ injections and was positively correlated with weight gain

In both male and female rats, increases in random blood glucose levels were positively associated with weight gain and HFD. In male rats, strong positive correlations were observed between initial weight and blood glucose levels (Figure 3A), after 6 weeks of HFD (Figure 3B), and at the time of wound induction (Figure 3C) (r^2 = 0.79, r^2 = 0.85, and r^2 = 0.95, respectively). In female rats, the correlation was weaker at the start of the experiment (r^2 = 0.40; Figure 3A) but strengthened after 6 weeks of HFD (r^2 = 0.81; Figure 3B) and at the time of wound induction (r^2 = 0.96; Figure 3C). These results suggest that weight gain is positively correlated with an increase in blood glucose. Since weight gain with HFD is associated with an increase in blood glucose, it is important to investigate why females require one extra dose of STZ.

4. Discussion

The results of this observational study suggest that STZ is effective in inducing T2D in rats, but female rats require an additional dose of STZ compared to males. In addition, the

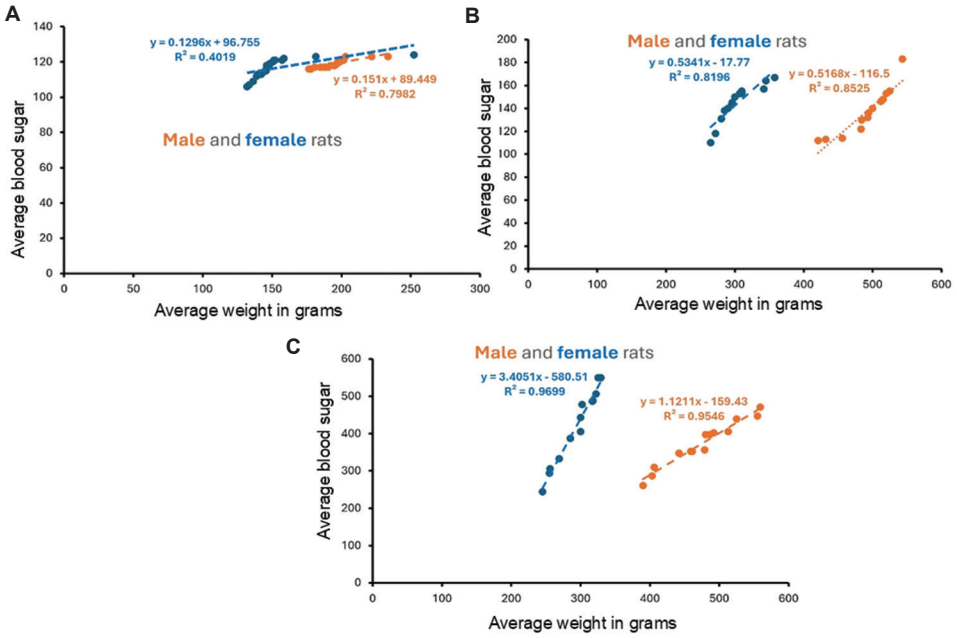


Figure 3. Correlation analysis between weight and random blood glucose levels in the female and male rats fed a high-fat diet. (A) At the start of the experiment (on arrival). (B) After 6 weeks of feeding with a high-fat diet. (C) Before inducing ulcers.

increase in blood glucose after STZ injection is positively correlated with weight gain after HFD administration in both males and females. Another important finding was that the increase in weight was greater in male rats compared to female rats after 6 weeks of HFD (weight gain of 200% in females and 250% in males) and remained stable until the end of the study (200% in females and 248% in males). The weight gain in both male and female rats may be due to the continuous HFD feeding. The increase in weight gain and a positive correlation in both male and female rats suggest that factors other than weight may affect the induction of diabetes after STZ injection, and these factors may be underlying causes of the need for increased STZ in female rats.

The first reason may be the differential STZ metabolism in male and female rats because sex plays a critical role in diabetes development. A recent study suggests that HFD and low-dose STZ induce hyperinsulinemia and IR in male but not female C57BL/6J mice,[32] where 30 mg/kg STZ was given once daily for 3 consecutive days. Another study reported different responses in IR and glucose metabolism in C57BL/6J female and male mice treated with low-dose STZ (40 mg/kg for 5 days).[33] Sex differences have also been monitored for fasting blood glucose and glucose tolerance between male and female outbred mice after 40 mg/kg STZ injection for 8 days.[34] Furman[20] concluded that C57BL/6, CD-1, or Balb/cJ mice develop T1D with 40 mg/kg STZ, intraperitoneal, injected for 5 days. A high dose of STZ (200 mg/kg) induces T1D within 48 h but has multiple toxic effects. Low-dose STZ is associated with milder toxicity. SD or Wistar male rats develop T1D after 65 mg/kg STZ injections. These studies primarily employed mouse models, focusing on T1D development in mice. These effects may be due to the sensitivity to STZ. Iwase *et al.*[35] reported sex differences in susceptibility to neonatal STZ treatment in inducing diabetes and hypertension in spontaneously hypertensive rats.

Second, the resistance of the female rats to the effects of low-dose STZ may be the reason for the increased number of injections we used in this study. Studies suggest that female rodents (mice and rats) are resistant to low doses of STZ, and this may be overcome by increasing the dose of STZ.[19,20] This notion is supported by the fact that most of the studies cited above used a higher dose of STZ for T1D and multiple injections of low-dose STZ to induce T2D in rodents. In this study, we administered only one additional dose of STZ (25 mg/kg) to induce T2D in rats fed with HFD. STZ sensitivity and resistance may also vary by strain: Wistar and SD rats are more sensitive, whereas Wistar–Kyoto rats exhibit lower sensitivity.[19] Among mice, DBA/2 is the most sensitive, followed by C57BL6, and Balb/cJ are resistant to STZ.[20] This suggests that it is not only the sex of the animal but also the strain that should be carefully considered when selecting the animals for DFU research.

Next, induction of T2D in males with a lower total dose of STZ may also be due to the increased toxicity of STZ to the pancreatic islets of males compared to females.[19,20,33]

This is supported by the fact that lower levels of beta-cell apoptosis occur in female mice compared to male mice after STZ treatment, with the fact that alpha-to-beta-cell transdifferentiation rates are higher in healthy female mice than in male mice, while control male mice have larger pancreatic islets than females.[36] The need for an increased dose or multiple injections of low-dose STZ in female rodents to induce T2D may also be due to the greater effect of STZ on glucose metabolism in male diabetic mice than in female diabetic mice.[34] Furthermore, the response in IR and glucose metabolism differs in male and female mice fed with HFD and treated with low-dose STZ.[32]

Another important aspect may be the effect of sex hormones on the metabolism of STZ and glucose. Female hormones may regulate STZ and glucose metabolism regardless of the genetic background.[33,34] This can be due to estrogen. Estrogen's role in glucose metabolism after STZ treatment is supported by the findings of restoration of nitrergic relaxation and estrogen receptor levels in STZ-induced diabetic female mice after 17β-estradiol supplementation.[33] The findings that 17β-estradiol supplementation suppresses gastric inflammatory and apoptotic stress responses in STZ-induced diabetic female mice[37] further support the role of estrogen. Another study further supported its role by reporting higher fasting blood glucose and hemoglobin A1c levels in ovariectomized female mice.[34] Furthermore, estrogen regulates various metabolic processes in tissues, including adipose, liver, muscle, and brain tissues, and dysregulation of estrogen signaling results in metabolic disorders, such as diabetes.[38] This again suggests that estrogen may play a role in the induction of diabetes. In addition, androgens may also sensitize the male mice to glucocorticoid-induced IR and fat accumulation.[39] It should also be noted that STZ-induced T2D rats, as compared to normoglycemic rats, showed comparatively higher levels of variation among biochemical, toxicological, and hematological parameters, suggesting alterations in normal glucose metabolism/homeostasis.[40] Although Rehman *et al.*[40] conducted the study using only male rats, the effects in female rats should be investigated. In the current observational study, induction of diabetes was positively correlated with weight gain in both male and female rats. The positive association in male and female rats may be supported by the secretion of androgens in males[41] and, to a lesser extent, in females.[42] We observed that both male and female rats gained weight on an HFD. However, there was no further weight gain after STZ injection. This aligned with previous observations that rodents experienced weight loss after STZ injections due to the lack of insulin, resulting in impaired glucose uptake and utilization in adipose tissue.[43] The need for a low dose of STZ in males to induce diabetes is supported by the notion that a higher level of body fat can exacerbate the effects of STZ-induced diabetes due to altered insulin sensitivity and changes in adipose tissue function, because adiposity significantly affects the STZ metabolism.[44,45] In our study, male rats gained more weight than female rats.

Altogether, the findings of this study suggest that female rats need a higher amount of STZ to induce T2D. Most studies related to DFU research include either male or female rodents. However, based on our results, both male and female rodents should be used while inducing T2D and working with diabetic ulcers. This is important because the healing pattern may differ due to the differential weight gain (adiposity) in male and female rats before the induction of T2D. One of the important findings we observed was that the rats (both male and female; one female and three males in the control group) were associated with mortality when anesthetized with 90 mg/kg ketamine and 10 mg/kg xylazine injected IP before the induction of the wound. We switched the anesthesia to isoflurane only and induced wound on the dorsal surface, and this resulted in no mortality. The mortality in rats injected with ketamine + xylazine may be due to the inhibition of the CYP-mediated ketamine metabolism by xylazine, resulting in a higher ketamine concentration in the body, as reported in humans, dogs, and horses.[46] However, this association and its relationship with adiposity in rodents warrant investigations. Of note, these rats died after inducing the full-thickness dermal wound and not during induction and maintenance of diabetes; thus, the mortality has not affected the statistical analysis of this study.

5. Conclusion

The results of this study suggest that the development of STZ-induced T2D in SD rats on an HFD is positively correlated with weight gain. The female rats need a higher amount of STZ to induce diabetes. These results underscore the importance of including both male and female rats in DFU research. This becomes increasingly important in translational research because the weight, amount of STZ, level of hyperglycemia, and the gender of the animal may influence the effect of the testing article (drug or small molecules).

There are also limitations in this study. Differences in hepatic metabolism, hormonal regulation, and islet cell structure are proposed as potential mechanisms underlying sex differences in STZ sensitivity. However, these remain hypothetical and lack experimental validation. A study evaluating the differences in hepatic metabolism, hormonal regulation, and islet cell structure can be conducted to support this hypothesis with observational

results. In addition, estrogen levels in female rats fluctuate significantly across the estrous cycle, which is known to influence insulin sensitivity and glucose metabolism.[47] Thus, future studies should monitor or control the estrous cycle to investigate the effects of estrogen on assessing STZ sensitivity. Future studies should also incorporate functional analyses focusing on sex hormones and the pharmacokinetics of STZ, as well as experimental designs that include monitoring of the estrous cycle. An additional issue is the difference in the dosage of STZ, which must be taken into consideration. For this study, we used the STZ from the same lot.

Acknowledgments

None.

Funding

The research was supported by a grant (25270C) from GMS Capitals LLC.

Conflict of interest

Vikrant Rai is an Editorial Board Member of this journal but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. Separately, other authors declared that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Author contributions

Conceptualization: Vikrant Rai
Formal analysis: Vikrant Rai
Investigation: Ugljesa Malicevic, Jacob Smith
Methodology: Ugljesa Malicevic, Jacob Smith, Vikrant Rai
Writing–original draft: Ugljesa Malicevic, Jacob Smith, Vikrant Rai
Writing–review & editing: All authors

Ethics approval and consent to participate

The study was reviewed and approved by the IACUC at Western University of Health Sciences with protocol approval # R24IACUC013.

Consent for publication

Not applicable.

Availability of data

All data related to this manuscript have been included in the manuscript. Further details may be obtained on request from the corresponding author.

References

1. Wang X, Yuan CX, Xu B, Yu Z. Diabetic foot ulcers: Classification, risk factors and management. *World J Diabetes* 2022;13:1049-1065.

   doi: 10.4239/wjd.v13.i12.1049

2. Volmer-Thole M, Lobmann R. Neuropathy and diabetic foot syndrome. *Int J Mol Sci* 2016;17:917.

   doi: 10.3390/ijms17060917

3. Armstrong DG, Tan TW, Boulton AJM, Bus SA. Diabetic foot ulcers: A review. *JAMA* 2023;330:62-75.

   doi: 10.1001/jama.2023.10578

4. Rai V, Le H, Agrawal DK. Novel mediators regulating angiogenesis in diabetic foot ulcer healing. *Can J Physiol Pharmacol.* 2023;101:488-501.

   doi: 10.1139/cjpp-2023-0193

5. Guo Q, Ying G, Jing O, *et al*. Influencing factors for the recurrence of diabetic foot ulcers: A meta-analysis. *Int Wound J*. 2023;20:1762-1775.

   doi: 10.1111/iwj.14017

6. Takahashi Y, Fukusato T. Animal models of liver diseases. In: *Animal Models for the Study of Human Disease*. Netherlands: Elsevier; 2017. p. 313-339.

   doi: 10.1016/B978-0-12-809468-6.00013-9

7. Rai V, Agrawal DK. Male or female sex: Considerations and translational aspects in diabetic foot ulcer research using rodent models. *Mol Cell Biochem.* 2023;478:1835-1845.

   doi: 10.1007/s11010-022-04642-7

8. Rai V, Moellmer R, Agrawal DK. Clinically relevant experimental rodent models of diabetic foot ulcer. *Mol Cell Biochem.* 2022;477:1239-1247.

   doi: 10.1007/s11010-022-04372-w

9. Tan MLL, Chin JS, Madden L, Becker DL. Challenges faced in developing an ideal chronic wound model. *Expert Opin Drug Discov.* 2023;18:99-114.

   doi: 10.1080/17460441.2023.2158809

10. Sanapalli BKR, Yele V, Singh MK, Thaggikuppe Krishnamurthy P, Karri V. Preclinical models of diabetic wound healing: A critical review. *Biomed Pharmacother.* 2021;142:111946.

    doi: 10.1016/j.biopha.2021.111946

11. Furman BL. Streptozotocin-induced diabetic models in mice and rats. *Curr Protoc Pharmacol.* 2015;70:5.47.41-5.47.20.

    doi: 10.1002/0471141755.ph0547s70

12. Diaz A, Lopez-Grueso R, Gambini J, *et al*. Sex differences in age-associated type 2 diabetes in rats-role of estrogens and oxidative stress. *Oxid Med Cell Longev.* 2019;2019:6734836.

doi: 10.1155/2019/6734836

13. Yan H, Yang W, Zhou F, *et al*. Estrogen improves insulin sensitivity and suppresses gluconeogenesis via the transcription factor Foxo1. *Diabetes.* 2019;68:291-304.

    doi: 10.2337/db18-0638

14. Pratchayasakul W, Chattipakorn N, Chattipakorn SC. Estrogen restores brain insulin sensitivity in ovariectomized non-obese rats, but not in ovariectomized obese rats. *Metabolism.* 2014;63:851-859.

    doi: 10.1016/j.metabol.2014.03.009

15. Kawakami M, Yokota-Nakagi N, Uji M, *et al*. Estrogen replacement enhances insulin-induced AS160 activation and improves insulin sensitivity in ovariectomized rats. *Am J Physiol Endocrinol Metab.* 2018;315:E1296-E1304.

    doi: 10.1152/ajpendo.00131.2018

16. Pratchayasakul W, Chattipakorn N, Chattipakorn SC. Effects of estrogen in preventing neuronal insulin resistance in hippocampus of obese rats are different between genders. *Life Sci.* 2011;89:702-707.

    doi: 10.1016/j.lfs.2011.08.011

17. Maric I, Krieger JP, van der Velden P, *et al*. Sex and species differences in the development of diet-induced obesity and metabolic disturbances in rodents. *Front Nutr* 2022;9:828522.

    doi: 10.3389/fnut.2022.828522

18. Gheibi S, Kashfi K, Ghasemi A. A practical guide for induction of type-2 diabetes in rat: Incorporating a high-fat diet and streptozotocin. *Biomed Pharmacother.* 2017;95:605-613.

    doi: 10.1016/j.biopha.2017.08.098

19. Ghasemi A, Jeddi S. Streptozotocin as a tool for induction of rat models of diabetes: A practical guide. *EXCLI J.* 2023;22:274-294.

    doi: 10.17179/excli2022-5720

20. Furman BL. Streptozotocin-induced diabetic models in mice and rats. *Curr Protoc.* 2021;1:e78.

    doi: 10.1002/cpz1.78

21. Akinlade OM, Owoyele BV, Soladoye AO. Streptozotocin-induced type 1 and 2 diabetes in rodents: A model for studying diabetic cardiac autonomic neuropathy. *Afr Health Sci.* 2021;21:719-727.

    doi: 10.4314/ahs.v21i2.30

22. Zhang M, Lv XY, Li J, Xu ZG, Chen L. The characterization of high-fat diet and multiple low-dose streptozotocin induced type 2 diabetes rat model. *Exp Diabetes Res.* 2008;2008:704045.

    doi: 10.1155/2008/704045

23. Melhorn SJ, Krause EG, Scott KA, *et al*. Acute exposure to a high-fat diet alters meal patterns and body composition. *Physiol Behav.* 2010;99:33-39.

    doi: 10.1016/j.physbeh.2009.10.004

24. Uckun FM, Orhan C, Tuzcu M, *et al*. RJX improves wound healing in diabetic rats. *Front Endocrinol (Lausanne).* 2022;13:874291.

    doi: 10.3389/fendo.2022.874291

25. Wang J, Zhao X, Tian G, Liu X, Gui C, Xu L. Down-regulation of miR-138 alleviates inflammatory response and promotes wound healing in diabetic foot ulcer rats via activating PI3K/AKT pathway and hTERT. *Diabetes Metab Syndr Obes.* 2022;15:1153-1163.

    doi: 10.2147/DMSO.S359759

26. Qi L, Ahmadi AR, Huang J, *et al*. Major improvement in wound healing through pharmacologic mobilization of stem cells in severely diabetic rats. *Diabetes.* 2020;69:699-712.

    doi: 10.2337/db19-0907

27. Sun X, Wang X, Zhao Z, Chen J, Li C, Zhao G. Paeoniflorin accelerates foot wound healing in diabetic rats though activating the Nrf2 pathway. *Acta Histochem.* 2020;122:151649.

    doi: 10.1016/j.acthis.2020.151649

28. Cheng KY, Lin ZH, Cheng YP, *et al*. Wound healing in streptozotocin-induced diabetic rats using atmospheric-pressure argon plasma jet. *Sci Rep.* 2018;8:12214.

    doi: 10.1038/s41598-018-30597-1

29. Yang P, Pei Q, Yu T, *et al*. Compromised wound healing in ischemic type 2 diabetic rats. *PLoS One.* 2016;11:e0152068.

    doi: 10.1371/journal.pone.0152068

30. Karalashvili L, Chakhunashvili D, Kakabadze M, Paresishvili T, Kakabadze Z. Treatment of a non-healing oral wound in diabetic-induced rats. *Explor Immunol.* 2023;3:565-573.

31. Begum F, Manandhar S, Kumar G, *et al*. Dehydrozingerone promotes healing of diabetic foot ulcers: A molecular insight. *J Cell Commun Signal.* 2023;17:673-688.

    doi: 10.1007/s12079-022-00703-0

32. Racine KC, Iglesias-Carres L, Herring JA, *et al*. The high-fat diet and low-dose streptozotocin type-2 diabetes model induces hyperinsulinemia and insulin resistance in male but not female C57BL/6J mice. *Nutr Res* 2024;131:135-146.

    doi: 10.1016/j.nutres.2024.09.008

33. Kim B, Kim YY, Nguyen PTT, Nam H, Suh JG. Sex differences in glucose metabolism of streptozotocin-induced diabetes inbred mice (C57BL/6J). *Appl Biol Chem.* 2020;63:59.

    doi: 10.1186/s13765-020-00547-5

34. Kim B, Park ES, Lee JS, Suh JG. Outbred mice with streptozotocin-induced diabetes show sex differences in glucose metabolism. *Int J Mol Sci.* 2023;24:5210.

doi: 10.3390/ijms24065210

35. Iwase M, Kikuchi M, Nunoi K, *et al*. Sex differences in diabetes induced by neonatal streptozotocin treatment in spontaneously hypertensive rats. *Diabetes Res Clin Pract.* 1987;4:61-65.

doi: 10.1016/s0168-8227(87)80034-7

36. Tanday N, Coulter-Parkhill A, Moffett RC, *et al*. Sex-based impact of pancreatic islet stressors in GluCreERT2/Rosa26-eYFP mice. *J Endocrinol.* 2023;259:e230174.

doi: 10.1530/JOE-23-0174

37. Sprouse J, Sampath C, Gangula P. 17beta-estradiol suppresses gastric inflammatory and apoptotic stress responses and restores nNOS-mediated gastric emptying in streptozotocin (STZ)-induced diabetic female mice. *Antioxidants (Basel).* 2023;12:758.

doi: 10.3390/antiox12030758

38. Tao Z, Cheng Z. Hormonal regulation of metabolism-recent lessons learned from insulin and estrogen. *Clin Sci (Lond).* 2023;137:415-434.

doi: 10.1042/CS20210519

39. Gasparini SJ, Swarbrick MM, Kim S, *et al*. Androgens sensitise mice to glucocorticoid-induced insulin resistance and fat accumulation. *Diabetologia.* 2019;62:1463-1477.

doi: 10.1007/s00125-019-4887-0

40. Rehman HU, Ullah K, Rasool A, *et al*. Comparative impact of streptozotocin on altering normal glucose homeostasis in diabetic rats compared to normoglycemic rats. *Sci Rep.* 2023;13:7921.

doi: 10.1038/s41598-023-29445-8

41. Kaler LW, Neaves WB. The androgen status of aging male rats. *Endocrinology.* 1981;108:712-719.

doi: 10.1210/endo-108-2-712

42. Turgeon JL, Waring DW. Androgen modulation of luteinizing hormone secretion by female rat gonadotropes. *Endocrinology.* 1999;140:1767-1774.

doi: 10.1210/endo.140.4.6642

43. Carpene C, Stiliyanov Atanasov K, Les F, Mercader Barcelo J. Hyperglycemia and reduced adiposity of streptozotocin-induced diabetic mice are not alleviated by oral benzylamine supplementation. *World J Diabetes.* 2022;13:752-764.

doi: 10.4239/wjd.v13.i9.752

44. Luo L, Liu M. Adipose tissue in control of metabolism. *J Endocrinol.* 2016;231:R77-R99.

doi: 10.1530/JOE-16-0211

45. Singla P, Bardoloi A, Parkash AA. Metabolic effects of obesity: A review. *World J Diabetes.* 2010;1:76-88.

doi: 10.4239/wjd.v1.i3.76

46. Capponi L, Schmitz A, Thormann W, Theurillat R, Mevissen M. *In vitro* evaluation of differences in phase 1 metabolism of ketamine and other analgesics among humans, horses, and dogs. *Am J Vet Res.* 2009;70:777-786.

doi: 10.2460/ajvr.70.6.777

47. Rebolledo-Solleiro D, Fernandez-Guasti A. Influence of sex and estrous cycle on blood glucose levels, body weight gain, and depressive-like behavior in streptozotocin-induced diabetic rats. *Physiol Behav.* 2018;194:560-567.

doi: 10.1016/j.physbeh.2018.06.033

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

Grewia tenax fruits as a traditional remedy for iron deficiency anemia: A comparative clinical study with ferrous salt

Randa Alsadig Almahdi[1]* and Sami Ahmed Khalid[2]*

[1]Department of Pharmacy Practice, Faculty of Pharmacy, University of Science and Technology, Omdurman, Khartoum State, Sudan

[2]Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Science and Technology, Omdurman, Khartoum State, Sudan

Abstract

Background: Iron deficiency anemia (IDA) remains a significant global public health concern, particularly among vulnerable populations such as children and menstruating women. Although standard oral iron supplements are effective in replenishing iron stores, their use is often limited by gastrointestinal side effects that negatively impact adherence. *Grewia tenax* fruits, a traditional Sudanese remedy, are widely used for managing IDA; however, their clinical efficacy has not been rigorously evaluated. **Method:** This open-label study compared *G. tenax* chewable tablets with ferrous gluconate in 34 adult females (18–50 years) with confirmed IDA (hemoglobin <12 g/dL). Participants received either *G. tenax* (five tablets twice daily) or ferrous gluconate (one tablet twice daily) for a period of 4 weeks. Hematological parameters were monitored weekly, and iron profile markers were assessed at baseline and at week 4. **Results:** Ferrous gluconate produced significantly greater increases in hemoglobin and serum iron levels. Although *G. tenax* showed more modest hemoglobin improvements, it yielded higher post-treatment ferritin levels (7.82 vs. 7.43 μg/L) and greater reductions in total iron-binding capacity, suggesting enhanced iron storage and regulation. A transient rise in reticulocyte counts observed in the *G. tenax* group indicates early erythropoietic stimulation. Variability in individual response to *G. tenax* may be attributed to differences in absorption or metabolism, underscoring the need for personalized approaches. **Conclusion:** These findings challenge the usual iron dose, highlighting the unique pharmacological effects of *G. tenax*. Further research is warranted to explore its mechanisms, long-term benefits, and role as a culturally acceptable adjunct or alternative in the management of IDA. **Relevance for patients:** The study population reflects women who are disproportionately affected by IDA in Sudan, making the findings highly relevant to communities that traditionally rely on *G. tenax* as a culturally rooted remedy.

Keywords: Iron deficiency anemia; *Grewia tenax*; Iron bioavailability; Iron supplementation; Traditional practice in Sudan

***Corresponding authors:**
Randa Alsadig Almahdi (Randa.almahdi@ust.edu.sd); Sami Ahmed Khalid (prof.sami@ust.edu.sd)

Citation: Almahdi RA, Khalid SA. *Grewia tenax* fruits as a traditional remedy for iron deficiency anemia: A comparative clinical study with ferrous salt. *J Clin Transl Res*. 2025;11(6):29-38. doi: 10.36922/JCTR025360061

Received: September 3, 2025

Revised: September 25, 2025

Accepted: October 9, 2025

Published online: November 7, 2025

1. Introduction

Iron is an essential mineral required for erythropoiesis and oxidative metabolism.[1] Its absorption is complex and tightly regulated by multiple physiological factors,

making iron supplementation a delicate balance between correcting deficiency and avoiding overload. As such, considerable research has been devoted to understanding the mechanisms governing gastrointestinal iron absorption and systemic iron metabolism.[1,2]

Iron deficiency is the most common cause of anemia worldwide,[3] with iron deficiency anemia (IDA) affecting approximately one-quarter of the global population.[4] Vulnerable groups include young children, pregnant women, and menstruating females,[5] and high prevalence has been reported in both developing and developed countries.[6]

Accurate diagnosis of IDA relies on laboratory evaluation, including complete blood count, serum iron, serum ferritin, total iron-binding capacity (TIBC), and transferrin saturation.[7] The diagnostic hemoglobin thresholds are <12 g/dL for menstruating women, <11 g/dL for pregnant women, and <13 g/dL for men.[8]

Iron hemostasis is critical for effective iron supplementation.[8] Iron absorption occurs primarily in the duodenum and proximal jejunum,[9] where dietary iron is transported across enterocytes and exported to the plasma through ferroportin.[8] However, oral iron has limited bioavailability, as only a small proportion of iron absorbed by enterocytes is released into the circulation and incorporated into hemoglobin.[9,10]

The regulation of iron absorption is tightly controlled at both cellular and molecular levels.[9] At the cellular level, iron is sequestered within enterocytes by ferritin, transferrin, and transferrin receptors, while at the molecular level, hepcidin plays a central regulatory role.[11]

The first-line treatment for IDA typically involves oral iron supplementation to replenish depleted iron stores. A commonly recommended regimen consists of an oral iron salt providing 60 mg of elemental iron daily for 3 months.[5] However, this conventional approach has been criticized for gastrointestinal adverse effects, inconsistent efficacy, and, most seriously, toxicity.[12] These limitations have spurred exploration into safer, more tolerable alternatives, including herbal therapies.[12-14]

Moreover, supplementing iron poses challenges for healthcare providers, particularly in striking a balance between iron replenishment and preventing the elevation of the free labile iron pool.[15-17] When the labile iron pool exceeds the transport capacity of ferroportin, it participates in a cascade of reactions that generate oxidative stress, a major pathogenic factor in IDA.[18]

Recently, radical scavenging activity has emerged as supporting evidence for the protective role of antioxidants against oxidative stress.[19-22] Therefore, the incorporation of antioxidants as adjunctive therapy in the management of IDA has been proposed as a safer and more effective approach in both iron deficiency and iron overload.[23]

In Sudan, IDA is highly prevalent, especially among women of childbearing age (35.5%).[24] A widely used traditional remedy is the fruit of *Grewia tenax*, commonly consumed as juice or porridge by pregnant and lactating women.[25] *G. tenax* is native to western Sudan, particularly in the Darfur and Kordofan regions.[26] The plant is considered to have multifaceted economic value, serving as a food source and exhibiting various pharmacological properties, such as tissue healing and bone strengthening.[27]

Several animal studies have demonstrated the positive effects of *G. tenax* fruits on iron absorption and hemoglobin synthesis.[28,29] An important *in vitro* experiment using the everted rat gut sac technique investigated the intestinal transport of iron across the intestinal membrane, and subsequent studies in rats demonstrated that the fruits possess significant hematinic effects.[30]

The traditional use of natural products as therapeutic agents has increasingly been recognized as a valid basis for scientific investigation, potentially leading to the discovery of new treatment mechanisms or novel bioactive compounds.[31] In line with this, advances in the understanding of the pathogenesis and treatment of IDA[14] prompted previous researchers to screen *G. tenax* fruits for primary and secondary metabolites and to assess their antioxidant activity, correlating these findings with their ability to improve hemoglobin levels and reverse IDA.[32] Phytochemical analysis revealed 22 secondary metabolites in the aqueous extract of *G. tenax* fruits, most of which are known to exhibit potent antioxidant and pharmacological effects.[32] For example, flavonoids exhibit dual actions by chelating excess iron and serving as antioxidants.[16] Iron chelators have been shown to protect the body from iron overload and inhibit oxidative reactions.[12] These findings coincide with the paradoxical role of iron as both an essential and potentially detrimental micronutrient.[11]

Although the ability of *G. tenax* fruits to reverse IDA has been widely reported in the literature,[33-36] the relatively low iron content of the fruits (4.5 mg/100 g) suggests that their therapeutic effects may be attributed more to their bioactive constituents than to direct iron supplementation.[37]

Therefore, the current clinical trial aimed to evaluate the clinical effectiveness of *G. tenax* fruits in women with IDA, while exploring their possible mechanisms of action.

2. Materials and methods

2.1. Materials

2.1.1. Study population

A total of 44 eligible adult menstruating females, aged between 18 and 50 years and previously diagnosed with IDA with hemoglobin levels below 12 mg/dL, were voluntarily recruited for the study. Of the initial 44 participants in the cohort, 34 completed the clinical trial. The participants were randomly assigned to one of two groups: one group received *G. tenax* chewable tablets (n = 20), while the other group received ferrous gluconate tablets (n = 14) for a period of 1 month. The flow of participant recruitment and retention throughout the study stages is presented in Figure 1.

(a) Inclusion criteria
Females aged ≥18 years, previously diagnosed with IDA, and with baseline hemoglobin levels ≤ 12 g/dL were included.

(b) Exclusion criteria
Individuals with a history of gastrointestinal disorders that may interfere with oral iron absorption, known gastrointestinal bleeding, anemia of causes other than IDA, or pregnancy were excluded from the study.

2.1.2. Interventions

G. tenax was administered in the form of chewable tablets containing 500 mg of extracts derived from the fruit pericarp.

2.1.3. Tablet authentication

(a) *G. tenax* chewable tablets

The crude *G. tenax* fruits were imported from Sudan (Batch No. D101001), donated by Dellas Natural Products Factory, Amman, Jordan (H/Reg 139), and registered by the Jordanian Ministry of Health and healthcare registration No. 5N/2000.

The tablets passed quality control test No. 08/05/04 NA, conducted by the Division of Organic Materials Food Laboratory in Jordan. Analysis of inorganic materials and constituents was also conducted. The ascorbic acid content was 0.1 mg/g, quantified using liquid chromatography against a standard ascorbic acid solution. Test for uniformity of weight, tablet diameter, and friability complied with Dellas Natural Products quality requirements.

The mineral content per 100 g was as follows: Potassium 750.9 mg, calcium 332.3 mg, sodium 12.3 mg, iron 2.95 mg,

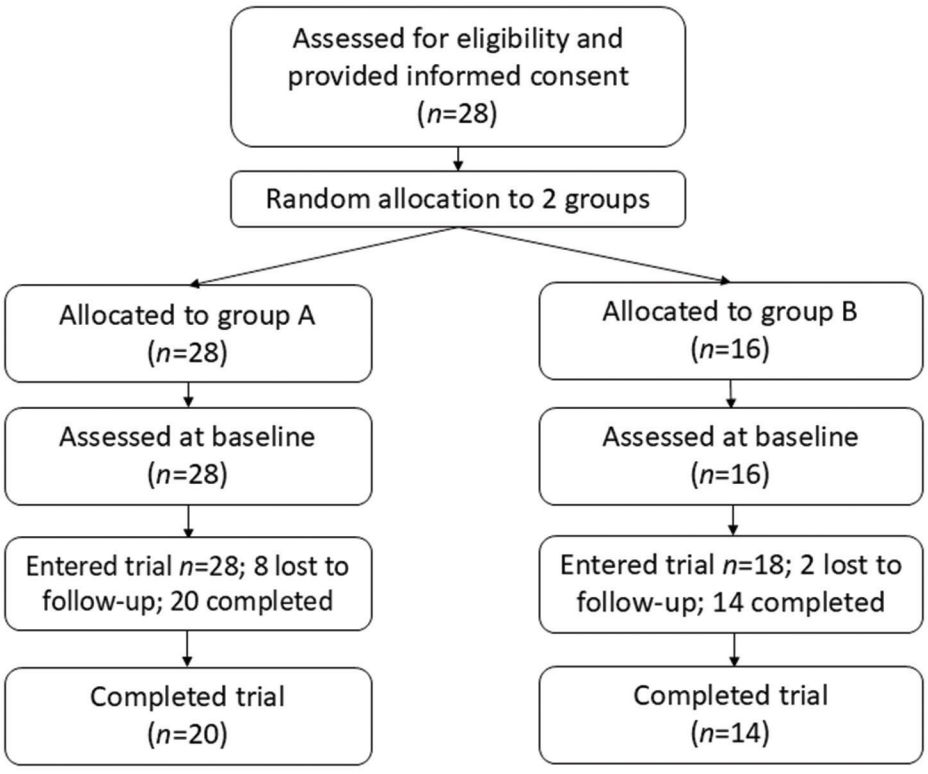


Figure 1. Flowchart of participant enrollment, allocation, follow-up, and completion

zinc 0.74 mg, manganese 0.49 mg, copper 0.39 mg, and chromium 0.25 mg.

Analysis of heavy metals and contaminants revealed that arsenic, mercury, cadmium, lead, and fluorine were each present at <0.01 mg/100 g, and no aflatoxins were detected.

(b) Ferrous gluconate tablets

The ferrous gluconate tablets were sugar-coated and manufactured by Chemical Industries Development Co., Giza, Arab Republic of Egypt (G.C.R. No. 19717). Each tablet contained 300 mg of ferrous gluconate, equivalent to 36 mg of elemental iron (Batch No. 103065T).

2.1.4. Sample size estimation

The study aimed to detect a 1–2 g/dL rise in hemoglobin over 1 month. The required sample size was calculated using Equation (1):

$$n = \frac{2(Z_{\frac{\alpha}{2}} + Z_{\beta})^2}{\Delta^2} \quad (1)$$

where:

- $\alpha = 0.05$ (two-sided), therefore $Z_{\alpha/2} = 1.96$
- Power = 80%, therefore $Z_{\beta} = 0.84$
- Δ = Expected mean difference in hemoglobin (1–2 g/dL in 1 month).

The calculated sample size was 35 participants per group. However, due to logistical constraints, the total sample size was limited to 34 participants, divided into two unequal groups.

2.1.5. Data analysis

Changes in hemoglobin levels over time were analyzed using SPSS version 25 (IBM, United States), employing analysis of variance (ANOVA). A $p<0.05$ was considered statistically significant.

2.2. Methods

This open-label clinical trial was conducted in adult female participants to compare the effects of *G. tenax* formulated as chewable tablets with ferrous gluconate tablets over 1 month.

Participants were randomly allocated into two groups:

- GT group ($n = 20$): Received *G. tenax* chewable tablets at a dose of five tablets twice daily, to be chewed after meals
- FG group ($n = 14$): Received ferrous gluconate tablets at a dose of one tablet twice daily for 4 weeks.

Each participant in the GT group received six packages of *G. tenax* chewable (Tonigrow®, Jordan), each containing 50 tablets, sufficient for the 4-week duration of the trial. Each participant in the FG group was supplied with 60 ferrous gluconate tablets for the same duration.

2.2.1. The outcome measures

The primary outcome measures included hematological parameters and iron profile indices:

- Hematological parameters (hemoglobin levels and reticulocyte count) were measured at baseline (week 0) and at weeks 1, 2, 3, and 4.
- Iron profile parameters (serum iron, serum ferritin, and TIBC) were measured at baseline (week 0) and at week 4.

Participants were instructed to maintain their regular diet, to discontinue the use of any tonics, nutritional supplements, or vitamin preparations, and to report any adverse reactions experienced during the 4-week duration.

2.2.2. Determination of the G. tenax daily dose

The daily dose of *G. tenax* chewable tablets was determined based on an ethnopharmacological survey carried out using a questionnaire administered to traditional users. The typical preparation method reported corresponded to approximately 5 g of *G. tenax* extract per day. This was standardized to 10 Tonigrow® chewable tablets, administered as 5 tablets twice daily after meals. This dose is equivalent to 5 g of *G. tenax* extract, containing approximately 0.9 mg of elemental iron.

2.2.3. Ethical clearance

Ethical approval for the trial was granted by the University of Science and Technology Ethics Committee (Reference No. UST/EC/2025/045).

All participants provided written informed consent upon enrollment and were informed of their right to withdraw from the trial at any time.

3. Results

3.1. Reported causes of IDA

Participants reported several causes of IDA, including heavy menstrual bleeding, poor nutrition, and unknown etiologies. The distribution of these reported causes is shown in Figure 2. Heavy menstrual bleeding accounted for approximately one-third of cases, while dietary factors and unknown causes comprised the remainder.

3.2. Hemoglobin response

Treatment response differed significantly between the two groups. In the *G. tenax* group, 10 out of 20 participants

(50%) exhibited a rise in hemoglobin levels, with changes ranging from 0.2 g/dL to 1.1 g/dL. In comparison, 11 out of 14 participants (79%) in the ferrous gluconate group responded to treatment, with a higher mean increase of approximately 1.2 g/dL (Figure 3).

Weekly mean hemoglobin levels increased from approximately 10.3 g/dL at baseline to 11.5 g/dL at week 4 in the GT group, and from 9.8 g/dL to 11.8 g/dL in the FG group (Figure 4).

A two-way repeated measures ANOVA (group × time) showed a significant between-group effect (F = 35.17, p<0.0001), a significant but small time effect (F = 2.46, p=0.048), and no significant group × time interaction (F = 1.65, p=0.166) (Table S1). Week-by-week analysis (Table S2) showed that significant between-group differences emerged at week 2 (p=0.009) and persisted through week 4 (p=0.005). These findings indicate that ferrous gluconate produced a more pronounced hemoglobin response compared to *G. tenax*, although the latter still resulted in modest, clinically meaningful improvements.

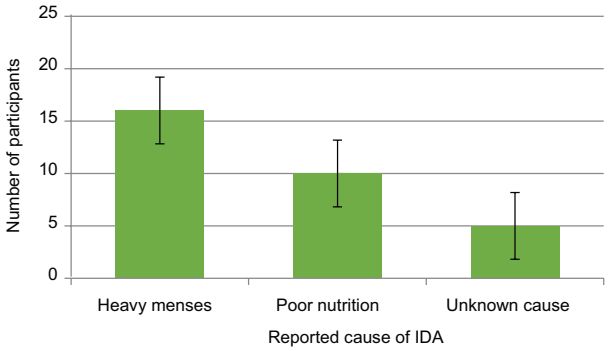


Figure 2. Distribution of reported causes of iron deficiency anemia

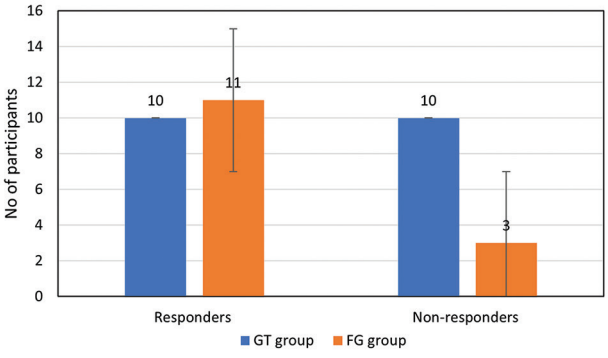


Figure 3. Comparisons of treatment responders and non-responders between the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups (Chi-square = 2.8464, p=0.09; not statistically significant)
Note: Responders were defined as participants who demonstrated an increase in hemoglobin levels; non-responders showed no increase.

3.3. Reticulocyte counts

Reticulocyte counts peaked early in the GT group and subsequently declined, whereas the FG group exhibited a gradual decrease over time (Figure 5 and Table S3). Within-group analysis using the Friedman test revealed no significant change over time in the GT arm (χ^2 = 7.56, p=0.109). A mixedeffects model showed no significant main effects or interaction between groups. Although *G. tenax* appeared to stimulate an early erythropoietic response, neither treatment resulted in statistically significant changes in reticulocyte counts over the 4-week period.

3.4. Serum iron

Serum iron levels increased in both groups; however, the responses diverged by week 4. Baseline values did not differ significantly (p=0.119), but by week 4, the FG group exhibited a significantly greater increase (F = 4.218,

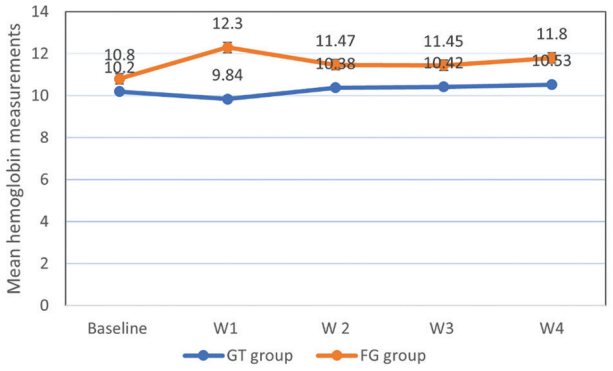


Figure 4. Mean hemoglobin levels in the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups among treatment responders (two-way repeated measures ANOVA)
Note: Responders were defined as participants who demonstrated an increase in hemoglobin levels.

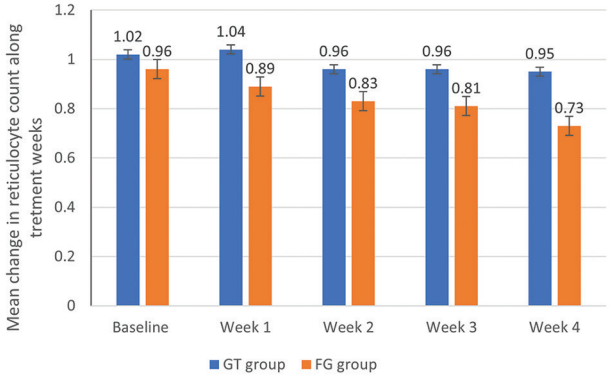


Figure 5. Changes in reticulocyte counts over the 4-week treatment period in the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups

p=0.048). As illustrated in Figure 6 (data corresponding to Table S4), ferrous gluconate more effectively replenished circulating iron, whereas *G. tenax* also produced a clinically meaningful increase of approximately 27 µg/dL in 65% of participants.

3.5. Serum ferritin

Serum ferritin levels increased significantly in both groups, indicating improved iron stores. Between-group differences were highly significant at both baseline and week 4 (F = 41.573 and F = 125.562, respectively; p<0.001 for both) (Table S5). The GT group exhibited a mean increase of 7.82 µg/L (45% of participants), whereas the FG group showed a mean increase of 7.43 µg/L (78% of participants). These results, depicted in Figure 7, suggest that *G. tenax* may enhance iron storage despite its relatively low elemental iron content.

3.6. TIBC

TIBC, an indicator of transferrin's iron-binding potential, typically ranges from 240 to 450 µg/dL. No significant between-group differences were observed at baseline (p=0.335) or at week 4 (p=0.137) (Table S6). However, the GT group showed a greater mean reduction from baseline (−37.2 µg/dL) compared with the FG group (−8.6 µg/dL), suggesting potentially greater transferrin saturation or regulatory feedback on iron absorption (Figure 8). Elevated TIBC is typically indicative of iron deficiency, whereas a decline suggests improving iron status or a physiological reduction in iron absorption.

3.7. Summary of key differences

Overall, ferrous gluconate produced greater and more rapid improvements in hemoglobin and serum iron levels. However, *G. tenax*, despite its low elemental iron content, still induced clinically meaningful hematological improvements in 50% of participants. It stimulated an early reticulocyte response, increased serum ferritin levels, and reduced TIBC more effectively than ferrous gluconate, as shown in Table S7. These findings highlight the potential of *G. tenax* to enhance iron bioavailability and stabilize iron storage through regulatory or antioxidant mechanisms, rather than through direct iron supplementation. Full numerical data are provided in Tables S1-S7.

4. Discussion

This open-label clinical trial was designed to compare the efficacy of *G. tenax* fruit extract in chewable tablet form with ferrous gluconate tablets for the treatment of IDA in menstruating females. The study included 34 participants, with 20 receiving *G. tenax* and 14 receiving ferrous gluconate for a duration of 4 weeks.

The response to iron supplementation varies considerably among individuals due to differences in baseline iron status, absorption efficiency, and other patient-specific factors.[5,38] This variability underscores the need for individualized treatment strategies in the management and monitoring of IDA.[39] Typically, hematological responses to iron supplementation begin to appear around day 14 following initiation of supplementation.[40]

In this study, hemoglobin levels were similar between the two groups at the baseline and remained comparable during the 1st week of intervention (p=0.154 and p=0.218, respectively). By weeks 2 (p=0.009) and 3 (p=0.019), the FG group showed significantly greater increases in hemoglobin levels compared with the GT group. This difference persisted at week 4 (p=0.005), indicating that ferrous gluconate produces faster and more robust improvements in hemoglobin levels. In contrast, *G. tenax* produced a modest increase that did not reach

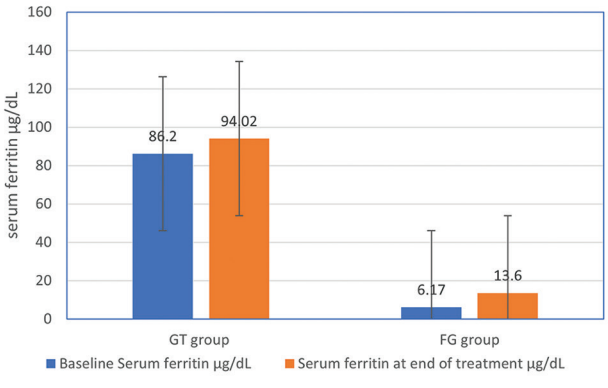


Figure 6. Changes in serum iron levels between baseline and week 4 in the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups (p=0.119 for GT; p=0.048 for FG)

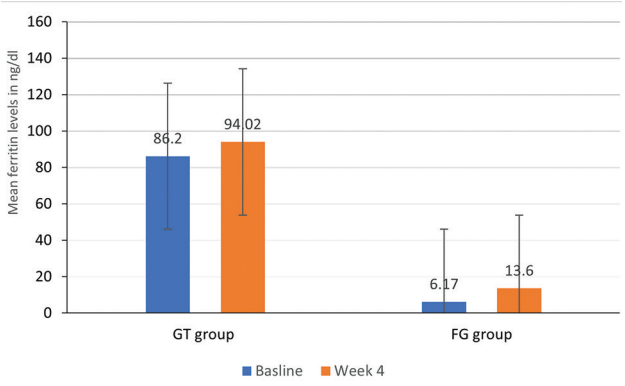


Figure 7. Mean serum ferritin levels at baseline and week 4 in the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups (p<0.001 for both comparisons; mean increase of 7.82 µg/L in GT versus 7.43 µg/L in FG)

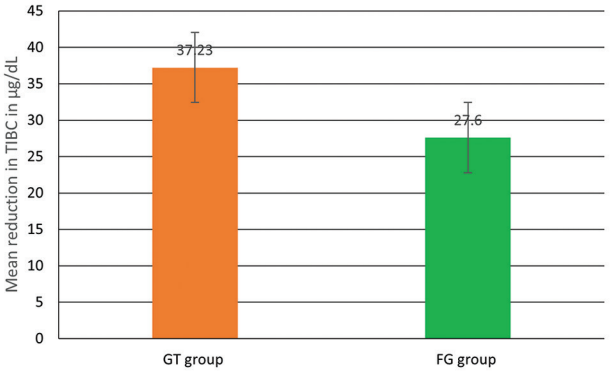


Figure 8. Comparison of mean total iron-binding capacity measurements before and after treatment in the GT (receiving *Grewia tenax* chewable tablets) and FG (receiving ferrous gluconate tablets) groups (p=0.335 at baseline; p=0.137 at week 4)

statistical significance, reflecting lower potency. However, the clinical improvement observed in some patients indicates potential bioactive mechanisms beyond iron supplementation alone.

Chi-square analysis revealed that 50% of participants in the GT group and 78% in the FG group responded positively to treatment (p=0.09). Although this difference did not reach statistical significance, the trend suggests greater efficacy of ferrous gluconate, which may be confirmed in future trials with larger sample sizes.

Reticulocyte dynamics further distinguished the two treatments. Participants receiving *G. tenax* displayed an early rise in reticulocyte counts within 5–10 days, followed by a decline, consistent with expected erythropoietic stimulation.[5] In contrast, reticulocyte counts declined steadily in the FG group. The early reticulocyte response in the GT group suggests a potential erythropoietic stimulus mediated by its phytochemical constituents—including flavonoids, β-carbolines, and ascorbic acid—which may enhance iron utilization and provide antioxidant protection. This potential of *G. tenax* to improve iron bioavailability and stabilize iron stores is an intriguing aspect that warrants further investigation and may offer new avenues in the treatment of IDA.

At baseline, serum iron levels were comparable between the two groups, with no statistically significant difference (p=0.119). At the end of treatment (4 weeks), the FG group demonstrated a significantly greater increase compared to the GT group (p=0.048), indicating that ferrous gluconate more effectively improves circulating serum iron.

At baseline, serum ferritin levels differed significantly between the two groups (p<0.001), with the GT group having significantly higher ferritin stores (approximately 81 µg/dL vs. 6 µg/dL in the FG group). This baseline imbalance may confound the interpretation of ferritin changes over time. At week 4 (end of treatment), serum ferritin levels remained significantly higher in the GT group, with mean increases of 7.82 µg/L in the GT group and 7.43 µg/L in the FG group (p<0.005). This suggests that participants in the GT group already had higher iron stores before treatment. While ferrous gluconate markedly improved hemoglobin levels, ferritin levels remained relatively low. Clinically, this indicates that *G. tenax* may contribute to maintaining or stabilizing iron stores, whereas ferrous gluconate primarily increases functional hemoglobin and circulating iron.

For TIBC, no significant differences were observed between the two groups at baseline (p=0.335 for the FG group; p=0.137 for the GT group), suggesting comparable initial iron-binding capacity. However, the GT group demonstrated a greater reduction in TIBC prior to full correction of hemoglobin levels, suggesting early suppression of iron absorption through increased transferrin saturation.[41,42] This finding raises the possibility that *G. tenax* not only enhances iron bioavailability but also exerts a regulatory effect to prevent excessive iron accumulation, in line with physiological iron homeostasis mechanisms.

Although ferrous gluconate demonstrated superior hematological response within 2–4 weeks, *G. tenax* produced clinically meaningful improvements despite delivering a substantially lower elemental iron dose (0.9 mg/day vs. 72 mg/day). This observation supports the concept that iron bioavailability, rather than absolute iron dose, plays a critical role in therapeutic outcomes, consistent with prior reports demonstrating an inverse relationship between iron dose and absorption efficiency.[43] Furthermore, while a higher proportion of responders in the FG group met the clinical response criteria (78% vs. 50%), this difference did not reach statistical significance (p=0.09), highlighting the need for larger studies to confirm these findings. The need for further research to investigate the therapeutic potential of *G. tenax* is therefore both urgent and highly relevant to the field of anemia treatment.

In summary, ferrous gluconate is more effective in rapidly increasing hemoglobin and serum iron levels, whereas *G. tenax* appears to enhance iron storage and regulate iron metabolism through mechanisms potentially independent of total iron content. The findings challenge the conventional assumption that efficacy is solely dependent on elemental iron dose and underscore the need for further investigation into the bioavailability-enhancing and regulatory properties of *G. tenax*. These results provide preliminary but clinically meaningful evidence supporting its therapeutic potential as an adjunct or alternative in the management of IDA.

5. Study limitation

This study has several limitations that should be acknowledged:

(i) The groups demonstrated markedly different baseline serum ferritin levels, which may confound the interpretation of treatment effects.
(ii) The small sample size reduces the statistical power of the study analyses.
(iii) The treatment duration (4 weeks) may be insufficient to observe the full hematinic effect of a natural product such as *G. tenax*, which may exert its benefits gradually.
(iv) Dietary iron consumption and nutrition variability among participants were not controlled, potentially influencing hematological measurements.

6. Conclusion

This exploratory clinical trial represents a significant milestone, providing the first human evidence that supports the potential efficacy of *G. tenax* fruits in the management of IDA. Although ferrous gluconate demonstrated greater potency in rapidly increasing hemoglobin and reticulocyte responses, *G. tenax* produced clinically meaningful hematological improvements despite its extremely low elemental iron content. These effects may be attributable to enhanced iron bioavailability through various mechanisms, early stimulation of erythropoietic activity, stabilization of iron stores, and modulation of iron absorption, as reflected by reductions in TIBC.

Importantly, ferrous gluconate demonstrated superior efficacy in raising hemoglobin and serum iron levels; however, *G. tenax* still produced clinically meaningful improvements, likely driven by enhanced bioavailability and antioxidant-mediated regulation. The greater reduction in TIBC observed in the GT group suggests more efficient transferrin saturation or feedback regulation of iron uptake, an effect not observed with ferrous gluconate. Reticulocyte kinetics also differed between the groups: participants receiving *G. tenax* showed an early rise followed by a decline, consistent with physiological erythropoietic patterns, whereas participants receiving ferrous gluconate exhibited a flat or declining trend, suggestive of potential oversaturation or oxidative stress. In addition, ferritin increases were higher in the GT group (≈7.78 ng/mL) compared with the FG group (≈ 7.43 ng/mL), highlighting its role in maintaining iron stores. The notable variability in hemoglobin and iron responses—especially among the GT group—further supports the need for individualized approaches to iron therapy.

Although the small sample size limits generalizability, this pilot study provides proof of concept and highlights the urgent need for larger, controlled clinical trials. If confirmed, *G. tenax* could represent a safe, natural, and potentially more tolerable alternative or adjunct to conventional iron supplements, particularly in resource-limited settings where plant-based remedies are readily accessible and culturally accepted.

Acknowledgments

The authors gratefully acknowledge the University of Science and Technology for providing the research facilities necessary to conduct this study. We extend our special thanks to Dr. Adnan Badwan, Vice President of the Jordanian Pharmaceutical Manufacturing Company (JPM), Amman, Jordan, for his generous support in supplying the *G. tenax* chewable tablets (Tonigrow®) used in the trial.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Sami Ahmed Khalid
Investigation: Randa Alsadig Almahdi
Methodology: Randa Alsadig Almahdi
Writing–original draft: Randa Alsadig Almahdi
Writing–review & editing: Sami Ahmed Khalid

Ethics approval and consent to participate

The ethics approval was granted by the IRB of the University of Science and Technology (Reference No: UST/EC/2025/045). Patients provided written informed consent before enrolling in the study.

Consent for publication

The data of the involved human subjects were anonymous and do not allow for identification of participants.

Availability of data

The data of this work are available upon request from the corresponding authors.

Further disclosure

This clinical trial was part of the work undertaken to complete a Ph.D. thesis; accordingly, it is archived at the University of Khartoum (http://khartoumpace.uok.edu).

References

1. Muñoz M, Villar I, García-Erce JA. An update on iron physiology. *World J Gastroenterol*. 2009;15(37):4617-4626.

doi: 10.3748/wjg.15.4617

2. Theil EC, Chen H, Miranda C, *et al*. Absorption of iron from ferritin is independent of heme iron and ferrous salts in women and rat intestinal segments. *J Nutr*. 2012;142(3): 478-483.

   doi: 10.3945/jn.111.145854

3. Warner MJ, Kamran MT. Iron deficiency anemia. In: *StatPearls*. Treasure Island, FL: StatPearls Publishing; 2023.

4. Pasricha SR, Drakesmith H, Black J, Hipgrave D, Biggs BA. Control of iron deficiency anemia in low- and middle-income countries. *Blood*. 2013;121(14):2607-2617.

   doi: 10.1182/blood-2012-09-453522

5. Umbreit J. Iron deficiency: A concise review. *Am J Hematol*. 2005;78(3):225-231.

   doi: 10.1002/ajh.20249

6. Fernandez-Jimenez MC, Moreno G, Wright I, Shih PC, Vaquero MP, Remacha AF. Iron deficiency in menstruating adult women: Much more than anemia. *Womens Health Rep (New Rochelle)*. 2020;1(1):26-35.

   doi: 10.1089/whr.2019.0011

7. O' Brayant CL, Sompson LA. Chapter 12. Anemias. In: Young LY and Koda-Kimble MA, editor. *Applied Therapeutics: The Clinical Use of Drugs*. 9th ed. Vancouver, WA: Applied Therapeutics Inc; 2008. p. 90-94.

8. Benito P, Miller D. Iron absorption and bioavailability: An updated review. *Nutr Res*. 1998;18:581-603.

   doi: 10.1016/S0271-5317(98)00044-X

9. Ems T, Kayla SL, Huechler MR. Biochemistry of iron absorption book. In: *StatPeals*. Treasure Island, FL: Stat Publishing; 2023.

10. Sheikh NA, Desai T, Kosalge BS. Natural Fe chelators as potential therapeutic agents for iron overload diseases. In: *Trace Elements and Their Effects on Human Health and Conditions*. London: IntechOpen; 2021.

    doi: 10.5.772/IntechOpen.98749

11. Jomova K, Valko M. Importance of iron chelation in free radical induced oxidative stress and human disease. *Curr Pharm Des*. 2011;17(31):3460-3473.

    doi: 10.2174/138161211798072463

12. Arollado EC, Osi MO. Hematinic activity of *Alternanthera sessils* (L.) R. Br. (Armarnthaceae) in mice and rats. *E Int Sci Res J*. 2010;2(2):110-161.

13. Wang X, Li Y, Han L, Li J, Liu C, Sun C. Role of flavonoids in the treatment of iron overload. *Front Cell Dev Biol*. 2021;9:685364.

    doi: 10.3389/fcell.2021.685364

14. Madhikarmi NL, Murthy KR. Antioxidant enzymes and oxidative stress in the erythrocytes of iron deficiency anemic patients supplemented with vitamins. *Iran Biomed J*. 2014;18(2):82-87.

15. Falade OS, Otemuyiwa IO, Oladipo A, Oyedapo OO, Akinpelu BA, Adewusi SR. The chemical composition and membrane stability activity of some herbs used in local therapy for anemia. *J Ethnopharmacol*. 2005;102(1):15-22.

    doi: 10.1016/j.jep.2005.04.034

16. Puntarulo S. Iron, oxidative stress and human health. *Mol Aspects Med*. 2005;26(4-5):299-312.

    doi: 10.1016/j.mam.2005.07.001

17. Pan Y, Qin R, Hon M. The Interactions of polyphenols with Fe and their applications in fenton/fenton-like reactions. *Sep Purif Technol*. 2022;300:121831.

    doi: 10.1016/j.seppur.2022.121831

18. Babayan S, Aslanyan G, Amroyan E, *et al*. Comparative study of Femineral® and Floradix® in women in childbearing ages and adolescent girls with iron deficiency anemia. *Sci Pharmaceutica*. 2008;76:725-742.

    doi: 10.3797//Scipharm.0807-17

19. Chobot V, Huber C, Trettenhahn G, Hadacek F. (+/-)-Catechin: Chemical weapon, antioxidant, or stress regulator?. *J Chem Ecol*. 2009;35(8):980-996.

    doi: 10.1007/s10886-009-9681-x

20. Isemura M. Catechin in human health and disease. *Molecules*. 2019;24(3):528.

    doi: 10.3390/molecules24030528

21. Abdallah FB, Fetoui H, Fakhfakh F, Keskes L. Caffeic acid and quercetin protect erythrocytes against the oxidative stress and the genotoxic effects of lambda-cyhalothrin *in vitro*. *Hum Exp Toxicol*. 2012;31(1):92-100.

    doi: 10.1177/0960327111424303

22. Benariba N, Djaziri R, Bellakhdar W, *et al*. Phytochemical screening and free radical scavenging activity of *Citrullus* colocynthis seeds extracts. *Asian Pac J Trop Biomed*. 2013;3(1):35-40.

    doi: 10.1016/S22211691(13)60020-9

23. Imam MU, Zhang S, Ma J, Wang H, Wang F. Antioxidants mediate both iron homeostasis and oxidative stress. *Nutrients*. 2017;9(7):671.

    doi: 10.3390/nu9070671

24. Elmardi KA, Adam I, Malik EM, *et al*. Prevalence and determinants of anaemia in women of reproductive age in Sudan: Analysis of a cross-sectional household survey. *BMC Public Health*. 2020;20:1125.

    doi: 10.1186/s12889-020-09252-w

25. Ahmed ME, Hamid HBB, Babiker HE, *et al*. Effects of *Grewia tenax* (Goddeim) as a natural food on the hemoglobin level and growth among displaced children of Darfur State,

Western Sudan. *J Med Med Sci.* 2012;3(11):729-733.

26. Ebrahim AM, Eltayeb MH, Khalid H, *et al*. Study on selected trace elements and heavy metals in some popular medicinal plants from Sudan. *J Nat Med*. 2012;66(4):671-679.

    doi: 10.1007/s11418-012-0630-6

27. Sharma N, Patni V. *Grewia tenax* (Frosk.) Fiori.- A traditional medicinal plant with enormous economic prospectives. *AJPCR*. 2012;5(3):28-32.

28. Osman MA. Nutrient composition and antinutritional factors of fiori (*Grewia tenax*) fruit. *J Saud Soc Agric Asci*. 2003;3(1):38-49.

    doi: 10.29328/journal.jhcr.1001026

29. Khemiss F, Ghoul-Mazgar S, Moshtaghie AA, Saidane D. Study of the effect of aqueous extract of *Grewia tenax* fruit on iron absorption by everted gut sac. *J Ethnopharmacol*. 2006;103(1):90-98.

    doi: 10.1016/j.jep.2005.07.017

30. Al Shaer MK, Ahmed AN, Mera NAM. The Effects of tinnas grewia on hemoglobin levels in the females of white rats. *J Home Econ.* 2021;31(3):74-86.

    doi: 10.21608/mkas.2021.87888.1062

31. Tanaka N, Kashiwada Y. Phytochemical studies on traditional herbal medicines based on the ethnopharmacological information obtained by field studies. *J Nat Med.* 2021;75(4):762-783.

    doi: 10.1007/s11418-021-01545-7

32. Almahdi RA, Ali HAR, Khalid SA. Mechanism of action and validation of the traditional medicinal use of *Grewia tenax* fruits in Sudan to encounter iron deficiency anemia. *J Hematol Clin Res*. 2023;7:29-38.

    doi: 10.29328/journal.jhcr.1001026

33. Freedman R. *Famine foods, 1998: Center for New Crops and Plant Products, Tilaceae*. United States: Purdue University. Available from: https://www.purdue.edu/purdue/search.php?q=Grewia+tenax [Last accessed on 2022 Jan 11].

34. Yagi SM, Al Hassan GOM. Nutritional composition of *Grewia tenax* species (*G.tenax* (Forsk.) Fiori, G.fl avescens Juss and *G. viollsa*) wild fruits". *IJFST*. 2010;2(3):159-162.

35. Abdel-Rahman NA, Awad IA, Abdelrahman EE. A study of some Sudanese Edible forest fruits and their nectars. *JAAS Journal*. 2014;22:39-44.

36. Abuagarib EAA, Yang R, Hua X, Siddeeg A. Chemical composition, nutritional properties and volatile compounds of goddeim (*G. tenax*. Forssk) Fiori fruits. *J Food Nutr Res*. 2014;2(4):187-192.

37. Almahdi RA, Khalid SA. An insight into the practice of iron therapy: Contribution to the on-going debate with special reference to low and middle-income countries. *SJMS*. 2021;16(1):8934.

    doi: 10.18502/sjms/16i1.8934

38. Gera T, Sachdev HP, Nestel P, Sachdev SS. Effect of iron supplementation on haemoglobin response in children: Systematic review of randomized controlled trials. *J Pediatr Gastroenterol Nutr*. 2007;44(4):468-486.

    doi: 10.1097/01.mpg.0000243440.85452.38

39. Alleyne M, Horne MK, Miller JL. Individualized treatment for iron-deficiency anemia in adults. *Am J Med*. 2008;121(11):943-948.

    doi: 10.1016/j.amjmed.2008.07.012

40. Okam MM, Koch TA, Tran MH. Iron supplementation, response in iron-deficiency anemia: Analysis of five trials. *Am J Med*. 2017;130(8):991.e1-991.e8.

    doi: 10.1016/j.amjmed.2017.03.045

41. Hawkins RC. Total iron binding capacity or transferrin concentration alone outperforms iron and saturation indices in predicting iron deficiency. *Clin Chim Acta*. 2007;380(1-2):203-207.

    doi: 10.1016/j.cca.2007.02.032

42. Ma J, Wen X, Mo F, Wang X, Shen Z, Li M. Effects of different doses and duration of iron supplementation on curing iron deficiency anemia: An experimental study. *Biol Trace Elem Res.* 2014;162(1-3):242-251.

    doi: 10.1007/s12011-014-0115-4

43. Salman Z, Yilmaz T, Mehmetcik G. The Relationship between ferritin levels and oxidative stress parameters in serum of β-thalassemia major patients. *Arch Biochem Biophys*. 2018;659(1):42-46.

    doi: 10.1016/j.abb.2018.09.020

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

The role of clinical and social criteria in intensive care unit admission decisions: Evidence from a medical decision-making tool

Filipa Madeira[1]*, João Miguel Ferreira[2], Dulce Correia[2], Nuno Gaibino[2], Renato Reis[2], and Cicero Roberto Pereira[1]

[1]Institute of Social Sciences, University of Lisbon, Lisbon, Portugal

[2]Department of Intensive Care Unit, Lisbon North University Hospital Center, Lisbon, Portugal

Abstract

Background: In critical care settings, especially during surges such as the COVID-19 pandemic, physicians are often required to make rapid triage decisions under resource constraints. While clinical indicators should ideally guide these decisions, emerging research indicates that non-clinical factors—such as a patient's race or gender—may inadvertently affect judgment. **Aim:** This study aims to validate clinical profiles using a decision-making tool and examine the predictive role of clinical and social factors in intensive care unit (ICU) admission decisions under contingency conditions, such as during the COVID-19 pandemic. **Methods:** A total of 432 ICU admission decisions (trials) were collected from a simulated task in which nine ICU physicians evaluated 48 fictional patient profiles under conditions of limited bed availability. Each participant reviewed all profiles and selected half of the profiles for admission. Each profile included six clinical criteria (e.g., prognosis, comorbidities, and respiratory failure severity) and two non-clinical features (i.e., gender and race). The trial served as the unit of analysis. Multilevel logistic regressions assessed the predictive power of clinical and social variables on acceptance decisions, omission errors (rejecting qualified candidates), and false alarms (accepting less qualified candidates). **Results:** The findings demonstrate that participants relied primarily on clinical information: high-scoring profiles were admitted more often (67.4% and 52.5%) than lower-scoring ones (37.5% and 13.5%) ($p<0.001$). However, social factors also shaped decisions. Male candidates were more likely to be admitted than females ($b = -0.51$; $t = -4.35$; $p<0.001$; 95% confidence interval [CI] = [−0.75, −0.27]), and Black candidates were admitted more than White candidates ($b = -0.52$; $t = -3.34$; $p<0.001$; 95% CI = [−0.85−−0.19]), even when less qualified, thereby suggesting possible overcorrection. **Conclusion:** Although clinical criteria primarily guided ICU admission decisions, social characteristics also subtly influenced outcomes. Together, these findings validate the novel decision-making paradigm as a valuable tool for assessing both clinical accuracy and the presence of social bias in triage contexts. They also provide empirical evidence that, under pressure and uncertainty, healthcare professionals may be susceptible to the influence of social cues. **Relevance for patient:** This study explores how critical care physicians decide who receives life-saving treatment in ICUs, especially during times of crisis when medical resources are limited. By simulating real-world triage situations, the research shows that even experienced clinicians may be influenced by non-clinical factors—such as a patient's race or gender—despite their intention to prioritize clinical indicators.

***Corresponding author:**
Filipa Madeira
(filipa.madeira@ics.ulisboa.pt)

Citation: Madeira F, Ferreira JM, Correia D, Gaibino N, Reis R, Pereira CR. The role of clinical and social criteria in intensive care unit admission decisions: Evidence from a medical decision-making tool. *J Clin Transl Res*. 2025;11(6):39-49. doi: 10.36922/JCTR025280040

Received: July 11, 2025

Revised: October 11, 2025

Accepted: October 17, 2025

Published online: December 3, 2025

These findings highlight the need for increased awareness and targeted training to reduce unconscious bias in clinical decision-making.

Keywords: Intensive care unit admission; COVID-19; Decision-making; Assessment tool; Social biases; Resource allocation

1. Introduction

The medical decision-making process in the context of intensive care unit (ICU) admission is inherently complex, particularly during times of crisis, such as the COVID-19 pandemic, when healthcare systems faced unprecedented pressure due to high patient volumes and limited resources.[1,2] The resulting imbalance between demand and ICU bed availability generated a scenario of resource scarcity, raising profound ethical and clinical challenges in prioritizing patients for life-saving care.[3]

In response, several international organizations issued guidelines aimed at promoting equity, transparency, and consistency in ICU admission protocols under resource-limited conditions.[4] However, despite these efforts, significant disparities in ICU capacity and emergency preparedness were evident across European healthcare systems. These structural differences contributed to uneven mortality outcomes and exposed vulnerabilities in healthcare infrastructures.[5] Moreover, early warning signs of the pandemic's spread were already visible in social media data before official responses were mobilized, highlighting discrepancies in institutional readiness and the timeliness of public health interventions.[6] Such variations not only shaped the trajectory of the pandemic across different regions but also exacerbated inequalities in ICU access and patient outcomes.

While clinical frameworks have been established to promote fairness in ICU admissions, particularly during crises such as the COVID-19 pandemic, evidence shows that medical decisions remain vulnerable to the influence of both patient-related and provider-related factors.[7] One underexplored yet critical contributor to inequities is the role of cognitive and social biases under contingency scenarios—an area well documented in social psychology.[8] When clinicians operate under intense pressure and cognitive load, they must process complex information rapidly, increasing their reliance on mental shortcuts or heuristics. While these strategies may improve efficiency, they can also unintentionally increase the likelihood of introducing social bias into clinical judgments.[9-11]

A growing body of interdisciplinary research suggests that resource scarcity may not only constrain logistical capacities in clinical environments but also influence the cognitive and social dynamics underlying decision-making. Under conditions of scarcity, individuals tend to exhibit increased sensitivity to social group distinctions, including heightened protection of perceived in-group members and increased vigilance or exclusion of out-group members.[12,13] Scarcity can therefore act as a psychosocial stressor that amplifies implicit social biases, particularly in time-sensitive and high-stakes contexts such as intensive care triage. During the COVID-19 pandemic, the acute shortage of ICU beds and the overwhelming demand for care reduced opportunities for deliberative, criterion-based reasoning—an essential safeguard in clinical decision-making. In such circumstances, non-clinical characteristics such as race, gender, or socioeconomic background may unconsciously influence clinical judgments.[14-16] Although clinicians are trained to prioritize clinical criteria, racial disparities in care have been documented across multiple domains. For instance, White patients are more likely to be referred for cardiac interventions than Black patients,[17] and disparities in pain management persist, with Black children receiving less analgesia than their White counterparts.[18] However, the literature also reports inconsistent findings, with some studies reporting no significant differences based on race, highlighting the complex and context-dependent nature of bias in clinical decision-making.[19]

More recently, attention has turned to racial disparities in access to and outcomes of intensive care, particularly among Black patients. Black racial groups experienced higher rates of ICU admission and poorer clinical outcomes compared to White patients.[20,21] For example, during the COVID-19 pandemic, Black patients were disproportionately represented in ICU admissions and experienced elevated in-hospital mortality. In addition, studies have shown that Black patients were less likely to receive invasive life-sustaining treatments, such as mechanical ventilation.[18] These inequities are not fully attributed to differences in socioeconomic status or comorbidities; as suggested by previous evidence, they are also likely to result from individual factors, such as implicit provider biases, and structural factors within the healthcare system.[22,23]

Previous studies have provided evidence of the complexity of ICU admission decisions[24,25]; however, many

have relied on qualitative or retrospective approaches that, although rich in descriptive detail, offer limited capacity to isolate causal effects. For example, Gopalan and De Vasconcellos[25] employed a structured interview method—the "20 Questions" approach—to explore the decision-making rationale of ICU physicians. This design was effective in uncovering key themes in clinical reasoning but lacked experimental control and did not permit systematic manipulation of patient attributes, thereby limiting its ability to identify underlying biases.

In contrast, the present study introduces a controlled experimental paradigm using a medical decision-making tool (MDMT) specifically designed to examine how physicians integrate clinical and social information when making triage decisions under resource scarcity. Developed during the COVID-19 pandemic through an interdisciplinary collaboration between critical care specialists and social scientists, the MDMT presents participants with multiple standardized fictional patient profiles that systematically vary across six clinical criteria (e.g., comorbidities, prognosis) and two social attributes (gender and race).

This design offers two key methodological advantages. First, it allows for direct comparison of decision patterns across controlled conditions, thereby minimizing confounding factors and enhancing internal validity. Second, by simulating real-time decisions under high-pressure, resource-limited scenarios, the paradigm makes it possible to detect subtle, potentially unconscious social biases that might remain unacknowledged in self-reports or interviews.

The MDMT was specifically developed to assess two complementary outcomes: (i) decision-making accuracy, defined as the physician's ability to consistently prioritize the most clinically appropriate candidates; and (ii) social bias, operationalized as systematic variation in decision patterns based on non-clinical characteristics such as gender or race. By combining clinical precision with sensitivity to social determinants of care, this study provides a robust and scalable tool to the growing body of research on fairness and equity in critical care decision-making.

2. Methods

2.1. Study design and ethical approval

This study employed an experimental, within-subjects design to examine how clinical and social factors influence ICU admission decisions under conditions of medical resource scarcity. The study protocol was approved by a local Institutional Review Board (Approval No. 30/21), and all participants provided informed consent before participation, in accordance with the Declaration of Helsinki.

2.2. Participants

Nine ICU physicians (five males, four females) aged 30–62 years volunteered to participate and provided informed consent. All participants completed an online experiment consisting of a medical decision-making task followed by a demographic questionnaire. Although the number of participants was nine, each physician evaluated 48 fictional patient profiles, yielding a total of 432 decision trials. The unit of analysis for all statistical models was at the trial level.

2.3. Procedure

The decision-making task simulated a scenario in which ICU resources were limited, and physicians were required to triage 48 fictional patient profiles for ICU admission. The task was divided into two phases:

(i) Viewing phase: Participants passively viewed all 48 fictional patient profiles for one second each in a randomized order to gain an overview of the range of clinical profiles.
(ii) Selection phase: Participants reviewed each patient profile one at a time, again in random order, and were given the following instructions:

Imagine that 48 clinical cases are being reviewed for ICU admission. However, there are only 24 beds available. Your task is to admit the 24 clinical cases for which ICU admission is most appropriate and reject the 24 clinical cases for which admission is less appropriate. You will first view all clinical cases. When you are done reviewing the clinical cases, your task is to select half (24 clinical profiles) for ICU admission.

Admission decisions were made by clicking either the "Accept" or "Reject" button. There was no time limit for decision-making.

2.3.1. Materials and clinical profiles

Each clinical profile consisted of six pieces of information relevant to ICU triage: age (30–40 vs. 80–90); obesity (body mass index: 20–45); comorbidities (conditions included non-Hodgkin's lymphoma, acute leukemia, arterial hypertension, metastatic esophageal cancer, multiple myeloma, and diabetes); degree of respiratory failure (minimal to moderate or severe); clinical progression (stable or worsening); and prognosis (presence of organ failure, shock, serum lactate levels twice the normal range, or anuric acute kidney injury). Participants were instructed to consider all clinical information when making their decisions.

To operationalize clinical qualification in a standardized manner, each fictional patient profile was scored based on six clinical attributes: age, comorbidities, oxygen saturation, respiratory rate, organ failure status, and prognosis. Each attribute was evaluated by an expert panel of critical care physicians and assigned a value of 1 (less favorable) or 2 (more favorable), based on established triage guidelines and clinical judgment. The total clinical score for each profile was computed by summing the six attribute values, resulting in scores ranging from 6 (lowest qualification) to 12 (highest qualification).

To enhance internal consistency and eliminate potential outliers, profiles with extreme total scores (6, 11, or 12) were excluded from the final set. This decision avoided ceiling and floor effects while maintaining comparability between profiles. The remaining profiles—those with total scores of 7 to 10—were grouped into two levels of clinical qualification for analytical purposes: less qualified (scores of 7 or 8) and more qualified (scores of 9 or 10). This structured scoring system allowed for controlled variation across profiles, ensuring that physicians were exposed to clinically plausible yet systematically differentiated clinical profiles.

2.3.2. Social variables and experimental design

To introduce social information, each profile was paired with one of 16 blurred facial images. These images represented four demographic categories: Black male, White male, Black female, and White female (n = 4 per group). The images were randomly assigned across profiles, ensuring equal distribution by qualification level and gender. In addition, a "neutral" condition (no image) was included to assess the independent effect of gender and race. The final experimental design was a 2 (gender: Male vs. female) × 3 (race: Black vs. White vs. neutral) × 2 (qualification: More qualified vs. less qualified) within-subjects model. Given the within-subjects experimental design, the primary unit of analysis was the ICU admission decision (trial), not the physician. This approach allowed for a robust estimation of patterns across multiple decision points.

2.4. Statistical analysis

All statistical analyses were performed using JASP version 0.19.3 (Jasp Team, The Netherlands). Multilevel logistic regression models were estimated to account for the hierarchical structure of the data (trial-level decisions nested within physicians). The unit of analysis was the individual ICU admission decision (trial), yielding a total of 432 observations.

To assess task accuracy, we first fitted a multilevel logistic regression model with profile qualification (more qualified vs. less qualified) as the independent variable and admission decision (accept = 1, reject = 0) as the binary outcome. This model tested whether participants systematically preferred more clinically qualified profiles.

Next, we fitted a second set of multilevel logistic models that included clinical variables (e.g., age category, comorbidities, respiratory failure severity) and social variables (gender, race) as fixed effects. We also examined interaction effects between clinical and social predictors to assess whether social characteristics moderated the impact of clinical qualification.

Each model included a random intercept for the physician to accommodate within-participant correlations. Model outputs are presented as regression coefficients (b), standard errors (SE), Wald z-statistics, p-values, and 95% confidence intervals (CI).

3. Results

3.1. Validation of the clinical profiles

A total of 432 decisions were analyzed to assess whether profile qualification influenced ICU admission decisions. We first examined the differences in participants' selection of more qualified patient profiles and rejection of less qualified profiles. This analysis was crucial for assessing task accuracy—specifically, whether participants systematically used clinical profile qualification as a criterion in their decision-making process.

A multilevel logistic regression analysis was conducted, with clinical profiles (Profiles 7–10) as the independent variable and ICU admission decision (accept vs. reject) as the dependent variable. The model revealed a statistically significant effect of profile qualification on candidate acceptance, χ^2 (1, 3) = 16.114, with p<0.001, suggesting that participants used the qualification level of the profiles as a critical factor in their decisions.

As shown in Table 1, the probability of acceptance increased with clinical qualification: Profile 10 had a 67.4% acceptance rate, followed by Profile 9 (52.5%). In contrast, less qualified profiles—Profile 8 and Profile 7—had lower acceptance rates of 37.5% and 13.5%, respectively. These differences indicate that participants reliably used clinical profile information when selecting ICU candidates.

3.2. Clinical criteria as predictors of ICU admission

To assess the impact of clinical variables on ICU admission, a logistic regression model was applied, including six clinical parameters: age, obesity, comorbidities, degree of respiratory failure, clinical evolution, and prognosis. As displayed in Table 2, five out of six clinical parameters significantly predicted the likelihood of

Table 1. Descriptives of acceptance rate for each clinical profile

Profile	Acceptance rate (%)	SD	95% CI	
			LI	LS
7	13.5	0.065	0.049	0.319
8	37.5	0.092	0.217	0.565
9	52.5	0.081	0.368	0.676
10	67.4	0.071	0.522	0.796

Abbreviations: CI: Confidence interval; LI: Lower bound of confidence interval; LS: Upper bound of confidence interval; SD: Standard deviation.

Table 2. Clinical factors predicting the likelihood of accepting ICU candidates

Predictor	*b*	SE	*t*	*p*-value	95% CI of *b*	
					Lower	Upper
Intercept	0.104	0.371	0.280	0.779	−0.624	0.832
Age	−0.815**	0.188	−4.342	<0.001	−1.184	−0.446
Obesity	0.063	0.131	0.484	0.629	−0.195	0.321
Comorbidities	−0.659**	0.137	−4.817	<0.001	−0.929	−0.389
Respiratory failure	−0.798*	0.281	−2.838	0.005	−1.351	−0.245
Clinical evolution	−0.343*	0.139	−2.460	0.014	−0.616	−0.070
Prognosis	−0.589**	0.167	−3.531	<0.001	−0.917	−0.261

Note: *p<0.05 and **p<0.01.
Abbreviations: CI: Confidence interval; SE: Standard error.

candidate acceptance. The findings demonstrate that age, comorbidities, respiratory failure severity, clinical evolution, and prognosis were all statistically significant. In contrast, obesity had no significant predictive value.

Younger candidates (mean = 0.64; standard deviation [SD] = 0.07) were more likely to be accepted than older candidates (mean = 0.39; SD = 0.06); b = −0.815; SE = 0.188; t = −4.342; p<0.001; 95% CI = [−1.184, −0.446]). Candidates with less severe comorbidities (mean = 0.61; SD = 0.07) were more likely to be accepted than those with severe conditions (mean = 0.42; SD = 0.06; b = −0.659; SE = 0.137; t = −4.817; p<0.001; 95% CI = [−0.929, −0.389]. Similarly, profiles with severe respiratory failure were more likely to be accepted (mean = 0.65; SD = 0.07) than those with moderate respiratory failure (mean = 0.39; SD = 0.06; b = −0.798; SE = 0.281; t = −2.838; p=0.005; 95% CI = [−1.351, −0.245]).

Candidates with worsening clinical progression were more likely to be accepted (mean = 0.57; SD = 0.07) than those with stable progression (mean = 0.47; SD = 0.06; b = −0.34; SE = 0.14; t = −2.46; p<0.01; 95% CI = [−0.616, −0.070]). Finally, profiles with prognoses involving multiple organ failure were more likely to be accepted (mean = 0.61; SD = 0.07) compared to candidates without such complications (mean = 0.43; SD = 0.06; b = −0.59; SE = 0.17; t = −3.53; p<0.001; 95% CI = [−0.917, −0.261]). However, obesity was not associated with any significant difference in acceptance rates (b = 0.063; SE = 0.131; t = 0.484; p=0.629).

Figure 1 illustrates the effect of clinical criteria on ICU admission decisions, highlighting how these parameters shaped acceptance probability.

3.3. Social criteria as predictors of ICU admission decisions

A second analysis examined the influence of gender, race, and clinical profile qualification. As expected, more qualified candidates (mean = 0.63; SE = 0.09) were accepted more often than less qualified ones (mean = 0.27; SE = 0.07; b = −0.76; t = − 6.34; p<0.001; 95% CI = [−0.99, −0.52]).

However, social factors also affected ICU admission decisions. Male candidates were significantly more likely to be accepted (mean = 0.57; SE = 0.09) compared to female candidates (mean = 0.32; SE = 0.08; b = −0.51; t = −4.35; p<0.001; 95% CI = [−0.75, −0.27]). Likewise, Black candidates were more likely to be accepted (mean = 0.51; SE = 0.10) than White candidates (mean = 0.32; SE = 0.09; b = −0.52; t = −3.34; p<0.001; 95% CI = [−0.85, −0.19]).

As displayed in Figure 2, a three-way interaction between profile qualification, gender, and race revealed a nuanced pattern. Female candidates reached acceptance rates above 50% when profiles were more qualified and presented with a neutral (no image) profile. In all other conditions, acceptance rates for female candidates fell below this threshold. In contrast, Black male candidates with less qualified profiles were accepted at rates similar to those with more qualified ones, suggesting potential overcorrection or compensatory bias. White male profiles with lower qualifications, and those with neutral images, had acceptance rates below 20%.

We then analyzed two types of decision errors:
(i) Omissions: Rejection of more qualified candidates
(ii) False alarms: Acceptance of less qualified candidates.

3.3.1. Omissions

More qualified female candidates were more likely to be rejected (mean = 0.24; SE = 0.04) compared to qualified male candidates (mean = 0.12; SE = 0.04; b = 0.44; t = −2.80; p=0.01; 95% CI = [0.13, 0.75]). Race had no significant effect when comparing Black versus White candidates (b = 0.29; t = 1.475; p=0.14; 95% CI = [−0.10, 0.68]), but candidates with no image (neutral profile) were

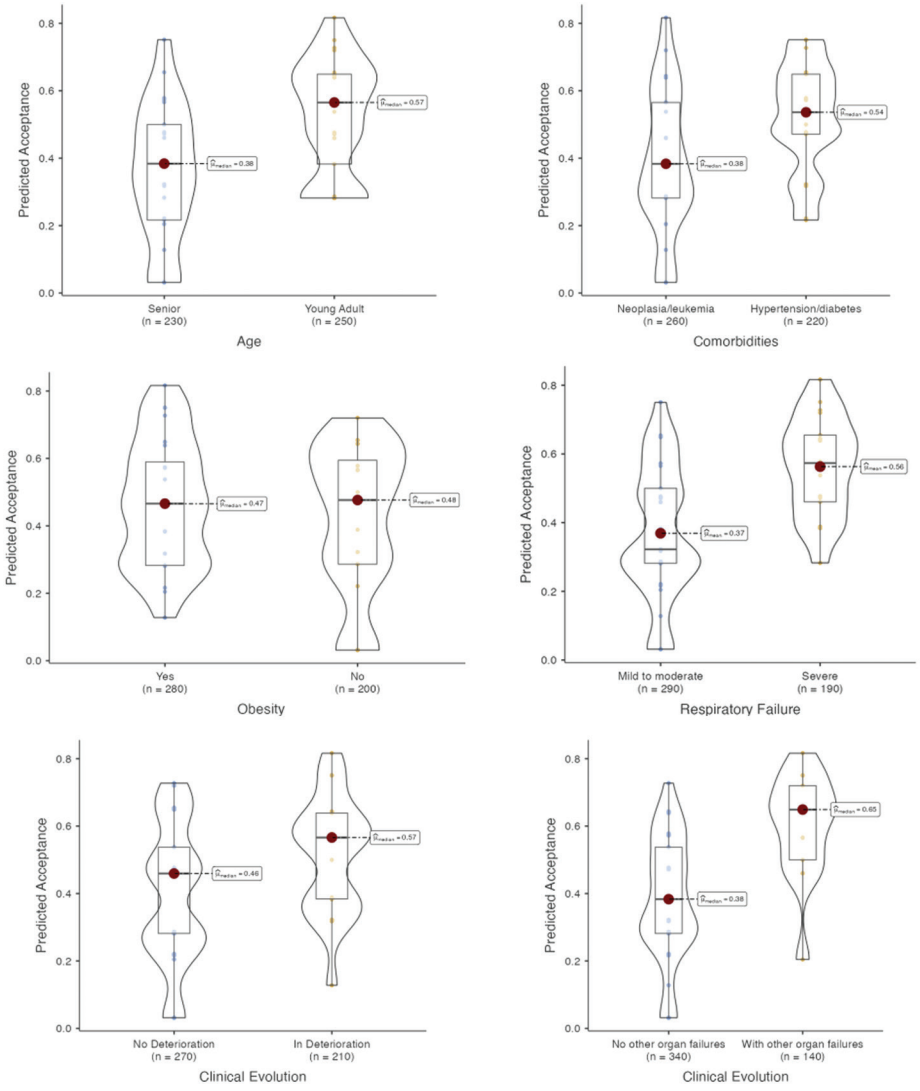


Figure 1. Effects of clinical criteria on intensive care unit admission decisions

less likely to be rejected ($b = -0.75$; $t = -2.66$; p=0.008; 95% CI [−1.30, −0.20].).

These findings are presented in Figure 3 and summarized in Table 3.

3.3.2. False alarms

No significant differences were observed in the acceptance rate of less qualified candidates based on gender ($b = -0.19$; $t = -0.79$; p=0.43; 95% CI = [−0.64, 0.26]) or race ($b = -0.19$; $t = -0.59$; p=0.56; 95% CI = [−0.84, 0.46]). However, as displayed in Figure 4 there was a near-significant interaction between gender and race ($b = 0.96$; $t = 1.79$; p=0.07; 95% CI [−0.09, 2.02]), suggesting a trend in which Black male candidates were more likely to be incorrectly accepted (mean = 0.28; SE = 0.06) than other groups (Table 4).

4. Discussion

The process of ICU admission decision-making is critical and complex, particularly in contexts of resource scarcity, such as during the COVID-19 pandemic. While guidelines are developed to ensure fairness, non-clinical factors—such as patient demographics—may unconsciously influence medical decisions. The present study examined the predictive role of clinical and social factors in ICU admission decisions using a validated MDMT, assessing both decision accuracy and potential biases.

The results provide strong evidence supporting the validity of the MDMT clinical profiles. Our findings indicate that participants consistently admitted more qualified profiles (Profiles 9 and 10), with acceptance rates above

Table 3. Predictors of omissions (rejection of qualified candidates)

Predictor	*b*	SE	*t*	*p*-value	95% CI for *b*	
					Lower	Upper
Intercept	−1.594	0.234	−6.812	<0.001	−2.053	−1.135
Gender	0.437*	0.156	2.798	0.005	0.131	0.743
Race (White vs. neutral)	−0.746*	0.280	−2.662	0.008	−1.296	−0.196
Race (White vs. Black)	0.293	0.198	1.475	0.140	−0.096	0.682
Gender-race (White vs. neutral)	0.345	0.255	1.352	0.176	−0.155	0.845
Gender-race (White vs. Black)	−0.012	0.194	−0.064	0.949	−0.392	0.368

Note: *$p<0.05$.
Abbreviations: CI: Confidence interval; SE: Standard error.

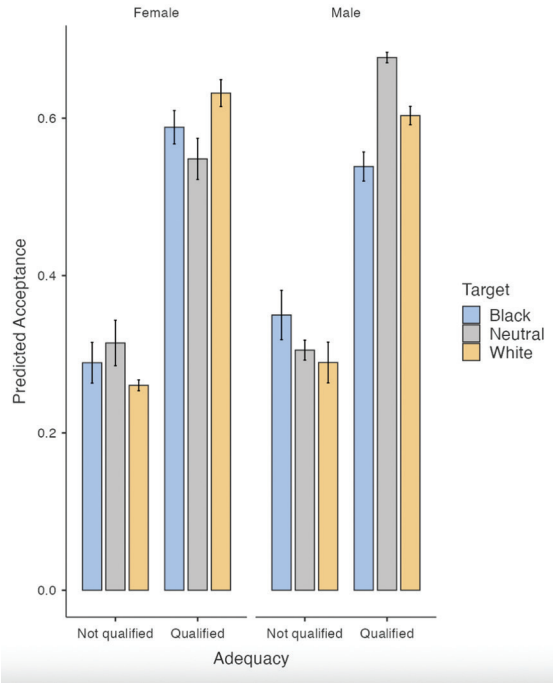


Figure 2. Effects of profile qualification, gender, and race on intensive care unit admission decisions

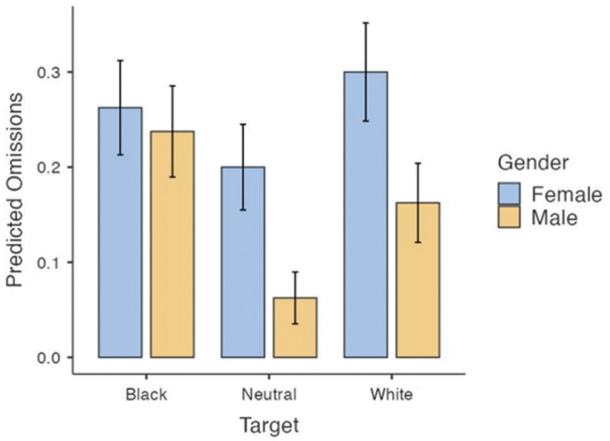


Figure 3. Effects of gender and race on omissions

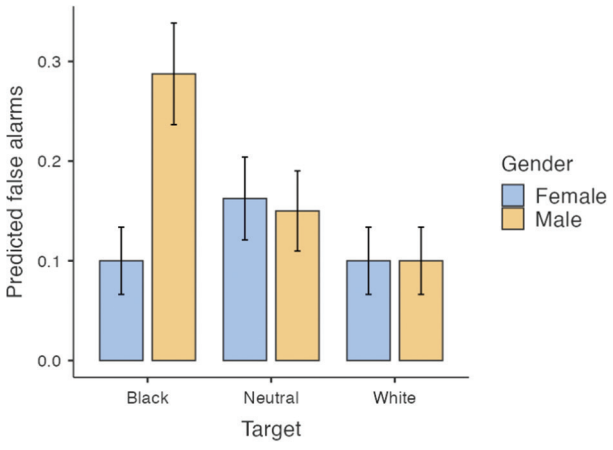


Figure 4. Effects of gender and race on false alarms

50%, while less qualified profiles (Profiles 7 and 8) were generally rejected, with acceptance rates below 50%. These findings demonstrate the tool's sensitivity in distinguishing clinical appropriateness, validating its use in simulating real-world ICU decision-making.

Beyond clinical criteria, the findings revealed subtle yet meaningful influences of social factors on decision-making outcomes. Male candidates were significantly more likely to be accepted (57%) compared to female candidates (32%), despite having similar clinical profiles. This gender disparity reflects patterns documented in prior research, where male patients often receive more aggressive interventions or are perceived as more clinically urgent.[26-28] These perceptions may be driven by implicit assumptions about gender and severity of illness, which can affect physician judgment even in controlled settings.

Racial-related differences were also observed. Black candidates had a higher acceptance rate (51%) than White candidates (32%), a finding that initially

Table 4. Predictors of false alarms (acceptance of less qualified candidates)

Predictor	*b*	SE	*t*	*p*	95% CI for *b*	
					Lower	Upper
Intercept	−2.412	0.438	−5.511	<0.001	−3.270	−1.554
Gender	−0.185	0.233	−0.795	0.427	−0.641	0.271
Race (White vs. neutral)	−0.282	0.411	−0.685	0.493	−1.088	0.524
Race (White vs. Black)	−0.194	0.332	−0.585	0.559	−0.844	0.456
Gender-race (White vs. neutral)	0.964	0.539	1.787	0.074	−0.092	2.020
Gender-race (White vs. Black)	−0.147	0.380	−0.387	0.699	−0.892	0.598

Abbreviations: CI: Confidence interval; SE: Standard error.

appears to contradict prior research documenting worse treatment outcomes and lower procedure rates for racial minorities.[29,30] However, further analysis suggests that this pattern may reflect compensatory behavior. Specifically, less qualified Black male candidates were admitted at rates similar to more qualified candidates, potentially indicating a motivation to avoid perceived racial bias. This interpretation aligns with aversive racism theory, which posits that individuals motivated to appear egalitarian may overcorrect their decisions in favor of historically marginalized groups in ambiguous contexts.[31]

These findings are consistent with the broader literature on healthcare disparities and highlight how contextual factors, such as medical contingency situations in high-stakes environments, can amplify the influence of cognitive heuristics in clinical judgment.[10,32] Importantly, such influences may not arise from deliberate and intentional bias but rather from scarcity-induced reliance on mental shortcuts—a common feature of decision-making under contingency circumstances.[8,12,33,34]

5. Limitations and future directions

While the findings provide important empirical evidence, some limitations should be acknowledged. Although the number of participating physicians was relatively small (n = 9), it is important to clarify that the unit of analysis in our study was the observed decision behavior in the randomized controlled trial, rather than the individual participant. Each of the nine ICU physicians made 48 independent ICU admission decisions, resulting in 432 unique trials. This trial-level dataset enables the identification of systematic patterns across diverse clinical scenarios and allows testing the effects of our experimental manipulations with adequate statistical power. This approach is consistent with established practices in cognitive neuroscience and decision science, where relatively small participant samples are often used, but a large number of within-subject observations provides sufficient power to detect meaningful effects.[35-37]

In the present study, concerns about generalizability refer more to the diversity of decision patterns observed than to the sample size per se. Using this methodological design, we were able to detect robust, statistically significant effects—including consistent patterns of clinical prioritization and evidence of social bias—at the behavioral level. While these effects were strong in our dataset, it is plausible that their magnitude may be even greater in larger and more heterogeneous physician samples, where factors such as training background, institutional culture, or other demographic characteristics could serve as important moderators.

Second, the MDMT employed an internal scoring system used to construct clinical profiles. Although the fixed scale (6–12) ensured consistency across conditions, it was not visible to participants and served only to categorize profiles based on expert-rated clinical attributes. Nonetheless, participants may have inferred implicit patterns across similar profiles, potentially relying on heuristics rather than naturalistic reasoning. To minimize this, profiles were randomized and diversified. Future studies could employ less structured or dynamic formats to further reduce perceived regularities and capture clinical reasoning more authentically.

Moreover, future studies may benefit from integrating training methodologies inspired by recent advances in medical artificial intelligence (AI).[38] While AI-based systems excel at extracting patterns from large-scale clinical data, they typically rely on retrospective observational inputs and lack experimental manipulation of social variables. Combining the predictive power of AI with controlled experimental paradigms like the MDMT could offer a more comprehensive understanding of how social information is cognitively processed and used in triage decisions, particularly under conditions of pressure and uncertainty.

Third, although the study identified patterns consistent with social bias, the underlying psychological

mechanisms—such as implicit prejudice, stereotyping, or aversive bias—were not directly assessed. To better understand the cognitive and affective drivers of these decisions, future research could incorporate behavioral data with implicit measures (e.g., implicit association test), allowing researchers a more nuanced understanding of how social cues are cognitively processed and translated into clinical decisions during contingency scenarios characterized by uncertainty and pressure.

A further methodological consideration concerns the constrained scoring format used to represent clinical severity. While the structured nature of the MDMT mirrors the logic of real-world triage tools, it may also have encouraged pattern-seeking or compensatory strategies among participants, potentially introducing artificial consistency in responses. Exploring alternative formats—such as dynamic clinical scenarios or open-ended justifications—could provide richer data and mitigate the influence of response heuristics.

Finally, although this study focused on race and gender as social dimensions of interest, other socially relevant variables—such as socioeconomic status, immigration background, or disability—may also influence clinical decision-making. Future research should expand the scope of analysis to include these intersecting factors to capture the multidimensional nature of healthcare disparities more comprehensively.

6. Conclusion

Our findings support the use of the MDMT as an effective experimental paradigm to assess how both clinical and social information influence ICU admission decisions under conditions of resource scarcity. While the findings indicate that participants consistently prioritized more clinically appropriate candidates—indicating accurate clinical judgment—systematic patterns of gender and racial bias were also observed. These findings highlight the need to address intersectional biases in critical care environments, particularly during high-pressure decision-making.

The findings suggest that efforts to promote fairness in ICU triage should include the implementation of structured decision-making protocols, evidence-based support tools, and targeted training programs in implicit bias and aversive racism mitigation. As healthcare systems continue to face global crises requiring ethically complex resource allocation, ensuring equitable access to life-saving care must remain a central ethical and clinical priority.

From a policy and practice standpoint, these findings underscore the need for integrating bias-mitigating safeguards into ICU triage processes. These may include transparent eligibility criteria, anonymized review mechanisms, and AI-assisted decision-support systems designed with embedded fairness constraints. Investing in such systemic reforms is essential to strengthening public trust in healthcare systems.

Acknowledgments

The authors would like to thank Alexandre Vieira for his assistance with study programming.

Funding

This research was supported by a research grant from "*la Caixa*" Social Observatory (O*bservatório Social* "*la Caixa*") (LL 20-30) awarded to FM.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: All authors
Data curation: Filipa Madeira
Investigation: Filipa Madeira, João Miguel Ferreira, Dulce Correia, Nuno Gaibino, Renato Reis
Methodology: All authors
Supervision: João Miguel Ferreira
Writing–original draft: Filipa Madeira
Writing–review & editing: All authors

Ethics approval and consent to participate

The study protocol was approved by a local Institutional Review Board (Approval No. 30/21), and all participants provided informed consent before participation, in accordance with the Declaration of Helsinki.

Consent for publication

All participants were informed about the nature of the study and provided consent to participate; however, no individual data are published that would require additional consent for publication.

Availability of data

The dataset used and/or analyzed during the study is available from the corresponding author upon reasonable request. All data generated or analyzed during this study are included in this published article.

Further disclosure

Part of or the entire set of findings have been presented in the conference *International Congress of Psychology*, *Prague*,

2024 (August, 2024).

References

1. Teixeira C, Rosa RG, Rodrigues Filho EM, Sganzerla D. The medical decision-making process in the time of the coronavirus pandemic. *Rev Bras Ter Intensiva*. 2020;32(3):373-377.

   doi: 10.5935/0103-507X.20200059.

2. Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. *JAMA*. 2020;323(16):1545-1546.

   doi: 10.1001/jama.2020.4031

3. Galvagni L, Raho JA. Ethical prioritization of critical care resources during COVID-19: Perspectives from Italy and the United States. *Theor Med Bioeth.* 2024;45:167-181.

   doi: 10.1007/s11017-024-09672-4

4. White DB, Lo B. A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA*. 2020;3237(18):1773-1774.

   doi: 10.1001/jama.2020.5046

5. Giancotti M, Lopreite M, Mauro M, Puliga M. The role of European health system characteristics in affecting COVID-19 lethality during the early days of the pandemic. *Sci Rep.* 2021;11(1):1-8.

   doi: 10.1038/s41598-021-03120-2

6. Lopreite M, Panzarasa P, Puliga M, Riccaboni M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci Rep.* 2021;11:2147.

   doi: 10.1038/s41598-021-81333-1

7. Hajjaj FM, Salek MS, Basra MK, Finlay AY. Non-clinical influences on clinical decision-making: A major challenge to evidence-based practice. *J R Soc Med*. 2010;103(5): 178-187.

   doi: 10.1258/jrsm.2010.100104

8. Madeira F, Do Bú EA, Freitas G, Pereira CR. Distributive justice criteria and social categorization processes predict healthcare allocation bias. *Br J Health Psychol*. 2023;28(2):552-566.

   doi: 10.1111/bjhp.12640

9. Chapman EN, Kaatz A, Carnes M. Physicians and implicit bias: How doctors may unwittingly perpetuate health care disparities. *J Gen Intern Med*. 2013;28(11):1504-1510.

   doi: 10.1007/s11606-013-2441-1

10. Burgess DJ, van Ryn M, Dovidio J, Saha S. Reducing racial bias among health care providers: Lessons from social-cognitive psychology. *J Gen Intern Med*. 2007;22(6):882-887.

    doi: 10.1007/s11606-007-0160-1

11. van Ryn M, Saha S. Exploring unconscious bias in disparities research and medical education. *JAMA*. 2011;306(9): 995-996.

    doi: 10.1001/jama.2011.1275

12. Krosch AR, Tyler TR, Amodio DM. Race and recession: Effects of economic scarcity on racial discrimination. *J Pers Soc Psychol*. 2017;113(6):892-909.

    doi: 10.1037/pspi0000103

13. Riek BM, Mania EW, Gaertner SL. Intergroup threat and outgroup attitudes: A meta-analytic review. *Pers Soc Psychol Rev*. 2006;10(4):336-353.

    doi: 10.1207/s15327957pspr1004_4

14. Axt JR, Ebersole CR, Nosek BA. The Judgment Bias Task: A flexible method for assessing individual differences in social judgment bias. *J Exp Soc Psychol*. 2018;79:337-355.

    doi: 10.1016/j.jesp.2018.06.006

15. Green AR, Carney DR, Pallin DJ, *et al*. Implicit bias among physicians and its prediction of thrombolysis decisions for Black and White patients. *J Gen Intern Med*. 2007;22(9): 1231-1238.

    doi: 10.1007/s11606-007-0258-5

16. van Ryn M, Burke J. The effect of patient race and socio-economic status on physicians' perceptions of patients. *Soc Sci Med*. 2000;50(6):813-828.

    doi: 10.1016/S0277-9536(99)00338-X

17. Schulman KA, Berlin JA, Harless W, *et al*. The effect of race and sex on physicians' recommendations for cardiac catheterization. *N Engl J Med*. 1999;340(8):618-626.

    doi: 10.1056/NEJM199902253400806

18. Sabin JA, Greenwald AG. The influence of implicit bias on treatment recommendations for 4 common pediatric conditions: Pain, urinary tract infection, attention deficit hyperactivity disorder, and asthma. *Am J Public Health*. 2012;102(5):988-995.

    doi: 10.2105/AJPH.2011.300621

19. Hagiwara N, Dovidio JF, Stone J, Penner LA. Applied racial/ethnic healthcare disparities research using implicit measures. *Soc Cogn*. 2020;38(Suppl):68–97.

20. Acosta AM, Garg S, Pham H, *et al*. Racial and ethnic disparities in rates of COVID-19-associated hospitalization, intensive care unit admission, and in-hospital death in the United States from March 2020 to February 2021. *JAMA Netw Open*. 2021;4(10):e2130479.

    doi: 10.1001/jamanetworkopen.2021.30479

21. Rubens M, Ramamoorthy V, Saxena A, *et al*. Racial disparities in access to health care infrastructure across US counties. *JAMA Netw Open*. 2023;6(5):e2311313.

    doi: 10.1001/jamanetworkopen.2023.11313

22. Mohammed S, Matos J, Doutreligne M, Celi LA, Struja T. Racial disparities in invasive ICU treatments among septic patients: High-resolution electronic health records analysis from MIMIC-IV. *Yale J Biol Med*. 2023;96(3):293-312.

doi: 10.59249/WDJI8829

23. Singh M, Venkataramani A. *Capacity Strain and Racial Disparities in Hospital Mortality.* (*NBER Working Paper No. 30380*). Cambridge, MA: National Bureau of Economic Research; 2022. Available from: https://www.nber.org/system/files/working_papers/w30380/revisions/w30380.rev0.pdf [Last accessed on 2025 Jul 25].

24. Gopalan PD, Pershad S. Decision-making in ICU - a systematic review of factors considered important by ICU clinician decision makers with regard to ICU triage decisions. *J Crit Care*. 2019;50:99-110.

doi: 10.1016/j.jcrc.2018.11.027

25. Gopalan PD, De Vasconcellos K. Factors influencing the decisions to admit or refuse patients into a South African tertiary intensive care unit. *S Afr Med J*. 2019;109(9):645-651.

doi: 10.7196/samj.2019.v109i9.13678

26. Hamberg K. Gender bias in medicine. *Womens Health (Lond)*. 2008;4(3):237-243.

doi: 10.2217/17455057.4.3.237

27. Arber S, McKinlay J, Adams A, Marceau L, Link C, O'Donnell A. Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: A video-simulation experiment. *Soc Sci Med*. 2006;62(1):103-115.

doi: 10.1016/j.socscimed.2005.05.028

28. Maserejian NN, Link CL, Lutfey KL, Marceau LD, McKinlay JB. Disparities in physicians' interpretations of heart disease symptoms by patient gender: Results of a video vignette factorial experiment. *J Womens Health*. 2009;18(10):1661-1667.

doi: 10.1089/jwh.2008.10

29. Smedley BD, Stith AY, Nelson AR. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press; 2003.

30. Penner LA, Blair IV, Albrecht TL, Dovidio JF. Reducing racial health care disparities: A social psychological analysis. *Policy Insights Behav Brain Sci*. 2014;1(1):204-212.

doi: 10.1177/2372732214548430

31. Dovidio JF, Gaertner SL. Aversive racism. *Adv Exp Soc Psychol*. 2004;36:1-52.

doi: 10.1016/S0065-2601(04)36001-6

32. Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Med Decis Making*. 2015;35(4):539-557.

doi: 10.1177/0272989X14547740

33. Phillips-Wren G, Adya M. Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty. *J Decis Syst*. 2020;29(3):173-186.

doi: 10.1080/12460125.2020.1768680

34. Younas S, Khanum S. Examining the role of stress and team support in decision making under uncertainty and time pressure. *MDM Policy Pract*. 2024;9(2):1–11.

doi: 10.1177/23814683241273575

35. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*. 2016;532(7600):453-458.

doi: 10.1038/nature17637

36. Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis-connecting the branches of systems neuroscience. *Front Syst Neurosci*. 2008;2:4.

doi: 10.3389/neuro.06.004.2008

37. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang*. 2008;59(4):390-412.

doi: 10.1016/j.jml.2007.12.005

38. Qin J, Liu C, Cheng S, Guo Y, Arcucci R. Freeze the backbones: A parameter-efficient contrastive approach to robust medical vision-language pre-training. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing/(ICASSP)*; 2024.

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

Task-related handwriting and drawing features for early detection of Alzheimer's disease: A pilot study

Maria Santina Ler[1]*, Miriam Veneziano[1†], Alfonsina D'Iorio[1†], Gennaro Cordasco[2†], Gabriella Santangelo[1†], and Anna Esposito[1†]*

[1]Department of Psychology, Faculty of Psychology, University of Campania Luigi Vanvitelli, Caserta, Italy

[2]Department of Computer Science, Faculty of Computer Science, University of Salerno, Salerno, Italy

Abstract

Background: Dementia causes significant disability worldwide and has no cure. The only way to improve the quality of life of those affected is through early intervention. For this reason, the development of effective diagnostic tools is a priority for healthcare systems and researchers. Handwriting and drawing, which engage multiple cognitive and motor areas, have shown promise in detecting early signs of dementia. However, findings in this field remain inconsistent, largely due to a lack of standardized protocols. **Aim:** This study aims to investigate the discriminatory power of graphomotor analysis in distinguishing individuals with Alzheimer's disease (AD) from healthy controls (HC) by examining the contribution of dynamic handwriting features and task-related characteristics within an easy-to-use and multi-task protocol. **Methods:** Patients with AD (n = 14) and HC (n = 25) were asked to complete five drawing and two writing tasks, and their online data were recorded using a digital tablet. **Results:** Significant differences (p<0.05) between groups were observed for time- and ductus-related features in almost all tasks, while pressure, space, and inclination features did not differ significantly. **Conclusion:** Although certain graphomotor characteristics are more sensitive than others, analyzing them together yields a detailed functional profile of patients. Overall, the study provides evidence of the effectiveness of handwriting analysis in identifying several symptoms associated with dementia. The protocol warrants further validation with a larger sample. **Relevance for patients:** The proposed protocol highlights the potential of a handwriting-based tool as an ecologically valid, objective, and accessible method for assessing and monitoring dementia. Adopting up-to-date digital approaches responds to the need for more sensitive tools that align with technological and cultural changes within the population. This could consequently simplify screening, improve access to treatment, and enhance the quality of life for patients and their caregivers.

Keywords: Dementia screening; Alzheimer's disease; Handwriting analysis; Online feature; Kinematic parameters

†These authors contributed equally to this work.

***Corresponding authors:**
Maria Santina Ler (mariasantina.ler@unicampania.it)
Anna Esposito (anna.esposito@unicampania.it)

Citation: Ler MS, Veneziano M, D'Iorio A, Cordasco G, Santangelo G, Esposito A. Task-related handwriting and drawing features for early detection of Alzheimer's disease: A pilot study. *J Clin Transl Res.* 2025;11(6):50-63. doi: 10.36922/JCTR025380063

Received: September 18, 2025

Revised: October 30, 2025

Accepted: November 19, 2025

Published online: December 4, 2025

1. Introduction

The growing prevalence of dementia worldwide poses a significant public health challenge, compounded by the economic and social strain it places on healthcare systems. Despite its profound impact, dementia diagnoses are often delayed due to a limited understanding of early markers. Recently, dynamic handwriting analysis has become increasingly relevant for early diagnosis of neurocognitive disorders because it is a non-invasive, economical, and reliable method of detecting dementia, especially Alzheimer's disease (AD).[1] Writing is a complex skill that requires the combined integration of executive, visuo-spatial, linguistic, and fine motor functions. Despite its potential, the lack of standardized and shared protocols hinders the systematic adoption of this approach in clinical practice. An important insight drawn from the literature is the substantial heterogeneity of the graphomotor characteristics selected for analysis, in contrast to an increasing convergence in the types of tasks proposed. The most frequent tasks can be categorized into three primary categories:

(i) Drawing simple geometric shapes, such as straight lines, circles, or spirals, is commonly used to assess fine motor skills and visuospatial coordination.[2-4]

(ii) Writing tasks which include various exercises, such as letter repetition (involving ascending and descending characters),[2,5,6] word writing (including irregular or non-sense words with distinct phonological and orthographic features),[7,8] and simple sentence writing. These tasks are proposed in different modalities, including copying, dictation, and spontaneous production.

(iii) Complex productions include the clock-drawing test (CDT), writing articulated sentences or paragraphs, performing backward tasks, and engaging in dual-task conditions.[3,4,9] These activities require the integration of a wider range of cognitive functions, including planning, working memory, attention, language, and visuospatial organization.

Although tasks are frequently categorized into broad groups, such as writing, drawing, and complex productions, their quantity and variability remain high, and the correlation among these tasks and discriminative graphomotor characteristics remains unclear. A noteworthy contribution to the digital writing domain is the multi-task protocol developed by Cilia *et al.*[10] for the DARWIN dataset,[11] which comprises 25 graphical and writing tasks of increasing complexity. They employed a task-informed machine learning approach that combines general and specific graphomotor features of individual writing tasks to enhance classification accuracy. Nevertheless, its extension has led numerous researchers to select specific components extracted from the whole protocol, tailoring them to their temporal and contextual requirements. Previous studies have shown that kinematic characteristics of handwriting are effective in identifying cognitive and mood disorders.[12-15] In particular, the pressure exerted on the paper and the time taken with the pen in flight have proven highly discriminatory, resulting in slower, less stable movements, and variable pressure during task performance in patients with AD. It is noteworthy that in patients with mild cognitive impairment (MCI), kinematic parameters, such as flight time, were particularly sensitive to cognitive status.[16-18] Nevertheless, the specific contribution of each graphomotor feature to task performance remains unclear, and further research is required to determine which parameters are most informative and the conditions under which they offer the greatest discriminatory power. Notably, people with AD have been observed to move more slowly and irregularly when performing complex activities, such as aiming tasks or curvilinear movements like circle drawing; however, these symptoms are less pronounced when simpler tasks, such as drawing straight lines or looped letters, are performed.[2] Longer execution times and slower writing speeds have also been observed in the early stages of the disease, particularly when writing in block letters.[16] Plonka *et al.*[8] investigated the relationship between cognitive demands and handwriting kinematics in different clinical populations, discovering that subjects with AD exhibited higher mean pressure in non-linguistic cognitive tasks, whereas subjects with primary progressive aphasia showed increased pressure in linguistic tasks, indicating a selective interaction between the cognitive domain elicited and the motor response. These findings underscore the importance of considering the task's impact on outcomes when examining graphomotor characteristics in writing and drawing, such as cognitive effort and the nature of movements performed.[17,18] In this regard, Garrè-Olmo *et al.*[17] used discriminant analysis to categorize participants based on their cognitive functioning as determined by graphomotor features. They observed that classification accuracy varied across task types and group comparisons, supporting the hypothesis that kinematic parameters were less informative when considered in isolation. In a multimodal integration study, Yamada *et al.*[19] analyzed graphomotor, vocal, and gait signals in individuals with cognitive decline, discovering that classification accuracy was the highest when combining the three behavioral modalities, with each modality outlining a distinct profile: graphical analyses revealed significant increases in pressure variability, decreases in speed, and prolonged pauses in the AD group during the writing process. These alterations were particularly pronounced during complex tasks, such as the Trail Making Test, and were also observed in MCI patients, highlighting the potential of complex tasks for early detection of dementia. All these

evidences[8,13-19] suggest that handwriting-related tasks are a promising tool for early dementia diagnosis. However, further investigation is needed, as the heterogeneity of protocols and outcomes limits the generalizability of results. Although specific graphomotor characteristics and task types have been identified, the most informative potential lies in their interactions, a topic that remains unexplored in the literature. Furthermore, many recent studies focus exclusively on classification models. Despite their ability to recognize complex patterns, these models have significant limitations, especially in interpretability, which can lead to the neglect of crucial information about the mechanisms of the graphomotor system and the underlying cognitive processes, both of which are clinically relevant. For this reason, a pilot study was conducted to evaluate the discriminative power of dynamic handwriting and drawing data in distinguishing between subjects diagnosed with probable AD and healthy controls (HCs). The main objective was to test a short-term screening method based on a combined writing and drawing protocol that could collect a wide range of information useful for the early identification of dementia. Specifically, we hypothesized that certain graphomotor features would be more sensitive to specific tasks. Therefore, the study examines the contribution of handwriting and drawing characteristics to the tasks performed, assessing whether their discriminatory capacity is task-specific or reflects more generalizable properties. These initial analyses aim to establish an empirical foundation for selecting tasks and key parameters for a future digital cognitive screening tool. By refining these tasks and identifying critical parameters, we hope to improve the accuracy and reliability of early dementia detection. Ultimately, this could lead to better outcomes through timely intervention and personalized support for those at risk.

In this regard, in line with the three-level functional structure (graphic, written, and complex tasks), this study proposes an independent and original protocol designed for future clinical application, which has been successfully validated in the Emotion Recognition from Handwriting and Drawing Project, where it was used to detect emotional states such as anxiety, depression, and stress through handwriting analysis on digital tablets by a sample of 129 subjects.[12] The current objective is to provide a preliminary validation of this protocol in detecting cognitive impairment. To this end, we explored the discriminatory power of online handwriting analysis in distinguishing between AD and HC subjects. Moreover, we investigated the specific contributions of individual tasks and related writing features to assess whether these features were task-dependent or task-independent. Although this is a pilot study with a small sample size and further large-scale validation is needed, our goal is to assess whether this approach shows potential as a feasible tool for future clinical applications.

2. Materials and methods

2.1. Selection criteria and participant recruitment

This pilot study recruited 39 participants (Participants were recruited between September 2024 and March 2025): 14 with probable AD (10 females and 4 males; mean age = 76; standard deviation [SD] = 6.42) and 25 healthy subjects (15 females and 10 males; mean age = 69.12; SD = 8.53). The sample size for this pilot study was determined through an a priori power analysis conducted using G*Power 3 software. Considering a comparison between two groups (clinical versus control) and the simultaneous analysis of an average of four dependent variables, it was hypothesized that the effect size would be medium to large, corresponding to an $f^2(V)$ value of 0.37. With a significance level of $\alpha = 0.05$ and a desired statistical power of 80% ($\beta = 0.20$), the minimum total sample size required was approximately 38 participants. The sample size adopted is consistent with that of comparable studies in the literature and accounts for the challenges of recruiting clinical subjects from sensitive populations. It is also proportionate to the exploratory and preliminary nature of the study.

All participants were right-handed and of Italian nationality. AD patients were recruited from the Memory Clinic at the University of Luigi Vanvitelli in Caserta. They were selected based on clinical and cognitive criteria consistent with mild-to-moderate disease, as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition[20] and the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association criteria.[21] A multidisciplinary team of healthcare experts made the diagnosis based on clinical assessment, caregiver interviews, and standardized cognitive tests. The assessment covered a wide range of cognitive domains, including executive functions, short-term and long-term memory, language, and attention. Aspects of affective emotional functioning were also considered, with particular attention paid to the presence of depression, apathy, and mood swings, using validated tools and clinical observations. Participants' functional state and level of autonomy were evaluated by collecting data on activities of daily living and instrumental activities of daily living, supplemented by caregiver information from the Caregiver's Inventory for Neuropsychological Diagnosis of Dementia.[22] Where possible, the clinical evaluation was supported by blood tests and neuroimaging. No time criteria were applied for

disease duration; however, participants taking psychotropic medications, including antidepressants or other treatments known to alter cognitive or motor performance, were excluded. Additional exclusion criteria included a known diagnosis of motor disorders (including Parkinson's disease, Parkinsonism, and other conditions that could affect writing ability) and severe cognitive impairment, as determined by specialists. Cognitively, HCs were recruited at the Memory Clinic and through local communities (e.g., associations). They reported no history of neurological or psychiatric disorders, or sensory or motor impairments. Cognitive profiles were assessed using the Mini-Mental State Examination (MMSE)[23] and the Montreal Cognitive Assessment (MoCA).[24] Only subjects with MMSE scores ≥24 and MoCA scores ≥23 were included.[25-27] Table 1 shows the summary of cognitive test scores, age, education, and other demographic variables in the two groups. The experiment adhered to the ethical principles of privacy and confidentiality and received approval from the Ethics Committee of the University of Campania Luigi Vanvitelli, under protocol number 16/2024. All participants received detailed information about the study and provided written informed consent. They received a copy of the consent form. For the clinical group, caregivers also reviewed and confirmed their comprehension of the study, its methods, and its purposes.

2.2. Data acquisition and apparatus

The experimental environment was designed to enhance the ecological validity of the interaction, maintaining the natural characteristics of the graphic gesture as much as possible, even in a digital acquisition setting. For this purpose, a digital tablet was utilized, equipped with an ink-stamped pen that allowed writing on traditional A4 sheets of paper superimposed on the active surface. This approach provided participants with consistent, familiar visual feedback, thereby mitigating the artificial effects of digitization and fostering fluid and spontaneous writing.

Table 1. Demographics and psychometrics of participants (*n*=39)

Demographics and psychometrics	Healthy control	Alzheimer's disease
Number of participants	25	14
Age (mean and SD)	69.12 (±8.53)	76 (±6.42)
Gender (female: male)	15: 10	10: 4
Years of education (mean)	11.6	7.29
MoCA (mean)	25.54	18.21
MMSE (mean)	27.40	21.07

Abbreviations: MMSE: Mini-mental state examination; MoCA: Montreal cognitive assessment; SD: Standard deviation.

The tablet used was the Intuos 4 (Wacom Co., Ltd., Japan), featuring a 224 × 148 mm active surface and a 5,080 lines per inch resolution. The pen had 8192 levels of pressure sensitivity and up to 60 levels of tilt detection, allowing for the precise capture of variations in stroke kinematics. The data were sampled at 8-millisecond intervals, recording real-time coordinates (x, y), pressure, tilt, azimuth angle, and pen-up/down states. The collected data was subsequently archived in CSV format for processing. The tablet was connected to a laptop, allowing the researcher to view the traces made by the participants in real time. This ensured continuous monitoring and facilitated prompt intervention as necessary.

2.3. Tasks and features

Participants were required to complete seven tasks, including two writing tasks and five drawing activities. These tasks were specifically selected from a wide range of already validated neuropsychological tests used to evaluate psychological and cognitive skills. Figure 1 illustrates the battery of tasks selected for the study, with the drawing tasks presented in the upper section and the writing tasks in the lower section. This protocol was meticulously designed to encompass a broad spectrum of executive complexity, ranging from basic geometric tracing to tasks requiring high-order cognitive functions, such as planning, semantic memory, and phonological integration. The proposed tasks include the following:

(i) Pentagon (Task 1): Copying two intersecting pentagons, taken from the MMSE,[23] is a classic visuo-constructive task requiring eye-hand coordination and spatial integration.

(ii) House (Task 2): Copying a three-dimensional house, inspired by personality projective tests, which combines the reproduction of horizontal, vertical, and oblique straight lines.

(iii) Word (Task 3): Writing four Italian words in capital letters; "*BIODEGRADABILE*," "*FLIPSTIM*," "*CHIUNQUE*," and "*SMINUZZAVANO*" (English translation: "biodegradable," "flipstrim," "anyone," and "they chop up"), selected to include regular, irregular and non-sense word, according to established paradigms in the literature on orthographic and phonological processing.

(iv) Circles (Tasks 4 and 5): Drawing circular loops first with the non-dominant hand and then with the dominant hand. This is a simple task useful for assessing motor fluency, control of curvilinear movements, and differences related to hemispheric dominance.

(v) Clock (Task 6): The CDT was adapted from the MoCA.[24] It represents a complex task requiring

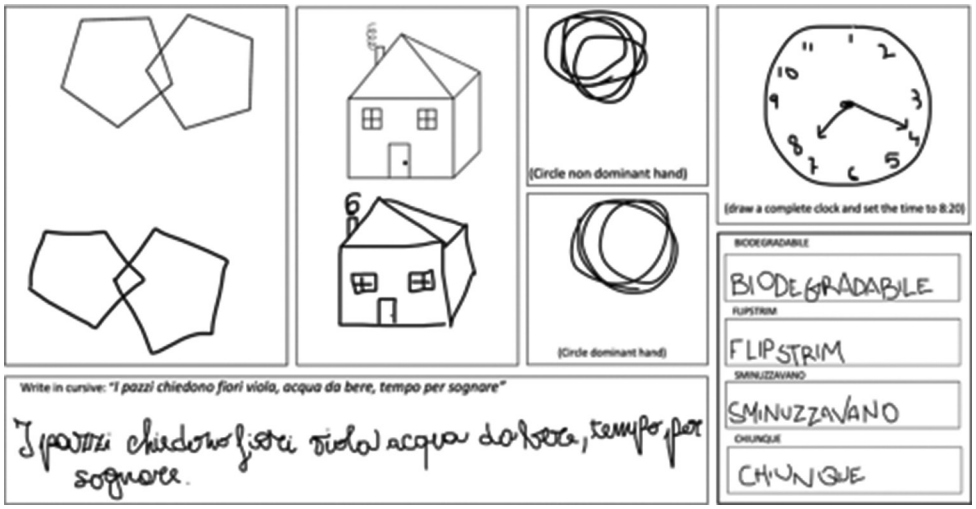


Figure 1. Illustration of the two task types. Drawing tasks are displayed on the top (pentagon, house, circle, and clock drawing) while writing tasks are shown on the bottom (word and sentence copy).

sequential planning, symbolic representation, and visuospatial integration.

(vi) Sentence (Task 7): Cursive copying of a phonologically complete Italian sentence: "*I pazzi chiedono fiori viola, acqua da bere, tempo per sognare*" (translated into English as "Crazy people ask for purple flowers, water to drink, and time to dream"), an exercise that simultaneously activates fine motor skills, sustained attention, linguistic processes, and working memory.

This heterogeneous battery enabled a sensitive examination of multiple motor, cognitive, and linguistic domains commonly affected by cognitive decline, thereby supporting a multi-level analysis of the graphomotor process. At the end of the experimental session, the raw data collected for each task were processed to extract a structured set of features. Specifically, 17 features were derived for each of the seven tasks, totaling 119 features, which were then grouped into five graphomotor categories: pressure, time, space, ductus, and inclination. Table 2 provides a detailed description of the features considered in the analysis, categorized by their type.

Table 2. Description of handwriting and drawing features according to their category

Category	Corresponding feature and description
Ductus	N_{up}: Length of movement performed in air
	N_{down}: Length of movement performed on paper
	N_{idle}: Strokes unrecognized by the tablet
Time	T_{up}: Time spent with the pen in the air
	T_{down}: Time spent with the pen on paper
	T_{idle}: Duration of pen status unrecognized by tablet
	T_{total}: Total time required to complete the task
Pressure	P_{min}: Minimum pressure value applied
	P_{max}: Maximum pressure value applied
	P_{avg}: Mean value of the applied pressure
	P_{sd}: Standard deviation of the applied pressure
	P_{10}: Lower 10th percentile of applied pressure
	P_{90}: Lower 90th percentile of applied pressure
Space and inclination	S_{bb}: Value derived from calculating the area of the smallest bounding box aligned to the axis that includes the stroke
	S_{avg}: Average length of the blank space between consecutive strokes
	S_{tot}: Total length of the blank space between consecutive strokes
	L_{avg}: Average inclination of the diagonals of the bounding boxes enclosing the strokes

2.4. Statistical analysis

Multivariate analysis of covariance (MANCOVA) was conducted using the Statistical Package for the Social Sciences (SPSS 21.0, IBM, United States) to evaluate the ability of graphomotor parameters to distinguish between individuals with cognitive impairment (AD) and HC subjects. Group affiliation (clinical versus control) was included as a fixed factor, while age and educational level were included as covariates to control for their potentially confounding effects on the dependent variables. The dependent variables were the graphic features extracted from each task and organized by graphomotor category (Table 2). A total of 28 separate MANCOVAs were performed, corresponding to an analysis of each of the four graphomotor categories (time, pressure, ductus, space, and inclination—the latter two analyzed jointly) for each of the seven proposed tasks. Each analysis included

all features belonging to the specific graphomotor category for that task as dependent variables. For example, the ductus-related features were analyzed in each of the seven tasks, and the same was performed for the other categories. Instead of conducting a single comprehensive analysis that included all 17 features simultaneously (17 features × 7tasks = 119 tests), we adopted an analytical design to limit the complexity of the statistical model, given the small number of participants. Such an analysis would have entailed a high risk of Type I error (false positives due to multiple comparisons) and a significant reduction in statistical power (Type II error), potentially compromising the reliability of the results and increasing the likelihood of misleading conclusions. Conversely, using 28 separate MANCOVAs enabled the simultaneous evaluation of the group's effect on multiple related variables, while accounting for interdependencies between features within each category and minimizing the risk of statistical errors. Where significant multivariate effects were present, *post hoc* univariate analyses of covariance were performed with Bonferroni correction for multiple comparisons to identify the features responsible for the differences between groups. Significance level was set at $p<0.05$. In the results sections, only the most relevant effects are reported. A complete, detailed view of the statistical results, including all *post hoc* analyses, is provided in the supplementary material (Table S1).

3. Results

3.1. Global effects of graphomotor categories across tasks

Multivariate analyses of covariance were used to estimate the overall effect of graphomotor categories (time, pressure, ductus, and space and inclination) on different groups for each task, while controlling for age and level of education. The results showed a significant difference between groups for time and ductus, whereas only one significant difference was observed for space and inclination. No statistically significant differences were found for pressure.

For time, the groups demonstrated a significant difference for Tasks 1 (F [4.32] = 3.675, p=0.014), 2 (F [4.32] = 9.335, p=0.000), 3 (F [4.32] = 6.559, p=0.001), 5 (F [3.33] = 5.181, p=0.005), 6 (F [4.32] = 3.274, p = 0.023), and 7 (F [4.32] = 2.926, p=0.036), indicating that participants with AD differed from the HC group in both drawing and writing tasks. In addition, a significant effect of age was observed in Task 6 (p=0.019), suggesting that age may influence execution times in clock drawing.

Regarding ductus, significant difference between groups was observed for Tasks 1 (F [3.32] = 3.955, p=0.017), 2 (F [3.33] = 3.921, p=0.017), 5 (F [2.34] = 7.606, p=0.002), 6 (F [3.33] = 3.348, p=0.031), and 7 (F [3.33] = 3.257, p=0.034), confirming graphical differences between groups in multiple writing and drawing conditions.

For space and inclination, a significant difference between groups was observed for Task 7 (F [4.32] = 2.691, p=0.049), whereas in Task 6, only the age covariate was significant (p=0.041). This suggests that spatial variations are more closely related to age than to clinical condition during CDT.

For pressure, no significant difference was observed between the groups.

Overall, the analyses indicate significant differentiation between groups in temporal and ductus measures, with age influencing clock drawing to some extent. Detailed statistical values for each graphomotor category and task are presented in Table 3.

3.2. Univariate analysis

Univariate analyses revealed several significant differences between the AD and HC groups across various writing features and tasks. While the space and inclination category showed an overall effect in Task 6 in the MANCOVA, this was not confirmed by the univariate analyses. Instead, statistically significant differences between the two groups emerged in the time and ductus categories. The results of the univariate analyses are reported in Table 4, which specifies the statistical significance, effect size, and mean values with SDs for each significant feature in the two groups (AD and HC). A complete overview of all results, including those that were not significant, is presented in the supplementary material (Table S1).

For the ductus category, the strokes unrecognized by the tablet (N_{idle}) feature were significant in all tasks, as indicated by the MANCOVAs. However, the N_{up} (length of movement performed in the air) and N_{down} (length of movement performed on paper) showed significant effects only in Tasks 5 and 7. Comparing the means revealed a homogeneous profile, with AD patients having higher scores than controls. Clinically, AD patients produced a greater number of strokes, both on paper and in the air, as well as more "empty" strokes. Notably, the N_{idle} feature represents movements that are not recognized by the tablet and can be interpreted as hand wandering, reflecting unstructured and non-purposeful motor activity. This phenomenon emerged consistently across tasks, regardless of whether they involved writing or drawing. During circle drawing with the dominant hand and cursive sentence writing, patients with AD exhibited not only increased N_{idle} but also a higher number of N_{up} and N_{down} strokes compared to HCs. This indicates an increased gesture segmentation, hesitation, and a loss of fluency in writing.

Table 3. Multivariate analysis of covariance results by group and covariates across all tasks and graphomotor categories

Category	Task	Wilks' Λ	*F* (df)	*p*	$\eta^2 p$	Age (*p*-value)	Education (*p*-value)
Ductus	1	0.729	3.955 (3.32)	0.017*	0.271	0.791	0.656
	2	0.737	3.921 (3.33)	0.017*	0.263	0.419	0.883
	3	0.898	1.243 (3.33)	0.310	0.102	0.201	0.851
	4	0.906	1.138 (3.33)	0.348	0.094	0.556	0.962
	5	0.691	7.606 (2.34)	0.002*	0.309	1.000	0.436
	6	0.767	3.348 (3.33)	0.031*	0.233	0.363	0.676
	7	0.772	3.257 (3.33)	0.034*	0.228	0.666	0.708
Time	1	0.685	3.675 (4.32)	0.014*	0.315	0.490	0.968
	2	0.461	9.335 (4.32)	0.000*	0.539	0.622	0.825
	3	0.549	6.559 (4.32)	0.001*	0.451	0.221	0.688
	4	0.796	2.054 (4.32)	0.110	0.204	0.234	0.914
	5	0.680	5.181 (3.33)	0.005*	0.320	0.693	0.370
	6	0.710	3.274 (4.32)	0.023*	0.290	0.019*	0.282
	7	0.732	2.926 (4.32)	0.036*	0.268	0.961	0.430
Pressure	1	0.737	1.783 (6.30)	0.136	0.263	0.626	0.235
	2	0.942	0.311 (6.30)	0.926	0.058	0.546	0.107
	3	0.761	1.567 (6.30)	0.191	0.239	0.181	0.114
	4	0.748	1.688 (6.30)	0.158	0.252	0.328	0.257
	5	0.785	1.370 (6.30)	0.259	0.215	0.755	0.755
	6	0.862	0.799 (6.30)	0.578	0.138	0.272	0.539
	7	0.845	0.920 (6.30)	0.494	0.155	0.844	0.617
Space and inclination	1	0.928	0.620 (4.32)	0.652	0.072	0.180	0.913
	2	0.867	1.225 (4.32)	0.320	0.133	0.447	0.710
	3	0.927	0.628 (4.32)	0.646	0.073	0.213	0.753
	4	0.835	1.579 (4.32)	0.204	0.165	0.322	0.245
	5	0.794	2.069 (4.32)	0.108	0.206	0.117	0.167
	6	0.896	0.930 (4.32)	0.459	0.104	0.041*	0.719
	7	0.748	2.691 (4.32)	0.049*	0.252	0.892	0.247

Notes: Task 1: Pentagon; Task 2: House; Task 3: Word; Tasks 4 and 5: Circles; Task 6: Clock; Task 7: Sentence. Statistical significance determined at $*p<0.05$.

Abbreviation: df: Degree of freedom.

For the time category, univariate analyses revealed significant group differences in almost all tasks. Notably, all temporal features were significant for Tasks 1, 2, and 3. No significant differences emerged for Task 4, aligning with the initial MANCOVA. In Tasks 5 and 7, all temporal features were significant, except for the time spent with the pen on paper (Tdown); in Task 6, only the duration of pen status unrecognized by the tablet (T_{idle}) was significant. Overall, a consistent pattern emerged from the mean scores, showing that the AD group exhibited higher values than the control group for all features and tasks, even when statistical significance was not achieved. A similar trend was also observed for the ductus category. These results suggest that patients with AD spend more time with the pen either in the air or in contact with the surface, thereby increasing total execution time. Furthermore, the T_{idle} parameter (i.e., the time not recognized by the system) was consistent with the increase in unrecognized strokes observed in the ductus category. Taken together, these two parameters suggest longer execution times and less fluid movement overall, which may indicate psychomotor slowing and reduced automation of graphomotor processes, even in simple tasks, such as drawing a circle with the dominant hand.

4. Discussion

The present study investigated a non-invasive method for distinguishing individuals with dementia from healthy subjects by integrating analyses of online drawing and writing tasks.

Table 4. Significant univariate results ($p<0.05$) for each feature and task, with group means and standard deviations

Category	Task	Feature	*F* (df)	*p*-value	η^2 *p*	AD mean±SD	HC mean±SD
Ductus	1	N_{idle}	12.091 (1.34)	0.001	0.262	14.890±2.201	4.622±1.545
	2	N_{idle}	11.627 (1.35)	0.002	0.249	16.241±2.348	5.505±1.648
	5	N_{up}	11.495 (1.35)	0.002	0.247	3.074±0.545	0.598±0.382
		N_{down}	11.495 (1.35)	0.002	0.247	4.074±0.545	1.598±0.382
		N_{idle}	7.181 (1.35)	0.011	0.170	1.925±0.480	0.202±0.337
	6	N_{idle}	7.597 (1.35)	0.009	0.178	16.350±2.615	6.684±1.836
	7	N_{up}	4.237 (1.35)	0.047	0.108	63.050±6.100	46.212±4.282
		N_{down}	3.957 (1.35)	0.050	0.102	63.629±6.186	47.128±4.342
		N_{idle}	5.430 (1.35)	0.026	0.134	17.090±2.963	7.830±2.080
Time	1	T_{up}	6.880 (1.35)	0.013	0.164	21,820.942±3,106.290	10,894.513±2,180.492
		T_{down}	5.503 (1.35)	0.025	0.136	17,132.331±2,126.358	10,443.255±1,492.619
		T_{idle}	6.527 (1.35)	0.015	0.157	19,021.266±4,492.789	3,628.291±3,153.760
		Total	13.832 (1.35)	0.001	0.283	57,976.077±6,618.315	24,966.757±4,645.794
	2	T_{up}	15.714 (1.35)	0.000	0.310	36,364.274±3,690.602	16,744.447±2,590.656
		T_{down}	7.112 (1.35)	0.012	0.169	24,180.190±2,769.910	14,273.774±1,944.367
		T_{idle}	11.534 (1.35)	0.002	0.248	12,366.581±2,192.672	2,380.035±1,539.168
		Total	20.558 (1.35)	0.000	0.370	72,917.563±6,499.348	33,398.245±4,562.284
	3	T_{up}	17.302 (1.35)	0.000	0.331	78,945.333±8,599.585	30,975.333±6,036.568
		T_{down}	12.929 (1.35)	0.001	0.270	37,686.454±3,296.460	21,791.026±2,313.984
		T_{idle}	5.849 (1.35)	0.021	0.143	32,822.920±6,340.780	12,257.445±4,450.976
		Total	16.966 (1.35)	0.000	0.326	149,465.100±15,286.424	65,025.264±10,730.463
	5	T_{up}	13.549 (1.35)	0.001	0.279	3,226.461±596.140	283.742±418.467
		T_{idle}	4.909 (1.35)	0.033	0.123	888.350±259.278	117.964±182.003
		Total	10.893 (1.35)	0.002	0.237	11,199.746±1,447.436	4,793.382±1,016.043
	6	T_{idle}	7.756 (1.35)	0.009	0.181	18,528.158±3,799.641	4,337.071±2,667.197
	7	T_{up}	11.746 (1.35)	0.002	0.251	42,545.297±5,243.194	18,446.794±3,680.514
		T_{idle}	6.197 (1.35)	0.018	0.150	13,352.760±2,628.629	4,577.254±1,845.194
		Total	12.131 (1.35)	0.001	0.257	81,485.894±8,358.319	42,445.139±5,867.208

Notes: Task 1: Pentagon; Task 2: House; Task 3: Word; Tasks 4 and 5: Circles; Task 6: Clock; Task 7: Sentence. Refer to Table 2 for a description of the features.
Abbreviations: AD: Alzheimer's disease; df: Degree of freedom; HC: Healthy control; SD: Standard deviation.

The results revealed clear differences between AD patients and controls in ductus- and temporal-related features. Regarding execution time, AD patients showed slower performance in both writing and drawing tasks, compared to controls. This slowdown is likely due to reduced processing speed and psychomotor retardation, which impair attention and action planning; these difficulties have been reported in the literature to be reflected in reduced autonomy in daily activities.[28] Particularly, pen-up time was highly effective at distinguishing between groups, even on simple tasks such as looping circles with the dominant hand. These observations confirm previous reports that in-air trajectories are an early indicator of cognitive impairment, which can distinguish between different stages of the disease.[8] Flight time may reflect a disorder in gesture planning, affecting the ability to decide when and how to initiate movement. Therefore, it is sensitive to tasks that require less complex integration between cognitive processes and motor control. Another consideration is the times observed in Task 6. Unlike the other tasks, no statistically significant differences in execution times were observed in the CDT, except for T_{idle}. The traditional CDT is highly sensitive to cognitive impairment but mainly assesses execution accuracy, i.e., the spatial arrangement of elements, the number of errors, and the overall quality of the final product. Therefore, time-related features may lack specificity for this task. Furthermore, the tasks in the present study involved imitating a model and provided clear guidance on what participants had to do. In contrast, CDT required

the participants to create the drawing from scratch without any prompts. This can create an "illusion of completion:" Patients may proceed with the drawing without fully understanding the necessary corrections, taking a similar amount of time to controls but producing a poor-quality result containing numerous errors. In other words, while execution times may be comparable, the difference lies in the accuracy and richness of detail, providing a coherent explanation for the observed statistical results. Regarding Task 4, the lack of statistical significance may be due to the effort required, given that it involves right-handed participants using their non-dominant left hand. This may cause even cognitively healthy participants to perform more slowly, thus attenuating the differences between groups. Therefore, any observed difficulty using the non-dominant hand may depend more on biomechanical and laterality factors than on specific cognitive deficits, which could explain the absence of statistical significance.

As for the ductus category, it was observed that AD patients tend to produce more strokes in writing and drawing tasks—both in the air and on paper—especially during sentence writing and in the circle loop with the dominant hand. Furthermore, they produced more unrecognizable strokes in most tasks. These results could arise from reduced motor control and altered spatial organization, leading to less precise, more fragmented writing, in line with findings observed for the temporal category.[29,30] Taken together, these two parameters help us to understand how hesitations and disfluencies in writing contribute to longer execution times. Conversely, such disfluencies and irregularities could result from a general psychomotor slowdown and impaired coordination.

Regarding pressure, as well as space and inclination, no statistically significant differences were found. This result can be interpreted as indicating that these parameters predominantly represent a biomechanical and automatic aspect of the graphic performance. These aspects may be less influenced by central cognitive processes and, therefore, could be more easily compensated for. Indeed, graphomotor skills are in part automated, meaning that parameters, such as pressure or stroke inclination, may remain relatively unaffected in the early and moderate stages of the disease. Conversely, kinematic and temporal characteristics—related to fluidity and execution speed—are highly susceptible to cognitive decline, since they could reflect higher cognitive functions, such as planning, executive control, and continuous movement monitoring.[31] This helps to explain why ductus features and execution times have greater discriminative power than other types of features, as supported by previous literature.[30-33] According to these studies, inclination does not appear to be a reliable or consistent parameter for distinguishing between normal and pathological aging. However, it could be effective in identifying mood alterations, such as symptoms of depression, where a more pronounced inclination could indicate reduced motor activation and cognitive effort.[34] In this context, inclination may have exploratory value and could be useful in differential diagnosis when considered in conjunction with other features. Regarding spatial aspects, it is important to recognize that spatial features reflect biomechanical, visuo-constructive, and spatial components.

In the present study, the absence of significant differences may be due to task performance being partially guided by a model to be copied, thereby reducing the cognitive demands placed on participants. Furthermore, using a bounding box likely provided external spatial cues, enabling participants to achieve a spatial distribution similar to that of the control group. In any case, the lack of statistical significance does not rule out genuine alterations; the small sample size in the AD group ($n = 14$) may have limited the analyses' statistical power. With a larger sample size, these measures may demonstrate a consistent trend with the visuospatial decline typically observed in AD. In addition, methodological differences across studies in the literature, including sample sizes and measurement tools, make direct comparison of results challenging and contribute to heterogeneity in reported data. Similarly, the same methodological considerations can be applied to pressure measurements, which are among the most controversial parameters in the literature. While our analyses do not contradict the results reported by Delazer *et al.*,[35] our data belong to a highly heterogeneous context in which empirical findings are often discordant and sometimes contradictory.[4,15] This variability can be attributed, at least in part, to differences in experimental protocols, measurement instruments, and sample selection criteria. These differences make direct comparisons between studies difficult and limit our ability to draw effective conclusions about the role of pressure in the analyzed graphomotor processes.

From a neurobiological perspective, although this study did not include biomarkers, literature suggests that alterations in fine motor skills are closely linked to underlying neurodegenerative processes, while the neural mechanisms remain poorly understood. Neuroimaging studies of motor function in patients with AD have revealed compensatory neural patterns during gross movement tasks (such as walking or gross coordination), while research into fine motor coordination remains limited.[36] Investigations into brain connectivity during manual dexterity tasks in individuals experiencing early stages of cognitive decline have not revealed significant differences compared to those without the condition, suggesting that

more subtle motor alterations emerge only in later stages of the disease.[37] The observed motor alterations in AD patients appear to result from an interaction between cognitive deficits and sensorimotor dysfunctions caused by neurodegenerative processes.[22,30] From this perspective, the observed slowing and graphomotor disfluency represent a loss of efficiency in the cognitive control mechanisms of movement, rather than a primary motor deficit. Understanding these differences is essential for improving our comprehension of the pathophysiological mechanisms of AD and for making differential diagnoses with other neurodegenerative diseases where movement disorders are predominant, such as Parkinson's disease.[5] Handwriting and drawing analysis may therefore represent a link between observable clinical manifestations and pathophysiological mechanisms, offering a multidimensional view of cognitive and functional decline. Moreover, graphomotor features are best interpreted in the context of the task, since certain motor patterns reflect the demands of the activity rather than specific deficits. This emphasizes the importance of considering the overall complexity of the graphomotor process rather than focusing on individual parameters in isolation.

Although several questions remain unanswered, the overall results confirm the existing literature on slowness and disfluency as distinctive features of AD patients.[35,38] Observed alterations, such as reduced execution speed, disorganized motor patterns, and an increased number of unrecognizable strokes, result in practical difficulties, including slower document completion, illegible handwriting, and impaired coordination of movements. These deficits directly affect autonomy in instrumental activities of daily living and overall quality of life, as they are often associated with reduced functional independence. Simple graphomotor tasks can therefore provide valuable insight into a patient's functional status. The proposed multi-task protocol is straightforward and fully non-invasive, with potential clinical applications, particularly for the early screening of dementia and the longitudinal monitoring of disease progression. Its use of short, easily replicable tasks enables frequent assessments in outpatient or home settings, supporting timely, patient-centered interventions to preserve autonomy, and quality of life. Furthermore, the quantitative and objective nature of the analysis minimizes human error, thereby enhancing the reliability and reproducibility of the results compared with those of traditional neuropsychological tests, which are based on subjective assessments and are susceptible to inter- and intra-examiner bias. In this context, the proposed protocol improves existing tools, such as the CDT, by offering a more detailed quantitative analysis. Continuous, objective measurements enable the detection of subtle variations that are often invisible in traditional assessments. This enhances sensitivity for long-term monitoring, supporting the early diagnosis of AD and the personalization of treatment pathways.

5. Limits and future direction

Several limitations should be considered when interpreting the results of this study. First, the small sample size ($n = 39$) and the imbalance between the AD and HC groups are significant methodological constraints. However, as this was a pilot study, the primary aim was to verify the feasibility of the experiment and to explore whether the proposed protocol could discriminate between the two groups by analyzing the relationships between different tasks and specific graphomotor features. The intention was not to draw definitive conclusions, but rather to identify potentially sensitive parameters and assess the consistency of the experimental paradigm with relevant clinical and neuropsychological assumptions. In this regard, the study advances our understanding of a field characterized by many uncertainties, including the neurodegenerative mechanisms underlying graphomotor skills, the specificity of graphomotor markers across different profiles of cognitive decline, and the relationship between higher cognitive processes and motor parameters in writing. Recruiting participants with neurodegenerative diseases poses logistical and organizational challenges due to strict inclusion criteria and limited access to vulnerable clinical groups. This justifies the use of a small but methodologically controlled sample, as is common in the literature. A similar issue arises with the age imbalance, with AD participants being older on average. Since age is a known confounding factor for motor slowing and cognitive decline, it may have influenced the results. While this was partially addressed by including age as a covariate in the MANCOVAs, future studies should aim to include larger, age- and education-matched samples to enhance statistical validity and generalizability. In addition, the homogeneity of the sample, comprising only Italian participants, prevented cross-cultural validation. Future studies should address the influence of cultural and educational factors on graphomotor and cognitive performance. Extending the protocol to different sociocultural contexts would enable the reproducibility of the results to be evaluated, as well as the impact of culture, language, and education on the generalizability of the findings.

In summary, the results of this study are a first step toward developing a complementary neuropsychological assessment tool based on graphomotor parameter analysis in writing and drawing tasks. However, as this is a pilot study with a small, unmatched sample, the results should be interpreted with caution. While the statistical analysis

controlled for confounding variables, such as age and education, future studies should validate these results using larger, more balanced populations that are more thoroughly characterized clinically. This may clarify some open methodological aspects, such as: (i) the stability of the parameters in relation to the method of acquisition (digital pen, tablet, or touchscreen); (ii) the sensitivity of the protocol in detecting early stages such as MCI; and (iii) the use of longitudinal approaches to observe the evolution of writing patterns over time and detect the progression of cognitive decline. Finally, future analyses may explore multivariate predictive models or classifiers that combine the most informative parameters to improve diagnostic accuracy and provide objective support for clinical practice. From this perspective, handwriting becomes a quantifiable window into cognitive functions and their impairment, rather than being observed only in qualitative terms.

6. Conclusion

This pilot study highlights the potential of a multi-task protocol, supported by dynamic graphic parameter analysis, as an innovative tool for early identification of neurodegenerative disease, particularly AD. Combining tasks and dynamic graphomotor features enables effective discrimination between clinical subjects and HCs, although a more in-depth evaluation is needed. Overall, the results suggest that time- and ductus-related features are particularly robust within the context of the proposed protocol. In contrast, no significant group differences were revealed for other features, such as pressure, as well as spatial and inclination parameters. This discrepancy does not imply that these features are uninformative; rather, it suggests that they vary in their discriminative power and likely reflect different underlying neurophysiological and cognitive processes. Clinically, patients with mild-to-moderate AD exhibited longer execution times and disfluencies with signs of uncertainty or hesitation in writing. This is consistent with a profile of psychomotor retardation and slower processing speed, which are typically associated with AD. As these graphomotor changes may appear before more evident cognitive symptoms, handwriting analysis could serve as a sensitive, non-invasive marker for early diagnosis. Furthermore, such measures could contribute to differential diagnosis by helping to distinguish AD from other neurodegenerative or psychiatric conditions that present with similar symptoms but different motor profiles. Longitudinal assessment of these parameters can also help clinicians track disease progression and evaluate treatment response. Finally, identifying specific motor and cognitive patterns enables the development of personalized rehabilitation strategies to preserve fine motor control and daily life autonomy.

From a technological perspective, identifying informative graphomotor parameters sensitive to cognitive decline is crucial for developing classification models and optimizing predictions. This paves the way for future digital developments and big data- and artificial intelligence-based applications. Technological progress and clinical relevance are closely linked, as the development of sensitive, accessible, affordable, and rapid screening tools is essential for developing person-centered, accessible medicine. As there is still no cure for AD, optimizing and streamlining diagnostic processes using cutting-edge tools is crucial to ensuring timely access to treatment and providing families with the right support. This is important because patients and carers often feel disorientated, confused, and socially isolated when facing the disease; receiving support to help them manage it can significantly improve their quality of life. While these objectives remain unattained, it is important to emphasize their prospective significance. These preliminary data provide a deeper understanding of the processes underlying handwriting in neurodegenerative diseases and offer the potential to integrate technological innovation into a clinical framework. While several questions remain unanswered, this research nonetheless represents a significant step toward developing dementia screening methods that combine scientific rigor, technological innovation, and clinical relevance.

Acknowledgments

We are grateful to the healthcare staff and research team of the Memory Clinic of L. Vanvitelli University in Caserta, Italy, for granting us access to the local patient community and for collecting the participants' clinical data. We acknowledge the valuable assistance of postdoctoral researchers Marialucia Cuciniello and Terry Amorese in improving the manuscript's English language.

Funding

This study received funding from the EU-H2020 program (grant no. 101182965, CRYSTAL), the EU NextGenerationE PNRR Mission 4 Component 2 Investment 1.1 (D.D 1409 del 14-09-2022) PRIN 2022 under the IRRESPECTIVE project (code P20222MYKE, CUP: B53D23025980001), and PNRR MUR under AI-PATTERNS FAIR Project (E63C22002150007).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Anna Esposito, Gennaro Cordasco
Formal analysis: Maria Santina Ler

Investigation: Maria Santina Ler, Miriam Veneziano, Alfonsina D'Iorio, Gabriella Santangelo
Methodology: Maria Santina Ler, Anna Esposito, Gennaro Cordasco
Writing–original draft: Maria Santina Ler
Writing–review & editing: Maria Santina Ler

Ethics approval and consent to participate

The experiment adhered to ethical principles of privacy and confidentiality and received approval from the Ethics Committee of the University of Campania Luigi Vanvitelli, under protocol number 16/2024. All participants received detailed information about the study and provided written informed consent to participate voluntarily. They received a copy of the consent form. For the clinical group, caregivers also reviewed and confirmed their comprehension of the study, its methods, and its purposes.

Consent for publication

All participants provided written consent for the publication of their data and its use for research and scientific purposes. To protect privacy, sensitive data were anonymized.

Availability of data

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

References

1. Smith AD. On the use of drawing tasks in neuropsychological assessment. *Neuropsychology*. 2009;23(2):231-239.

   doi: 10.1037/a0014184

2. Yu NY, Chang SH. Kinematic analyses of graphomotor functions in individuals with Alzheimer's disease and amnestic mild cognitive impairment. *J Med Biol Eng*. 2016;36:334-343.

   doi: 10.1007/s40846-016-0143-y

3. Ghaderyan S, Abbasi A, Saber S. A new algorithm for kinematic analysis of handwriting data; towards a reliable handwriting-based tool for early detection of Alzheimer's disease. *Expert Syst Appl*. 2018;114:428-440.

   doi: 10.1016/j.eswa.2018.07.052

4. Kairamkonda DD, Mandeleeka PS, Favaro A, *et al*. Analysis of interpretable handwriting features to evaluate motoric patterns in different neurodegenerative diseases. In: *2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. Philadelphia, PA, USA: IEEE; 2022. p. 1-6.

   doi: 10.1109/spmb55497.2022.10014966

5. Slavin MJ, Phillips JG, Bradshaw JL, Hall KA, Presnell I. Consistency of handwriting movements in dementia of the Alzheimer's type: A comparison with Huntington's and Parkinson's diseases. *J Int Neuropsychol Soc*. 1999;5:20-25.

   doi: 10.1017/s135561779951103x

6. Kahindo C, El-Yacoubi MA, Garcia-Salicetti S, Rigaud AS, Cristancho-Lacroix V. Characterizing early-stage Alzheimer through spatiotemporal dynamics of handwriting. *IEEE Signal Process Lett*. 2018;25(8):1136-1140.

   doi: 10.1109/lsp.2018.2794500

7. Impedovo D, Pirlo G, Barbuzzi D, Balestrucci A, Impedovo S. Handwritten processing for pre diagnosis of Alzheimer disease. In: *Proceedings of Biodevices 2014*; 2014. p. 193-199.

   doi: 10.5220/0004900701930199

8. Plonka A, Mouton A, Macoir J, *et al*. Primary progressive aphasia: Use of graphical markers for an early and differential diagnosis. *Brain Sci*. 2021;11(9):1198.

   doi: 10.3390/brainsci11091198

9. Bensalah A, Parziale A, De Gregorio G, Marcelli A, Fornés A, Lladós J. I can't believe it's not better: In-air movement for Alzheimer handwriting synthetic generation. In: *Proceedings of the International Graphonomics Conference*. Cham, Switzerland; 2023. p. 136-148.

   doi: 10.48550/arXiv.2312.05086

10. Cilia ND, De Stefano C, Fontanella F, Di Freca AS. An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis. *Procedia Comput Sci*. 2018;141:466-471.

    doi: 10.1016/j.procs.2018.10.141

11. Cilia ND, De Gregorio G, De Stefano C, Fontanella F, Marcelli A, Parziale A. Diagnosing Alzheimer's disease from on-line handwriting: A novel dataset and performance benchmarking. *Eng Appl Artif Intell*. 2022;111:104822.

    doi: 10.1016/j.engappai.2022.104822

12. Likforman-Sulem L, Esposito A, Faundez-Zanuy M, Clémençon S, Cordasco G. EMOTHAW: A novel database for emotional state recognition from handwriting and drawing. *IEEE Trans Hum-Mach Syst*. 2017;47(2):273-284.

    doi: 10.1109/thms.2016.2635441

13. Impedovo D, Pirlo G. Dynamic handwriting analysis for the assessment of neurodegenerative diseases: A pattern recognition perspective. *IEEE Rev Biomed Eng*. 2019;12:209-220.

    doi: 10.1109/rbme.2018.2840679

14. Schröter A, Mergl R, Bürger K, Hampel H, Möller HJ, Hegerl U. Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects. *Dement Geriatr Cogn Disord*. 2003;15(3):132-142.

    doi: 10.1159/000068484

15. Qi H, Zhang R, Wei Z, *et al*. A study of auxiliary screening for

Alzheimer's disease based on handwriting characteristics. *Front Aging Neurosci*. 2023;15:1117250.

doi: 10.3389/fnagi.2023.1117250

16. Kawa J, Bednorz A, Stępień P, Derejczyk J, Bugdol M. Spatial and dynamical handwriting analysis in mild cognitive impairment. *Comput Biol Med*. 2017;82:21-28.

doi: 10.1016/j.compbiomed.2017.01.004

17. Garre-Olmo J, Faúndez-Zanuy M, López-De-Ipiña K, Calvó-Perxas L, Turró-Garriga O. Kinematic and pressure features of handwriting and drawing: Preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. *Curr Alzheimer Res*. 2017;14(9):960-968.

doi: 10.2174/1567205014666170309120708

18. Müller S, Preische O, Heymann P, Elbing U, Laske C. Diagnostic value of a tablet-based drawing task for discrimination of patients in the early course of Alzheimer's disease from healthy individuals. *J Alzheimers Dis*. 2016;55(4):1463-1469.

doi: 10.3233/jad-160921

19. Yamada Y, Shinkawa K, Kobayashi M, *et al*. Combining multimodal behavioral data of gait, speech, and drawing for classification of Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis*. 2021;84(1):315-327.

doi: 10.3233/jad-210684

20. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*. 1984;34(7):939-944.

doi: 10.1212/wnl.34.7.939

21. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association; 2013.

22. Petersen RC, Jack CR Jr., Xu YC, *et al*. Memory and MRI-based hippocampal volumes in aging and AD. *Neurology*. 2000;54(3):581-587.

doi: 10.1212/wnl.54.3.581

23. Folstein MF, Robins LN, Helzer JE. The mini-mental state examination. *Arch Gen Psychiatry*. 1983;40(7):812.

doi: 10.1001/archpsyc.1983.01790060110016

24. Nasreddine ZS, Phillips NA, Bédirian V, *et al*. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.

doi: 10.1111/j.1532-5415.2005.53221.x

25. Conti S, Bonazzi S, Laiacona M, Masina M, Coralli MV. Montreal Cognitive Assessment (MoCA)-Italian version: Regression based norms and equivalent scores. *Neurol Sci*. 2015;36(2):209-214.

doi: 10.1007/s10072-014-1921-3

26. Carson N, Leach L, Murphy KJ. A re-examination of montreal cognitive assessment (MoCA) cutoff scores. *Int J Geriatr Psychiatry*. 2018;33(2):379-388.

doi: 10.1002/gps.4756

27. Measso G, Cavarzeran F, Zappala G, *et al*. The mini-mental state examination: Normative study of an Italian random sample. *Dev Neuropsychol*. 1993;9(2):77-85.

doi: 10.1080/87565649109540545

28. De Paula JJ, Albuquerque MR, Lage GM, Bicalho MA, Romano-Silva MA, Malloy-Diniz LF. Impairment of fine motor dexterity in mild cognitive impairment and Alzheimer's disease dementia: Association with activities of daily living. *Braz J Psychiatry*. 2016;38:235-238.

doi: 10.1590/1516-4446-2015-1874

29. Neils-Strunjas J, Groves-Wright K, Mashima P, Harnish S. Dysgraphia in Alzheimer's disease: A review for clinical and research purposes. *J Speech Lang Hear Res*. 2006;49(6):1313-1330.

doi: 10.1044/1092-4388(2006/094)

30. Yan JH, Rountree S, Massman P, Doody RS, Li H. Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *J Psychiatr Res*. 2008;42(14):1203-1212.

doi: 10.1016/j.jpsychires.2008.01.006

31. Guarino A, Favieri F, Boncompagni I, Agostini F, Cantone M, Casagrande M. Executive functions in Alzheimer disease: A systematic review. *Front Aging Neurosci*. 2019;10:437.

doi: 10.3389/fnagi.2018.00437

32. Fernandes CP, Montalvo G, Caligiuri M, Pertsinakis M, Guimarães J. Handwriting changes in Alzheimer's disease: A systematic review. *J Alzheimers Dis*. 2023;96(1):1-11.

doi: 10.3233/JAD-230438

33. Werner P, Rosenblum S, Bar-On G, Heinik J, Korczyn A. Handwriting process variables discriminating mild Alzheimer's disease and mild cognitive impairment. *J Gerontol B Psychol Sci Soc Sci*. 2006;61(4):P228-P236.

doi: 10.1093/geronb/61.4.p228

34. Greco C, Raimo G, Amorese T, *et al*. Discriminative power of handwriting and drawing features in depression. *Int J Neural Syst*. 2024;34(2):2350069.

doi: 10.1142/s0129065723500697

35. Delazer M, Zamarian L, Djamshidian A. Handwriting in Alzheimer's disease. *J Alzheimers Dis*. 2021;82(2):727-735.

doi: 10.3233/jad-210279

36. Koppelmans V, Silvester B, Duff K. Neural mechanisms of motor dysfunction in mild cognitive impairment and

Alzheimer's disease: A systematic review. *J Alzheimers Dis Rep.* 2022;6(1):307-344.

doi: 10.3233/adr-210065

37. Crockett RA, Hsu CL, Best JR, Beauchet O, Liu-Ambrose T. Head over heels but I forget why: Disruptive functional connectivity in older adult fallers with mild cognitive impairment. *Behav Brain Res*. 2019;376:112104.

doi: 10.1016/j.bbr.2019.112104

38. Kobayashi M, Yamada Y, Shinkawa K, Nemoto M, Nemoto K, Arai T. Automated early detection of Alzheimer's disease by capturing impairments in multiple cognitive domains with multiple drawing tasks. *J Alzheimers Dis*. 2022;88(3):1075-1089.

doi: 10.3233/jad-215714

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

Association between the neutrophil percentage-to-albumin ratio and breast cancer risk: Evidence from the National Health and Nutrition Examination Survey 1999–2016

Xiaoying He[1†], Tang Xiao[2†], Haifang Zhao[2], Xinyue Huang[2], Sheng Xu[1], Junfeng Liu[1], Yunxiang Wang[1], Jie Yin[2], Yong Zhou[1], Rui Qi[3], Ruijuan Heng[1*], Pan Qi[1*]

[1]Department of Head and Neck and Breast Surgery, Xinxiang Central Hospital, Xinxiang, Henan, China

[2]Department of Pharmacology, School of Pharmacy, Xinxiang Medical University, Xinxiang, Henan, China

[3]Department of Pharmacology, Zhengzhou Shuqing Medical College, Zhengzhou, Henan, China

[†]These authors contributed equally to this work.

***Corresponding authors:**
Ruijuan Heng
(15836035526@163.com)
Pan Qi
(18637377356@163.com)

Citation: He X, Xiao T, Zhao H, *et al*. Association between the neutrophil percentage-to-albumin ratio and breast cancer risk: Evidence from the National Health and Nutrition Examination Survey 1999–2016. *J Clin Transl Res*. 2025;11(6):64-75.
doi: 10.36922/JCTR025260030

Received: June 23, 2025

1st revised: August 1, 2025

2nd revised: August 19, 2025

3rd revised: September 6, 2025

Accepted: December 1, 2025

Published online: December 17, 2025

Abstract

Background: Increasing evidence underscores the association between chronic inflammation and carcinogenesis. **Aim:** The aim of the study is to investigate the association between the neutrophil percentage-to-albumin ratio (NPAR) and breast cancer (BC) risk. **Methods:** This analysis utilized cross-sectional data from 16,993 participants enrolled in the National Health and Nutrition Examination Survey from 1999 to 2016. To investigate the link between NPAR and BC risk, weighted multivariate logistic regression models were applied. Nonlinear associations were explored using restricted cubic spline (RCS) models. Furthermore, a nomogram incorporating NPAR was constructed for risk stratification, and its predictive accuracy was evaluated using receiver operating characteristic curve (ROC) analysis. **Results:** After adjusting for potential confounders, elevated NPAR levels showed a significant positive association with BC (odds ratio [OR] = 1.12, 95% confidence interval [95% CI]: 1.08–1.16; $p<0.001$). Compared with the reference quartile (Q1), progressively increasing ORs associated with BC were observed across ascending NPAR quartiles: Q1: OR 1.05 (95% CI: 0.78–1.42), Q2: OR 1.35 (95% CI: 1.02–1.79), and Q3: OR 1.47 (95% CI: 1.12–1.93). Analysis using RCS indicated a linear relationship where increasing NPAR levels were associated with a higher risk of BC. The predictive model incorporating NPAR demonstrated strong accuracy for BC prediction, achieving an area under the ROC curve of 0.8068 (95% CI: 0.7904–0.8232). **Conclusion:** Our findings reveal a significant dose-dependent association between NPAR and BC risk in a nationally representative sample from the United States. Although cross-sectional designs preclude causal inference, longitudinal studies should validate these observations and explore underlying biological mechanisms. **Relevance for patients:** NPAR could potentially serve as a novel biomarker to predict BC risk in patients.

Keywords: National Health and Nutrition Examination Survey; Neutrophil percentage-to-albumin ratio; Breast cancer; Inflammatory marker

1. Introduction

Breast cancer (BC), a major threat to women's health and the leading cause of cancer-related deaths in women globally, poses substantial public health challenges due to its high incidence and mortality rates.[1,2] Its development stems from intricate interplays among genetic, hormonal, and environmental, inflammatory, and metabolic factors.[3,4] Early and accurate risk stratification combined with prognostic assessment may facilitate personalized therapeutic strategies to enhance patients' clinical outcomes.[5-7]

Among the multifactorial contributors, the dual role of chronic inflammation and metabolic dysregulation in tumorigenesis has garnered increasing attention.[8,9] The functional states of immune cells within the tumor microenvironment, particularly neutrophils, are known to exert a profound impact on disease progression,[10-12] while nutritional status and metabolic homeostasis further modulate cancer-related outcomes.[13,14] These insights have spurred the development of composite biomarkers that better reflect the inflammatory–metabolic axis underlying carcinogenesis.

The neutrophil percentage-to-albumin ratio (NPAR) has emerged as a novel composite biomarker that integrates a proxy for systemic inflammation (neutrophil percentage) with a key indicator of nutritional and metabolic status (albumin).[15,16] This combination theoretically enables NPAR to capture aspects of both pathways simultaneously,[17,18] suggesting its utility as an indicator of inflammatory–nutritional dysregulation. However, despite this potential, evidence regarding the association between NPAR and BC risk remains limited. Its potential added value over established inflammatory markers in the context of BC is especially unclear and represents a significant knowledge gap, particularly given how frequently metabolic dysregulation coexists with and fuels chronic inflammation in this disease.

Leveraging the National Health and Nutrition Examination Survey (NHANES) cross-sectional data spanning 1999–2016, we examined the association between NPAR and BC risk among women in the United States. Our primary aim was to assess whether this association exists independently of potential confounders and explore the dose–response relationship. The findings from this investigation are intended to generate novel hypotheses regarding the role of inflammatory–nutritional dysregulation in BC and provide a foundation for future prospective studies.

2. Methods

2.1. Study design and data source

Conducted by the National Center for Health Statistics, the NHANES is a nationwide initiative designed to assess the health status and nutritional behaviors of noninstitutionalized civilians in the United States.[19] Through biennial data collection cycles, NHANES gathers information to identify contemporary disease patterns and trends, thereby informing public health policies. All NHANES data are publicly accessible at https://www.cdc.gov/nchs/nhanes/index.htm.

This study collected information from 92,062 individuals over nine successive NHANES periods (1999–2016). To define the final analytical cohort, several exclusion criteria were applied. First, male participants (n = 45,336) were excluded to focus on the at-risk female population. Second, individuals aged less than 20 years (n = 20,985) were removed, aligning the cohort with the typical age range for BC risk assessment. Third, exclusion was applied to pregnant individuals (n = 1,489) to avoid potential physiological confounders. Finally, a total of 7,259 participants were excluded due to missing essential data points, namely neutrophil count values or BC status, which are fundamental for the study's primary analyses. Following rigorous screening, 16,993 participants were included in this research.[20] Figure 1 illustrates the flow chart detailing participant selection.

2.2. Measurement of NPAR

Hematological parameters (red and white blood cell counts, hemoglobin, hematocrit, and red blood cell indices) were measured using the NHANES complete blood count protocol and a Beckman Coulter DxH 900 analyzer (Beckman-Coulter, United States). This instrument utilizes automated dilution/mixing for cell counts and sizing, as well as a single-beam photometer for hemoglobin measurement. White blood cell differentials, including lymphocytes and neutrophils, were determined using the Coulter VCS system (United States).[21]

Serum albumin concentration was measured using biochemical detection methods such as the bromocresol green method. Fasting venous blood was drawn from the subjects, and after serum separation, bromocresol green reagent was added to the serum. This reagent specifically binds to albumin to form a colored complex. The absorbance was measured using a spectrophotometer, and the serum albumin concentration was calculated based on the standard curve.[22] The NPAR was calculated using Equation (1):[23]

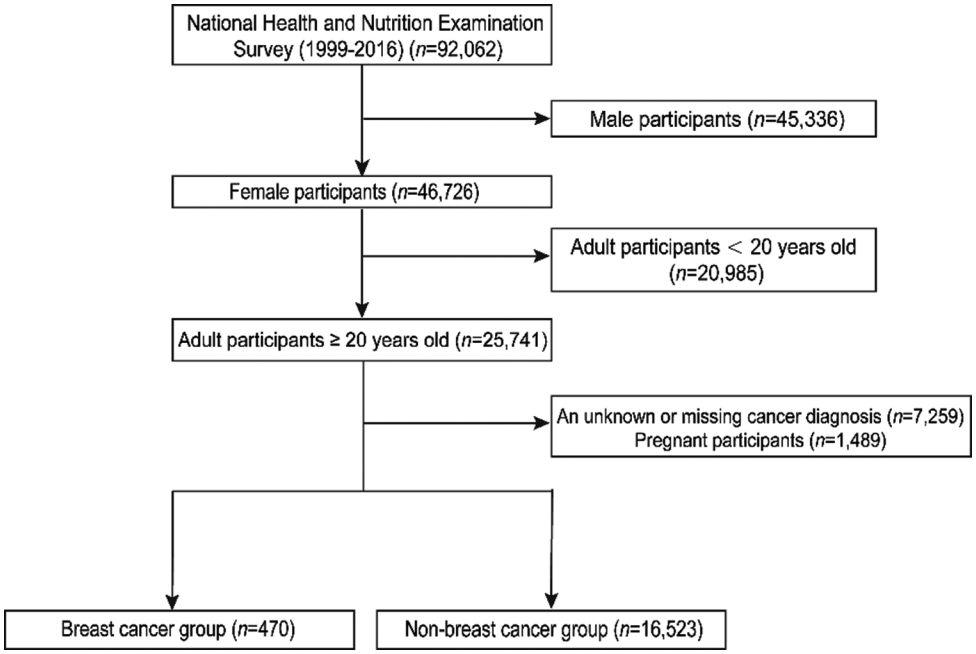


Figure 1. Comprehensive flowchart of participant recruitment. This diagram illustrates the stepwise exclusion process applied to derive the final analytic cohort for the study. Total initial female participants: 46,726. Exclusions comprised (i) participants aged <20 years (n = 20,985), (ii) those with missing/unknown cancer diagnosis (n = 7,259), and (iii) pregnant women (n = 1,489). The final analytical sample was 16,993 women, including 470 women with BC and 16,523 women without BC.
Abbreviation: BC: Breast cancer.

$$\text{Neutrophil percentage} \times 100/\text{Albumin (g/dL)} \quad (1)$$

This study classified participants into four groups according to their NPAR quartiles: <12.33, 12.33–13.92, 13.92–15.53, and ≥15.53, designating the lowest NPAR category (<12.33) as the control group.[24]

2.3. Definition of BC

BC cases were self-reported by participants in response to the question: "What kind of cancer was it?" Based on this, individuals were divided into two groups: those who reported BC and those who did not. The relevant data can be obtained from the publicly available datasets curated by the United States National Public Health Agency.[25]

2.4. Covariate selection

Drawing on prior studies and biological factors, our objective was to include a wide range of covariates known to confound BC outcomes. Key demographic characteristics collected included age, self-reported race and ethnicity, educational attainment, current marital status, history of tobacco use, and patterns of alcohol intake, as well as the family poverty–income ratio (PIR), which was collected using standardized questionnaires and face-to-face interviews. Per NHANES protocols, standardized physical exams were administered by licensed medical staff at mobile examination centers (MECs). Self-reported race/ethnicity included five categories: non-Hispanic White, Mexican American, non-Hispanic Black, other Hispanic, and other race. Education levels were grouped as less than high school (Grades <9), some high school (9–11), high school diploma/General Educational Development, some college, and college degree or higher. Marital status comprised married, divorced, widowed, separated, never married, or unmarried couples. Smoking status was defined by a lifetime consumption of ≥100 cigarettes, irrespective of current usage. Alcohol use was categorized based on a consumption of ≥12 standard alcoholic drinks during the 12 months before survey administration. Body mass index (BMI) was calculated as weight (kg)/height (m^2) and categorized according to the World Health Organization guidelines. A BMI of 25–29.9 kg/m^2 was considered overweight, and a BMI ≥30 kg/m^2 was considered obese. Laboratory tests were performed following standardized procedures to measure neutrophil counts. Given the important roles of hypertension and diabetes as factors that may affect BC, it was essential to account for their possible confounding effects. Physician-diagnosed hypertension and diabetes were defined based on positive responses to standardized NHAHES questions (e.g., "Have you ever been told by a doctor or health professional that you have hypertension, diabetes, or sugar diabetes?").[26]

2.5. Statistical analysis

Statistical significance was defined as $p<0.05$ for all analyses. Analyses were performed using R software (version 4.4.3), with the tidyverse and stats libraries employed. Continuous variables were assessed using Shapiro–Wilk normality testing. Normally distributed data are reported as mean ± standard deviation, while non-normal data are reported as median (interquartile range), and categorical variables are reported as frequency (percentage). Descriptive analyses were conducted for all variables.[27]

Normally distributed continuous variables were compared between groups using Student's *t*-test. Categorical variables were analyzed using the χ^2 test,[28] employing Fisher's exact test when expected cell counts fell below 5. Variance inflation factors (VIFs) were assessed before modeling to evaluate multicollinearity, with VIFs >5 indicating significant collinearity. We employed multivariable logistic regression to evaluate the association between NPAR and BC risk, reporting odds ratios (ORs) alongside their 95% confidence intervals (CIs). Three progressively adjusted models were constructed:

(i) Model 1: Unadjusted
(ii) Model 2: Included covariate adjustments for race/ethnicity and age
(iii) Model 3: In addition, adjusted for marital status, PIR, diabetes, smoking status, and hypertension.[29]

Group differences in survival distributions were evaluated employing the log-rank test. In addition, to flexibly model the association between NPAR and BC risk across its spectrum, we incorporated restricted cubic splines (RCS) into the logistic regression framework. Knot locations for the RCS were set at the 10th, 50th, and 90th percentiles of the NPAR distribution.[30]

Stratified analyses were conducted to examine whether the association between NPAR and BC differed across subgroups defined by age group (20–40, 40–60, ≥60 years), BMI category (<25, 25–30, >30 kg/m^2),[31] PIR level (<1, 1–3, ≥3), alcohol consumption (yes/no), smoking status (yes/no), diabetes (yes/no), and hypertension (yes/no). To investigate potential effect modification, we introduced interaction terms into the models to examine if the NPAR relationship with BC risk differed significantly across these strata.[32]

3. Results

3.1. Demographic attributes of participants

This study enrolled 16,993 female participants, a relatively large sample for similar epidemiological studies, providing a solid foundation for the stability and reliability of the analyses. Among all participants, 470 women reported a diagnosis of BC, accounting for 2.7% of the total number. This incidence proportion offers important population data support for exploring factors related to BC onset.

To more clearly describe the association between BC and various baseline characteristics, we conducted a stratified analysis of the baseline characteristics of the study population based on whether they had BC, with specific data as detailed in Table 1. Comparative analysis showed that there were significant statistical differences in multiple baseline characteristics between participants with BC and those without.

In terms of age, the average age of the BC patient group was significantly higher than that of the non-patient group. This result is consistent with findings from previous studies that age is an important risk factor for BC. This is because, as age increases, the cumulative risk of DNA damage in somatic cells increases, which may, to a certain extent, raise the likelihood of having BC.

Regarding ethnic composition, the proportion of non-Hispanic White women was significantly higher among those with BC than among other racial groups. This difference suggests that ethnic factors may play a role in BC onset, but the specific mechanisms (such as differences in genetic background and living habits) still need further research and verification.

In terms of weight-related indicators, the average BMI of women with BC was significantly lower than that of women without BC. It is worth noting that the association between BMI and BC has always been a research hotspot. Different studies may draw different conclusions due to differences in population characteristics and follow-up time. The association between lower BMI and BC observed in this study provides new population-level data for this field.

In terms of marital status, the proportion of unmarried individuals among BC patients was lower than controls. Some studies have shown that marital status may influence health outcomes through social support and lifestyle stability. Whether the observed differences are related to these mechanisms requires further exploration.

The economic status indicator showed that the PIR of women with BC was significantly higher than that of controls, indicating relatively lower economic pressure. This may be related to the fact that groups with high economic levels are more likely to access regular physical examinations and other health services, thus detecting diseases earlier. It may also involve other potential influencing factors, such as diet and environment.

Table 1. Participants' characteristics by BC status

Variables	Overall (*n*=16,993)	BC (*n*=470)	Non-BC (*n*=16,523)	*p*-value
Age				<0.001***
20–40 years	5,459 (32.12)	7 (1.49)	5,452 (33.00)	
40–60 years	5,730 (33.72)	99 (21.06)	5,631 (34.08)	
>60 years	5,804 (34.16)	364 (77.45)	5,440 (32.92)	
Race				<0.001***
Mexican American	2,918 (17.17)	42 (8.94)	2,876 (17.41)	
Non-Hispanic Black	3,357 (19.76)	71 (15.11)	3,286 (19.89)	
Non-Hispanic White	8,065 (47.46)	305 (64.89)	7,760 (46.96)	
Other races, including multi-racial	1,229 (7.23)	26 (5.53)	1,203 (7.28)	
Other Hispanic	1,424 (8.38)	26 (5.53)	1,398 (8.46)	
Education level				0.294
Less than 9th Grade	1,799 (10.59)	44 (9.36)	1,755 (10.62)	
9–11th Grade	2,436 (14.34)	63 (13.40)	2,373 (14.36)	
Some college or an Associate in Arts degree	5,271 (31.02)	136 (28.94)	5,135 (31.08)	
High school graduate/GED or equivalent	3,807 (22.40)	106 (22.55)	3,701 (22.40)	
College graduate or above	3,680 (21.66)	121 (25.74)	3,559 (21.54)	
Drinking				0.442
Yes	10,138 (59.66)	270 (57.45)	9,868 (59.72)	
No	6,855 (40.34)	200 (42.55)	6,655 (40.28)	
Hypertension				<0.001***
Yes	6,113 (35.97)	270 (57.45)	5,843 (35.36)	
No	10,880 (64.03)	200 (42.55)	10,680 (64.64)	
Diabetes				<0.001***
Yes	2,003 (11.79)	96 (20.43)	1,907 (11.54)	
No	14,990 (88.21)	374 (79.57)	14,616 (88.46)	
Smoking				0.002**
Yes	6,467 (38.06)	208 (44.26)	6,259 (37.88)	
No	10,526 (61.94)	262 (55.74)	10,264 (62.12)	
NPAR				<0.001***
1	4,242 (24.96)	88 (18.72)	4,154 (25.14)	
2	4,252 (25.02)	93 (19.79)	4,159 (25.17)	
3	4,250 (25.01)	131 (27.87)	4,119 (24.93)	
4	4,249 (25.00)	158 (33.62)	4,091 (24.76)	
BMI				0.765
<25	5,343 (31.44)	152 (32.34)	5,191 (31.42)	
25–30	4,941 (29.08)	140 (29.79)	4,801 (29.06)	
>30	6,709 (39.48)	178 (37.87)	6,531 (39.53)	
Household poverty ratio				0.003**
<1	3,588 (21.11)	71 (15.11)	3,517 (21.29)	
1–3	7,168 (42.18)	202 (42.98)	6,966 (42.16)	
>3	6,237 (36.70)	197 (41.91)	6,040 (36.56)	

(Cont'd...)

Table 1. Participants' characteristics by BC status

Variables	Overall (*n*=16,993)	BC (*n*=470)	Non-BC (*n*=16,523)	*p*-value
Marital status				<0.001***
Married	8,165 (48.05)	232 (49.36)	7,933 (48.01)	
Widowed	2,172 (12.78)	130 (27.66)	2,042 (12.36)	
Divorced	2,169 (12.76)	62 (13.19)	2,107 (12.75)	
Separated	651 (3.83)	12 (2.55)	639 (3.87)	
Never married	2,745 (16.15)	24 (5.11)	2,721 (16.47)	
Living with a partner	1,091 (6.42)	10 (2.13)	1,081 (6.54)	

Notes: Data are presented as *n* (%), unless stated otherwise. **/*** indicates $p<0.01$, and $p<0.001$, respectively.
Abbreviations: BC: Breast cancer; BMI: Body mass index; GED: General Educational Development; NPAR: Neutrophil percentage-to-albumin ratio.

In terms of behavioral habits, the smoking rate of BC patients was significantly higher than that of non-patients, while their alcohol intake was lower. As a clear carcinogenic risk factor, smoking's association with BC onset is reflected in this study. There is a dose–effect difference in the association between alcohol intake and BC. The finding that the alcohol intake was lower in the BC group in this study provides supplementary evidence for related research.

In terms of chronic disease burden, the proportion of BC patients with diabetes and hypertension was higher than controls. These two chronic diseases and BC may share common risk factors (such as unhealthy lifestyles and inflammatory reactions). The mechanism of their interaction needs in-depth research.

3.2. Correlation between the NPAR and BC risk

Before constructing the multivariable logistic model, collinearity between NPAR and the remaining covariates was assessed. All VIFs were below 5, suggesting no significant collinearity issues. Results from multivariable logistic regression models exploring NPAR's association with BC are detailed in Table 2. ORs and 95% CIs are provided for three models: an unadjusted model (Model 1), a model adjusted for race and age (Model 2), and a model additionally adjusted for hypertension, diabetes, smoking, marital status, and PIR (Model 3). After confirming the absence of problematic collinearity, we examined the link between NPAR and BC outcomes. Within the fully adjusted Model 3, elevated NPAR levels demonstrated a significant positive association with the odds of BC. Specifically, upon quartile stratification, participants in NPAR Q4 exhibited a 47% greater likelihood of BC compared to those in Q1 (adjusted OR = 1.47; 95% CI: 1.12–1.93; $p=0.006$).

3.3. Linearity assessment of the NPAR association with BC risk in patients

To characterize the dose–response relationship between NPAR and BC risk beyond simple linear assumptions, we employed RCS with three knots, placed at the 10th, 50th, and 90th percentiles of the NPAR distribution as per standard practice. This flexible modeling approach revealed a statistically significant overall association ($p<0.001$). However, the test for non-linearity yielded a non-significant result ($p=0.970$), providing strong statistical evidence against a complex curvilinear pattern and instead supporting a predominantly linear relationship across the observed range of NPAR values. This linear trend is visually confirmed in Figure 2, which demonstrates a consistent, monotonic increase in the adjusted OR for BC, even after accounting for potential confounders, including age, BMI, and smoking status, with progressively higher NPAR levels. Importantly, the entire range of estimated OR values at various NPAR points was statistically significant, as evidenced by their 95% CIs uniformly excluding the null value (OR = 1.0). The fact that the CIs exclude 1.0 indicates a robust positive association. Furthermore, the graded, increasing relationship between NPAR and BC risk observed in the RCS analysis is consistent with a potential dose–response pattern. This pattern likely reflects the continuous, underlying pathophysiological processes involving systemic inflammation (captured by the percentage of neutrophils), where neutrophil-derived pro-inflammatory cytokines may promote the formation of the tumor microenvironment and nutritional/metabolic status (reflected by albumin, a marker whose levels correlate with tissue repair capacity and metabolic homeostasis), which contribute to oncogenesis.

3.4. Subgroup analysis

Subgroup analysis included race, age categories, BMI, educational attainment, alcohol consumption patterns, smoking status, diabetes mellitus, and hypertension. The results, visually summarized in Figure 3, revealed that the positive association between elevated NPAR levels and increased BC risk was statistically more pronounced within specific subgroups. These subgroups encompassed

Table 2. Assessing the link between NPAR and BC using logistic regression

NPAR quartile	Model 1		Model 2		Model 3	
	OR (95% CI)	*p*	OR (95% CI)	*p*	OR (95% CI)	*p*
Q1	1.00 (Ref)		1.00 (Ref)		1.00 (Ref)	
Q2	1.06 (0.79–1.42)	0.719	1.07 (0.79–1.44)	0.668	1.05 (0.78–1.42)	0.735
Q3	1.50 (1.14–1.97)	0.004	1.38 (1.04–1.82)	0.026	1.35 (1.02–1.79)	0.036
Q4	1.82 (1.40–2.37)	<0.001	1.49 (1.13–1.95)	0.004	1.47 (1.12–1.93)	0.006

Abbreviations: BC: Breast cancer; CI: Confidence interval; NPAR: Neutrophil percentage-to-albumin ratio; OR: Odds ratio.

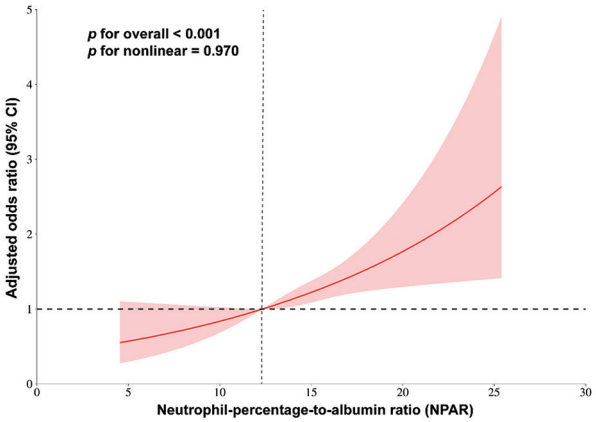


Figure 2. Restricted cubic splines curves of association between the NPAR and BC in all study participants. Solid lines: adjusted ORs (multivariate logistic regression); shaded bands: 95% CIs. Reference: NPAR=12.325 (median of 1st quartile; vertical dashed line). Horizontal line at OR=1: no association. Model adjusted for age, race/ethnicity, body mass index, education, marital status, smoking/alcohol history, household poverty rate (poverty–income ratio), hypertension, and diabetes. Junctions at the 10th (10.82%), 50th (13.92%), and 90th (17.09%) percentiles of the NPAR distribution. Non-linearity: p=0.970 (likelihood ratio test).
Abbreviations: BC: Breast cancer; CI: Confidence interval; NPAR: Neutrophil percentage-to-albumin ratio; OR: Odds ratio.

individuals identifying as belonging to racial groups classified as "Others," participants with a high school graduation or equivalent educational level, younger adults (20–40 years old), individuals classified as overweight, non-smokers, abstainers from alcohol, and individuals diagnosed with either diabetes or hypertension. Importantly, despite these observed variations in association strength across subgroups, formal statistical testing for interaction effects indicated no significant multiplicative interaction between the NPAR variable and any of the stratification variables examined.

The forest plot (Figure 3) illustrates the ORs and their 95% CIs for the association between the exposure of interest and the outcome, analyzed across different subgroups. An OR > 1 implies that the exposed group has higher odds of the outcome compared to the unexposed group, while an OR < 1 indicates lower odds. Subgroups were defined based on the following factors: (i) ethnicity, including Mexican-American, non-Hispanic, non-Hispanic Black, non-Hispanic White, and other races; (ii) education level such as below grade 9, grades 9–11, high school graduates, college graduates; (iii) alcohol consumption status (yes/no); (iv) diabetes status (yes/no); (v) hypertension status (yes/no); (vi) smoking status (yes/no); (vii) age group (20–40, 40–60, ≥60); BMI category (<25, 25–30, ≥30); PIR category (<1, 1–3, ≥3); NPAR quartiles (Q1–Q4); and marital status, such as married, widowed, divorced/separated, never married, and living with a partner.

3.5. Nomogram model for predicting BC based on participants' baseline characteristics

Using a retrospective analysis of a substantial cohort, we developed a predictive nomogram for BC risk, as illustrated in Figure 4. This tool integrates readily available baseline clinical and demographic characteristics with the NPAR score. To demonstrate its application, the nomogram was used to assess a representative case: a 69-year-old White woman with an NPAR score of 11.372. The model predicted that her odds of having BC were significantly elevated, approximately 2.86-fold higher than the reference profile. The nomogram's discriminative power, reflecting its ability to distinguish between individuals with and without BC, was rigorously quantified using receiver operating characteristic (ROC) curve analysis. This evaluation yielded an area under the curve (AUC) of 0.8068 (95% CI: 0.7904–0.8232). This high AUC signifies that the model has strong predictive capability, effectively stratifying BC risk based on the integrated assessment of clinical and inflammatory status.

4. Discussion

As a novel inflammatory–nutritional composite biomarker, NPAR integrates information from routine blood tests, reflecting the balance between pro-inflammatory neutrophils and nutrition-related albumin levels. This integration may capture aspects of systemic inflammatory and metabolic dysregulation more comprehensively than isolated markers.[33] Emerging research suggests

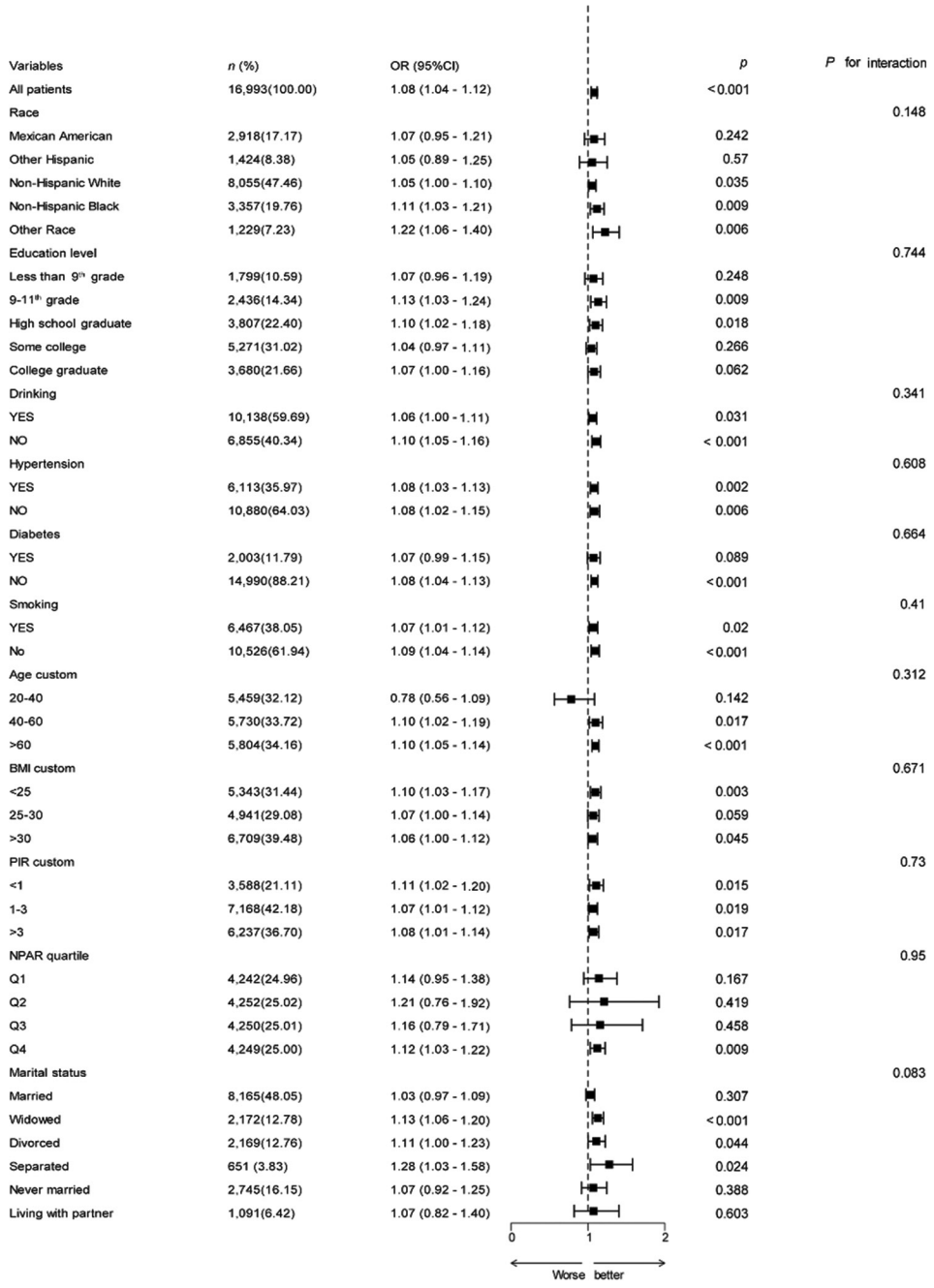


Figure 3. Subgroup analysis of the association between NPAR and BC. Each square represents the OR for a specific subgroup. The horizontal line extending from the square denotes the 95% CI.

Abbreviations: BC: Breast cancer; CI: Confidence interval; NPAR: Neutrophil percentage-to-albumin ratio; OR: Odds ratio.

that composite biomarkers, such as NPAR, may offer advantages over traditional inflammatory indices, like the neutrophil-to-lymphocyte ratio or the systemic immune-inflammation index, by integrating multifaceted pathological processes—specifically, the simultaneous pro-inflammatory component (neutrophil percentage) and nutritional–metabolic component (albumin) status.[34] Although direct comparative studies in BC are currently lacking, NPAR's promising diagnostic and prognostic utility is evidenced across a spectrum of chronic conditions. It achieved an AUC of 0.810 (95% CI: 0.794–0.825) in discriminating individuals with and without non-alcoholic fatty liver disease[35] and demonstrated strong discriminatory power for diabetic retinopathy risk stratification, as well as

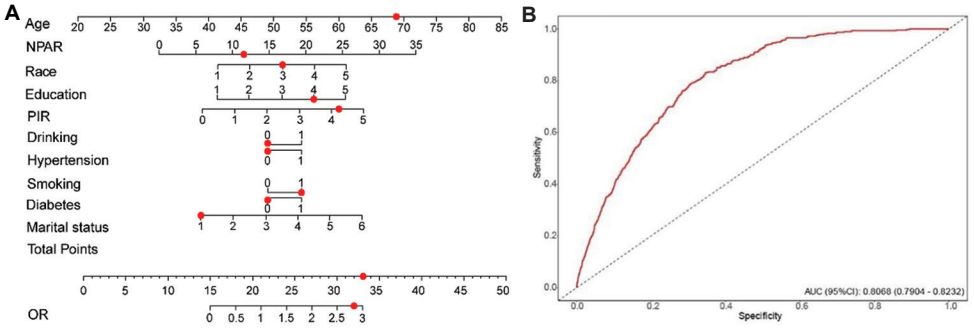


Figure 4. Development and evaluation of the BC risk prediction nomogram. (A) Nomogram for predicting the odds ratio (OR) of BC risk, integrating baseline clinical and demographic characteristics and the NPAR score. Red markers indicate reference values or illustrative examples, depending on the context. (B) Receiver operating characteristic (ROC) curve depicting the discriminative ability of the predictive model for BC risk. The red solid line represents the ROC curve of the model, illustrating the trade-off between sensitivity and specificity across different prediction thresholds. The dashed diagonal line serves as a reference for a random classifier, for which the area under the curve (AUC) would be 0.5 (no discriminative power).
Abbreviations: BC: Breast cancer; CI: Confidence interval; NPAR: Neutrophil percentage-to-albumin ratio; PIR: Poverty–income ratio.

for predicting mortality after hip fracture surgery in older adults.[36] It also showed favorable sensitivity/specificity in assessing severity in anti-N-methyl-D-aspartate receptor encephalitis.[37] This consistent performance across diverse disease contexts underscores NPAR's potential as a sensitive indicator of systemic inflammatory–metabolic–immune dysregulation. However, its applications in oncology, particularly in BC, remain nascent. The mechanistic underpinnings of NPAR in cancer pathogenesis and its clinical translatability necessitate systematic investigation, especially through direct comparisons with established indices within the same cohorts.

Using a large, nationally representative multi-stage stratified sample from NHANES, this study identified a significant positive association between higher NPAR levels and increased risk of BC among women in the United States. RCS analyses revealed a monotonically increasing trend in BC risk with rising NPAR levels (p<0.001), indicating a graded relationship consistent with a potential dose–response pattern, although the cross-sectional design precludes causal inference. This association remained robust across subgroups stratified by age and BMI, supporting the generalizability of the finding.

The observed association may be interpreted in the context of BC pathophysiology. Chronic inflammation is implicated in tumor initiation and progression; neutrophils, in particular, are known to facilitate invasion and metastasis through mechanisms including epithelial–mesenchymal transition and the formation of pre-metastatic niches.[38] Concurrently, low albumin levels may reflect not only malnutrition but also ongoing systemic inflammation, potentially contributing to immunosuppression within the tumor microenvironment.[39] Furthermore, breast tumors, often driven by estrogen signaling, may exhibit enhanced neutrophil recruitment through estrogen-mediated upregulation of chemokines such as chemokine (C-X-C motif) ligand 8.[40] This hormone–inflammation interplay might be particularly relevant in estrogen receptor-positive subtypes, suggesting a plausible biological context for the association observed in this study. Nevertheless, the exact mechanisms linking NPAR to BC remain speculative and warrant further investigation.

The clinical appeal of NPAR is underpinned by its low cost and high accessibility, as it is derived from routine blood parameters obtained through standard venipuncture. Its estimated cost is a fraction of that of mammography. While mammography remains the gold standard for BC detection, its limitations—including reduced sensitivity in dense breasts and higher associated costs—highlight the need for complementary screening tools. We propose a stratified screening strategy in which NPAR could serve as an initial pre-screening tool to identify individuals at higher risk (e.g., NPAR ≥ 2.5), who could then be prioritized for mammographic examination. Modeling studies based on similar approaches suggest that this strategy could reduce unnecessary mammograms by 30–40% and improve early-stage cancer detection rates.[41,42] In addition, the ease of serial measurement makes NPAR a candidate for ongoing risk monitoring in high-risk populations, potentially bridging the intervals between routine imaging. Consequently, our findings suggest that NPAR, as a low-cost and accessible measure, could be a candidate for further investigation in BC risk assessment. If its value is validated in prospective studies, it could be implemented in resource-limited settings, helping to address healthcare access disparities among low-income women.

This study has several strengths, including the use of a nationally representative sample with complex survey weighting, which enhances the generalizability

of the findings. Extensive adjustment for confounders strengthens the validity of the observed associations. Limitations include the lack of temporal sequencing inherent in cross-sectional designs, the absence of detailed cancer characteristics (e.g., stage, molecular subtype), and insufficient data on other inflammatory markers, which would be necessary for direct comparison with NPAR. Future research should prioritize prospective validation of NPAR in independent cohorts, ideally with repeated measurements and accompanied by molecular subtyping. Mechanistic studies exploring the role of neutrophils and albumin in the breast tumor microenvironment—particularly in relation to hormone receptor status—would help clarify the biological relevance of NPAR. Finally, comparative analyses with other inflammatory indices are needed to determine whether NPAR offers substantive advantages in clinical prediction.

5. Conclusion

In conclusion, this study explored, for the first time, the association between NPAR and BC risk, highlighting a potential synergistic link between inflammation and nutritional status in the context of BC. As a convenient and cost-effective composite indicator, NPAR, combined with the high predictive performance of the nomogram model, offers novel insights for early screening and risk stratification of BC. Future studies should further validate its clinical applicability and explore its value in personalized prevention strategies. These findings not only enrich the etiological understanding of BC but also provide a scientific foundation for future public health interventions.

Acknowledgments

None.

Funding

This study was supported by the Henan Provincial Medical Science and Technology Research Plan (Joint Construction Projects, Grant Numbers: LHGJ20200957 and LHGJ20210911), issued by the Henan Provincial Health Commission (Yu Wei Ke Jiao Han [2021] No. 1 and No. 39).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Ruijuan Heng, Pan Qi
Formal analysis: Xiaoying He, Tang Xiao, Haifang Zhao
Investigation: Xiaoying He, Tang Xiao
Methodology: Xinyue Huang, Sheng Xu, Junfeng Liu, Rui Qi
Writing–original draft: Yunxiang Wang, Jie Yin, Yong Zhou
Writing–review & editing: Xiaoying He, Pan Qi

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data used in this study are publicly available from the National Health and Nutrition Examination Survey (NHANES) database, which is maintained by the U.S. Centers for Disease Control and Prevention (CDC). The dataset is de-identified and approved for unrestricted public access by the CDC/National Center for Health Statistics Institutional Review Board. Researchers can access the data through the official NHANES portal: https://www.cdc.gov/nchs/nhanes/index.htm.

References

1. Kim J, Harper A, McCormack V, *et al*. Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nat Med*. 2025;31(4):1154-1162.
doi: 10.1038/s41591-025-03502-3
2. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D, *et al*. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. *JAMA Oncol*. 2019;5:1749-1768.
doi: 10.1001/jamaoncol.2019.2996
3. Samavat H, Kurzer MS. Estrogen metabolism and breast cancer. *Cancer Lett*. 2015;356(2 Pt A):231-243.
doi: 10.1016/j.canlet.2014.04.018
4. Brody JG, Rudel RA. Environmental pollutants and breast cancer. *Environ Health Perspect*. 2003;111(8):1007-1019.
doi: 10.1289/ehp.6310
5. Rodrigues-Ferreira S, Nahmias C. Predictive biomarkers for personalized medicine in breast cancer. *Cancer Lett*. 2022;545:215828.
doi: 10.1016/j.canlet.2022.215828
6. Kearney MR, McGuinness JE, Kalinsky K. Clinical trial data and emerging immunotherapeutic strategies: Hormone receptor-positive, HER2- negative breast cancer. *Breast Cancer Res Treat*. 2021;189(1):1-13.
doi: 10.1007/s10549-021-06291-8

7. Laborda-Illanes A, Sanchez-Alcoholado L, Dominguez-Recio ME, *et al*. Breast and gut microbiota action mechanisms in breast cancer pathogenesis and treatment. *Cancers (Basel)*. 2020;12(9):2465.

doi: 10.3390/cancers12092465

8. Pesic M, Greten FR. Inflammation and cancer: Tissue regeneration gone awry. *Curr Opin Cell Biol*. 2016;43:55-61.

doi: 10.1016/j.ceb.2016.07.010

9. Gobel A, Dell'Endice S, Jaschke N, *et al*. The role of inflammation in breast and prostate cancer metastasis to bone. *Int J Mol Sci*. 2021;22(10):5078.

doi: 10.3390/ijms22105078

10. Dai Q, Wu W, Amei A, Yan X, Lu L, Wang Z. Regulation and characterization of tumor-infiltrating immune cells in breast cancer. *Int Immunopharmacol*. 2021;90:107167.

doi: 10.1016/j.intimp.2020.107167

11. Bernardo G, Le Noci V, Di Modica M, *et al*. The emerging role of the microbiota in breast cancer progression. *Cells*. 2023;12(15):1945.

doi: 10.3390/cells12151945

12. Wong GL, Manore SG, Doheny DL, Lo HW. STAT family of transcription factors in breast cancer: Pathogenesis and therapeutic opportunities and challenges. *Semin Cancer Biol*. 2022;86(Pt 3):84-106.

doi: 10.1016/j.semcancer.2022.08.003

13. Hibino S, Kawazoe T, Kasahara H, *et al*. Inflammation-induced tumorigenesis and metastasis. *Int J Mol Sci*. 2021;22(11):5421.

doi: 10.3390/ijms22115421

14. Ozga AJ, Chow MT, Luster AD. Chemokines and the immune response to cancer. *Immunity*. 2021;54(5):859-874.

doi: 10.1016/j.immuni.2021.01.012

15. Li X, Wu M, Chen M, *et al*. The Association between neutrophil-percentage-to-albumin ratio (NPAR) and mortality among individuals with cancer: Insights from national health and nutrition examination survey. *Cancer Med*. 2025;14(2):e70527.

doi: 10.1002/cam4.70527

16. Li J, Xiang T, Chen X, Fu P. Neutrophil-percentage-to-albumin ratio is associated with chronic kidney disease: Evidence from NHANES 2009-2018. *PLoS One*. 2024;19(8):e0307466.

doi: 10.1371/journal.pone.0307466

17. Guo Y, Wei L, Patel SH, *et al*. Serum albumin: Early prognostic marker of benefit for immune checkpoint inhibitor monotherapy but not chemoimmunotherapy. *Clin Lung Cancer*. 2022;23(4):345-355.

doi: 10.1016/j.cllc.2021.12.010

18. Tang Y, Hou H, Li L, *et al*. Neutrophil percentage-to-albumin ratio: A good parameter for the evaluation of the severity of anti-NMDAR encephalitis at admission and prediction of short-term prognosis. *Front Immunol*. 2022;13:847200.

doi: 10.3389/fimmu.2022.847200

19. Xu B, Wu Q, La R, *et al*. Is systemic inflammation a missing link between cardiometabolic index with mortality? Evidence from a large population-based study. *Cardiovasc Diabetol*. 2024;23(1):212.

doi: 10.1186/s12933-024-02251-w

20. Pan Y, Yuan Y, Yang J, *et al*. U-shaped relationship between frailty and non-HDL-cholesterol in the elderly: A cross-sectional study. *Front Nutr*. 2025;12:1596432.

doi: 10.3389/fnut.2025.1596432

21. Kim H, Hur M, Yi JH, *et al*. Detection of blasts using flags and cell population data rules on Beckman Coulter DxH 900 hematology analyzer in patients with hematologic diseases. *Clin Chem Lab Med*. 2024;62(5):958-966.

doi: 10.1515/cclm-2023-0932

22. Virag D, Schlosser G, Borbely A, *et al*. A mass spectrometry strategy for protein quantification based on the differential alkylation of cysteines using iodoacetamide and acrylamide. *Int J Mol Sci*. 2024;25(9):4656.

doi: 10.3390/ijms25094656

23. Liang H, Pan K, Wang J, Lin J. Association between neutrophil percentage-to-albumin ratio and breast cancer in adult women in the US: Findings from the NHANES. *Front Nutr*. 2025;12:1533636.

doi: 10.3389/fnut.2025.1533636

24. Boakye E, Erhabor J, Obisesan O, *et al*. Comprehensive review of the national surveys that assess E-cigarette use domains among youth and adults in the United States. *Lancet Reg Health Am*. 2023;23:100528.

doi: 10.1016/j.lana.2023.100528

25. Li Y, Yu M, Yang M, Yang J. The association of systemic immune-inflammation index with incident breast cancer and all-cause mortality: Evidence from a large population-based study. *Front Immunol*. 2025;16:1528690.

doi: 10.3389/fimmu.2025.1528690

26. Teng TQ, Liu J, Hu FF, Li QQ, Hu ZZ, Shi Y. Association of composite dietary antioxidant index with prevalence of stroke: Insights from NHANES 1999-2018. *Front Immunol*. 2024;15:1306059.

doi: 10.3389/fimmu.2024.1306059

27. Qi X, Li Y, Fang C, *et al*. The associations between dietary fibers intake and systemic immune and inflammatory biomarkers, a multi-cycle study of NHANES 2015-2020. *Front Nutr*. 2023;10:1216445.

doi: 10.3389/fnut.2023.1242115

28. Vaudin A, Wambogo E, Moshfegh AJ, Sahyoun NR. Sodium and potassium intake, the sodium to potassium ratio, and associated characteristics in older adults, NHANES 2011-2016. *J Acad Nutr Diet*. 2022;122(1):64-77.

doi: 10.1016/j.jand.2021.06.012

29. Cao Y, Li P, Zhang Y, *et al*. Association of systemic immune inflammatory index with all-cause and cause-specific mortality in hypertensive individuals: Results from NHANES. *Front Immunol*. 2023;14:1087345.

doi: 10.3389/fimmu.2023.1087345

30. Ma Y, Xu Y, Du L, Bai L, Tang H. Association between systemic immune inflammation index and short term prognosis of acute on chronic liver failure. *Sci Rep.* 2024;14(1):21535.

doi: 10.1038/s41598-024-72447-3

31. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. *JAMA*. 2005;293(15):1861-1867.

doi: 10.1001/jama.293.15.1861

32. Jiang Y, Wang X, Xia L, *et al*. A cohort study of diabetic patients and diabetic foot ulceration patients in China. *Wound Repair Regen*. 2015;23(2):222-230.

doi: 10.1111/wrr.12263

33. Lu Z, Long Y, Li J, *et al*. Simultaneous inhibition of breast cancer and its liver and lung metastasis by blocking inflammatory feed-forward loops. *J Control Release*. 2021;338:662-679.

doi: 10.1016/j.jconrel.2021.08.047

34. Wang Y, Chen S, Tian C, *et al*. Association of systemic immune biomarkers with metabolic dysfunction-associated steatotic liver disease: A cross-sectional study of NHANES 2007-2018. *Front Nutr*. 2024;11:1415484.

doi: 10.3389/fnut.2024.1415484

35. Liu CF, Chien LW. Predictive role of neutrophil-percentage-to-albumin ratio (NPAR) in nonalcoholic fatty liver disease and advanced liver fibrosis in nondiabetic US adults: Evidence from NHANES 2017-2018. *Nutrients*. 2023;15(8):1892.

doi: 10.3390/nu15081892

36. Jiao S, Zhou J, Feng Z, *et al*. The role of neutrophil percentage to albumin ratio in predicting 1-year mortality in elderly patients with hip fracture and external validation. *Front Immunol*. 2023;14:1223464.

doi: 10.3389/fimmu.2023.1223464

37. Wang M, Wang S, Hu J, Wang X, Pang Y, Sun X. Associations between neutrophil percentage-to-albumin ratio with all-cause and cause-specific mortality among US cancer survivors: Evidence from NHANES 2005-2018. *Front Nutr*. 2025;12:1541609.

doi: 10.3389/fnut.2025.1541609

38. Lakritz JR, Poutahidis T, Mirabal S, *et al*. Gut bacteria require neutrophils to promote mammary tumorigenesis. *Oncotarget*. 2015;6(11):9387-9396.

doi: 10.18632/oncotarget.3328

39. Gago-Dominguez M, Castelao JE, Pike MC, Sevanian A, Haile RW. Role of lipid peroxidation in the epidemiology and prevention of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2829-2839.

doi: 10.1158/1055-9965.epi-05-0015

40. Haim K, Weitzenfeld P, Meshel T, Ben-Baruch A. Epidermal growth factor and estrogen act by independent pathways to additively promote the release of the angiogenic chemokine CXCL8 by breast tumor cells. *Neoplasia*. 2011;13(3):230-243.

doi: 10.1593/neo.101340

41. Catarzi S, Giuseppetti GM, Rizzatto G, *et al*. Studio multicentrico per la valutazione dell'efficacia diagnostica della mammografia e dell'ecografia nelle neoplasie mammarie non palpabili [A multicenter study for the evaluation of the diagnostic efficiency of mammography and echography in nonpalpable breast neoplasms]. *Radiol Med*. 1992;84(3):193-197.

42. Li C, Wu J, Jiang L, *et al*. The predictive value of inflammatory biomarkers for major pathological response in non-small cell lung cancer patients receiving neoadjuvant chemoimmunotherapy and its association with the immune-related tumor microenvironment: A multi-center study. *Cancer Immunol Immunother*. 2023;72(3):783-794.

doi: 10.1007/s00262-022-03262-w

ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

SHORT COMMUNICATION

Enhancing binary dose–response analysis in clinical and translational research: Leveraging grouped data techniques in R

Jenny-Hoa Q. Nguyen*, Fudan Zheng, Yuan Xiong, and Mahesh N. Samtani

Department of Clinical Pharmacology and Pharmacometrics, Johnson and Johnson, Spring House, Pennsylvania, United States of America

Abstract

Background: In pharmacometric analyses, binary dose–response outcome data are used to understand drug potency through the pharmacologic parameter, Effective dose 50 (ED50). Optimal treatment strategies can be developed by characterizing a drug's dose–response curve, which provides insights into the theoretical maximum effect and the steepness of the curve in response to changes in dose or exposure. Approaches for analyzing group-level response data have not been systematically described in pharmacometric literature, although they are commonly applied in the statistical literature. **Aim:** This study demonstrates the use of R to analyze grouped or ungrouped binary data, with a focus on pharmacometric applications. **Methods:** Simulated data were generated to represent a hypothetical Phase 2, dose-ranging, placebo-controlled, randomized clinical trial of drug X, consisting of 250 participants randomized into five distinct cohorts, including a single placebo arm. Linear and non-linear Emax models were fit to the simulated data. **Results:** Both grouped and ungrouped data approaches produced identical final parameter estimates in logistic regression using the linear and Emax models. The same ED50 value for drug X was obtained from both approaches in the Emax model. **Conclusion:** This study demonstrates the various methods by which summary- or subject-level binary data can be analyzed using R to model binary response data. **Relevance for patients:** This work helps bridge the gap between statistical and pharmacometric analysis techniques in the context of binary data analysis. This type of data may facilitate comparative assessments of drug potency and maximal effect using publicly available information from scientific publications or regulatory approval documents.

Keywords: Generalized linear models; Generalized non-linear models; Logistic regression; R software; Binary data; Dose–response analysis

***Corresponding author:**
Jenny-Hoa Q. Nguyen
(jnguye58@its.jnj.com)

Citation: Nguyen JQ, Zheng F, Xiong Y, Samtani MN. Enhancing binary dose–response analysis in clinical and translational research: Leveraging grouped data techniques in R. *J Clin Transl Res*. 2025;11(6):76-86.
doi: 10.36922/JCTR025300045

Received: July 23, 2025

Revised: August 25, 2025

Accepted: September 11, 2025

Published online: October 22, 2025

1. Introduction

In Phase 2 dose-finding studies, dose–response information is often expressed as a binary outcome versus dose, where the achievement of an endpoint of interest—such as a PASI90 response (representing a 90% improvement from baseline in the Psoriasis Area and Severity Index) in psoriasis clinical trials—is reported at each dose level evaluated

for a given drug. Generalized linear models (GLMs), such as logistic regression, can be used to effectively analyze this type of data, regardless of whether the data are programmed at a grouped or individual (i.e., ungrouped) level.[1,2] In addition, GLMs for binary data can be further extended to generalized non-linear models (GNMs), where non-linear relationships between the logit transformation of the response and the predictors of interest can be introduced into the model, including the standard Emax relationship frequently used in pharmacometric analyses.[3] In the context of logistic regression, the logit link function transforms the predicted probabilities using log-odds, allowing binary outcome data to be modeled while ensuring that predicted probabilities remain within the interval (0, 1). While alternative link functions—such as the probit link function—can also be specified and applied, the logit link function is used here due to its ease of interpretation through odds ratios.

This study demonstrates the use of the "stats::glm()" and "gnm::gnm()" functions in R for analyzing summary-level binary response data with a grouped data approach.[4,5] The simulated data used in this study represent a landmark analysis where only one observation is collected per patient. As a result, complex models accounting for hierarchical or correlated data dependencies were unnecessary, as the simulated data did not require inclusion of random effects in the model structure; therefore, the use of packages such as "glmmTMB" and "nlme" was not required.[6,7] Furthermore, results from the grouped and ungrouped approaches will be provided to illustrate that the results from either approach are identical, thereby giving pharmacometricians the flexibility to work with a variety of data structures to achieve the same goal. The grouped data approach may be a valuable tool for model-based meta-analysis when analyzing summary-level binary data.

2. Materials and methods

2.1. Study data

In this study, simulated data representing a hypothetical Phase 2, placebo-controlled, dose-ranging randomized clinical trial for drug X were generated using R version 4.2 (R Foundation for Statistical Computing, Austria). A total of 250 participants were assigned in the simulation to one of five possible dose levels (0 mg [placebo], 1 mg, 5 mg, 10 mg, and 20 mg), with 50 participants per cohort. The probability of response at each dose level was assigned based on real-life clinical data. To generate the individual response data, the probability of response at each dose level was used to assign a 0 or 1 value representing failure or success, respectively, from a binomial distribution.

2.2. Efficacy data

Efficacy results describing participants' responses to drug X as a binary variable (i.e., responder or non-responder) were simulated at the landmark time point, defined here as the primary endpoint evaluated at week 12. To capture elements of a real clinical trial and perform a landmark analysis at the pivotal endpoint, the efficacy data were assigned based on real-life clinical data. The guidelines used to generate the hypothetical clinical trials are as follows:

(i) The placebo response rate was allowed to be >0%
(ii) The maximum response rate was not allowed to exceed 90%
(iii) In general, higher doses of drug X were expected to lead to higher response rates; however, the response rates between placebo and 1 mg may be similar, and the response rates between 10 mg and 20 mg may be similar.

The 90% upper limit for maximum response reflects clinical reality, as responses in many diseases rarely reach 100% due to their complex and multifactorial nature. The similarity observed between the 10 mg and 20 mg doses aligns with the Emax relationship demonstrated in recent studies, which show a plateau in efficacy where responses flatten at higher doses and do not approach complete remission.[8] This pattern is consistent with pharmacodynamic behavior across various diseases, suggesting that increasing the dose beyond a certain point provides minimal additional benefit.

2.3. Fitting GLM with "stats::glm()"

Data were formatted as illustrated in Figure 1A for the grouped data approach using the proportion of successes, or in Figure 1B for the grouped data approach using the number of successes and failures (i.e., non-responders). For the ungrouped data approach, which is the more standard method, data were provided as individual-level responses, as illustrated in Figure 1C. Appendix 1 provides the complete R code used for model implementation, including the embedded datasets defined within the script. It illustrates both the ungrouped "long format" and grouped ("summary-level") data structures, demonstrating how the ungrouped data can be more cumbersome to prepare, whereas the grouped data are more compact and easier to format.

To fit a logistic regression model with dose as the predictor, the following arguments in the "stats::glm()" function were used (Expression I):

```
glm.mod1 <- glm(                                    (I)
formula = <response> ~ <predictor>,
```

A

dose_group	dose_val	time	endpoint	n_subjs	prop_resp
0 mg	0	12	Endpoint A	50	0.08
1 mg	1	12	Endpoint A	50	0.24
5 mg	5	12	Endpoint A	50	0.54
10 mg	10	12	Endpoint A	50	0.84
20 mg	20	12	Endpoint A	50	0.84

B

dose_group	dose_val	time	endpoint	n_success	n_fail
0 mg	0	12	Endpoint A	4	46
1 mg	1	12	Endpoint A	12	38
5 mg	5	12	Endpoint A	27	23
10 mg	10	12	Endpoint A	42	8
20 mg	20	12	Endpoint A	42	8

C

ID	dose_group	dose_val	time	endpoint	response
1	0 mg	0	12	Endpoint A	1
2	0 mg	0	12	Endpoint A	1
3	0 mg	0	12	Endpoint A	1
4	0 mg	0	12	Endpoint A	1
5	0 mg	0	12	Endpoint A	0

Figure 1. Input data formats for fitting models using the logit transformation of the response. (A) Group-level data showing the proportion of successes for each group. (B) Group-level data showing total counts of successes and failures for each dose group. (C) Individual-level data indicating response outcomes as failure (0) or success (1).

```
weights = <total number of subjects per group>,
family = binomial(link = "logit"),
data = <dataframe>
)
```

When fitting the proportion of successes as the response (i.e., data formatted as in Figure 1A), a warning message from R may arise, alerting the user that the specification of "family = binomial(link = "logit")" contains non-integer (i.e., proportion) values. This occurs because the binomial distribution is the sum of *n* Bernoulli trials, where each trial may only take the value 0 or 1 to represent failure or success, and therefore, non-integer values are not expected.[9] Here, the *n* Bernoulli trials refer to the *n* total participants per dosing group; therefore, the "weights" argument provides the complementary information needed to fit the model. This message can be safely ignored, as it does not impact the results returned from R when the "weights" argument is supplied. The model can thus be fit as follows (Expression II):

```
glm.mod1.fig1a <- glm(                                    (II)
formula = prop_resp ~ dose_val,
weights = n_subjs,
family = binomial(link = "logit"),
data = df_input_summary
)
```

If proportions are not used, the "weights" argument does not need to be supplied, since the total number of successes and failures can be used instead. This results in the following function (Expression III) for data formatted as in Figure 1B:

```
glm.mod1.fig1b <- glm(                                    (III)
formula = cbind(n_success, n_fail) ~ dose_val,
family = binomial(link = "logit"),
data = df_input_summary
)
```

To demonstrate that the grouped and ungrouped responses yield the same model results, the function used

to model individual-level data (Figure 1C) is as follows (Expression IV):

```
glm.mod1.fig1d <- glm                                    (IV)
formula = response ~ dose
family = binomial(link = "logit"),
data = df_input_individual
)
```

2.4. Fitting GNMs with "gnm::gnm()"

Data were presented in one of the forms described in the previous section and illustrated in Figure 1. Arguments such as "formula," "weights," "family," and "data" were provided for "gnm::gnm()" in the same manner as in "stats::glm()," depending on the format of the data. Custom model functions were defined and subsequently supplied to the "formula" argument of the function. Figure 2 illustrates how a custom function was defined to fit the standard Emax function, where the Hill shape parameter was assumed to equal 1. It should be noted that the Emax model may not have been appropriate in cases where the data did not support an Emax-like relationship, such as inverted U-shaped responses or scenarios lacking a clear plateau.

In the function definition for "fx_emax()" provided in Figure 2, "predictors" refers to the list of named parameters (referred to as "predLabels" in the "term" argument) that will be estimated by the model and are provided as a named list with an initial value of 1, while "variables" refers to the independent variable used for the model fit, such as dose (referred to as "varLabels" in the "term" argument). The "term" argument requires the definition of the custom function to be used, and the "sprinf()" function facilitates the definition of custom functions by utilizing the indices provided in the "predictors" and "variables." Additional guidance can be found in the R help documentation for this function (accessible via typing `?sprintf` in the R console, or via https://www.rdocumentation.org/packages/base/versions/3.6.2/topics/sprintf). Finally, the "class" of the function must be defined as "non-lin" before fitting the "fx_emax()" model using "gnm::gnm()."

To fit the data in the formats presented in Figure 1, the guidelines provided in the previous section for logistic regression models were followed, ensuring that the predictors are provided in the formula call with the defined "fx_emax()" function (Figure 2). For example, the function call for the data provided in Figure 1A is as follows (Expression V):

```
gnm.mod2.fig1a <- gnm(                                   (V)
formula = prop_resp ~ fx_emax(dose_val),
weights = n_subjs,
family = binomial(link = "logit"),
data = df_input_summary
)
```

3. Results

3.1. Comparison of grouped and ungrouped approaches

Tables 1 and 2 present the results of the linear (GLM) and non-linear (GNM) model fits, respectively. In addition, Figure 3 illustrates the results of the linear and Emax models. The grouped and ungrouped approaches provided identical model estimates for each model type fitted to the data. Researchers are advised to consult outside texts for information regarding goodness-of-fit assessments for logistic regression, such as deviance residuals and the HosmerLemeshow test.[10]

A Standard Emax function

$$\text{logit}(\text{p} = \text{probability of response}) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \frac{E_{max} \cdot \text{Dose}}{ED_{50} + \text{Dose}}$$

B R code

```
fx_emax <- function(x){
  list(
    predictors = list(Emax = 1, ED50 = 1),
    variables  = list(substitute(x)),
    term       = function(predLabels, varLabels) {
      sprintf(
        "%s * %s / (%s + %s)",
        predLabels[1], varLabels[1],
        predLabels[2], varLabels[1]
      )
    }
  )
}

class(fx_emax) <- "nonlin"
```

Figure 2. Defining custom functions in R for use with the "gnm::gnm()" function based on the standard Emax function. (A) Standard Emax model equation. (B) R code.

Table 1. Final model parameter estimates from the linear model (logistic regression) fit using a generalized linear model

Parameter	Grouped data with the proportion of success		Grouped data with failure and success counts		Ungrouped data	
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Intercept	−1.270	0.2130	−1.270	0.2130	−1.270	0.2130
Slope	0.204	0.0284	0.204	0.0284	0.204	0.0284
AIC[a]	45.857		45.857		266.358	

Notes: [a]AICs are not directly comparable between the grouped and ungrouped approaches, as the definitions of the likelihood functions optimized in the different model fits differ. As a metric, AIC combines the number of parameters with a likelihood measure, such as negative two times the log-likelihood. Figure 3A for the model-predicted results using this generalized linear model.
Abbreviation: AIC: Akaike information criterion.

Table 2. Final model parameter estimates from the non-linear model (standard Emax) fit using a generalized non-linear model

Parameter	Grouped data with the proportion of success		Grouped data with failure and success counts		Ungrouped data	
	Estimate	Standard error	Estimate	Standard error	Estimate	Standard error
Intercept	−2.28	0.423	−2.28	0.423	−2.28	0.423
Emax	5.01	0.742	5.01	0.742	5.01	0.742
ED50	4.29	2.310	4.29	2.310	4.29	2.310
AIC[a]	27.944		27.944		248.715	

Notes: [a]AICs are not directly comparable between the grouped and ungrouped approaches, as the definitions of the likelihood functions optimized in the different model fits differ. As a metric, AIC combines the number of parameters with a likelihood measure, such as negative two times the log-likelihood. Figure 3B for the model-predicted results using this generalized non-linear model.
Abbreviations: AIC: Akaike information criterion; ED50: Effective Dose 50.

3.2. Comparison of Effective Dose 50 (ED50) estimates between the generalized linear and GNMs

In the GLM, the ED50 value can be estimated by algebraically solving for the dose level at which 50% of the response is achieved. Using the output from Table 1, the derived ED50 value for GLM was calculated to be 6.23 mg, whereas the estimated ED50 value for the GNM was 4.23 mg (Table 2). However, Figure 3 shows that the GNM provides a better fit to the simulated data, which was further confirmed by the lower Akaike information criterion value for the GNM relative to the GLM when both models were fit using the same data structure.

The difference in the ED50 values obtained from the two model approaches may be attributed to several factors that demonstrate the inadequacy of the linear model for the simulated data. First, the maximum response rate was not allowed to exceed 90% in the simulated data and therefore cannot approach 100%, as shown in Figure 3A. It should be noted that the underlying assumption in a logistic regression model is that there is a monotonic increase toward 100% (or a monotonic decrease toward 0%) and, therefore, it cannot plateau at an intermediate value. Second, a placebo response was observed in the hypothetical trial, which could not be adequately captured by the GLM approach; however, as shown in Figure 3B, this placebo response was well represented in the GNM approach.

When applied to clinical trial data, the GNM offers the capability to analyze situations in which the observed maximum response rate does not approach 100% and a placebo response is observed. While the intercept in the GLM can represent the placebo response, its ability to model such data are limited as the shape parameter is not estimable, which likely contributes to the overestimation of the placebo effect in the GLM fit (Figure 3A). In addition, it should be noted that the GLM offers the flexibility to estimate a polynomial fit (e.g., $y \sim x + x^2 + x^3 \ldots$), but such a model lacks pharmacological interpretability with respect to its parameters.

3.3. Applications within pharmacometrics and drug development

In this study, using simulated data for drug X, the ED50 value estimated from the GNM—whether using either the grouped or ungrouped approach—was 4.29 mg (Table 2). In the context of drug development, the ED50 value can be used to compare different drugs within the same class to determine which drug may be more potent and thus requires a lower dose to achieve the same therapeutic effect. If several candidates within the same drug class are

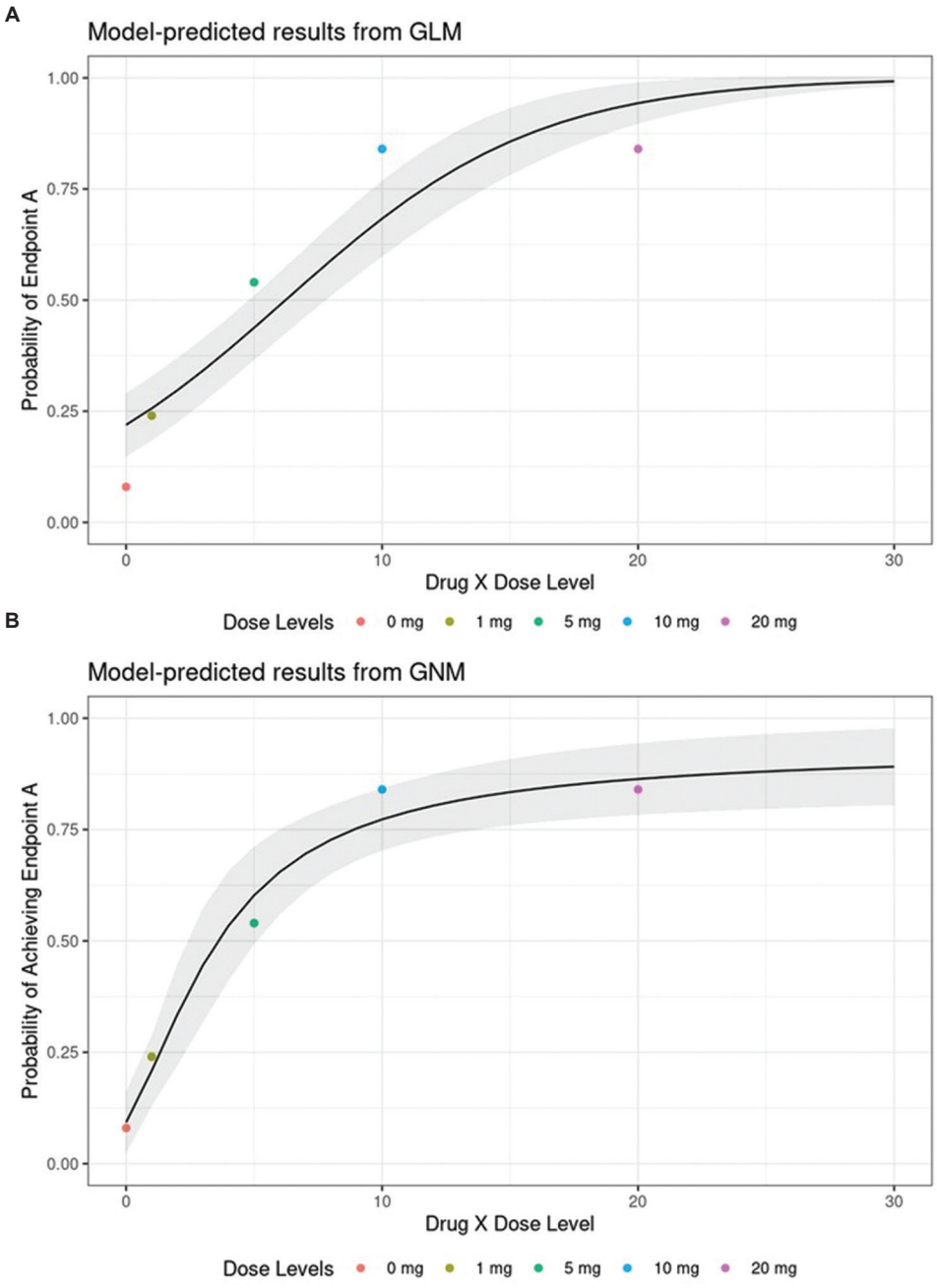


Figure 3. Model-predicted probability of achieving endpoint A at week 12 using linear and non-linear models. (A) Model-predicted results obtained using a standard logistic regression model implemented in "stats::glm()." (B) Model-predicted results obtained using a logit-transformed Emax model implemented in "gnm::gnm()."
Abbreviations: GNM: Generalized non-linear model; GLM: Generalized linear model.

being considered for further development, and assuming all candidates demonstrate similar safety profiles, the candidate with a lower ED50 value may be a more suitable option from a cost-of-goods or dosing perspective. This would allow a lower injection volume for a subcutaneously administered drug, assuming similar pharmacokinetic half-lives across drugs.

However, if several candidates from different classes (e.g., biologics versus small-molecule drugs) are being considered for development, and the Emax value can help determine which drug class is more efficacious based on the maximal effect of the drug, where a higher Emax value indicates greater efficacy. For example, biologics are generally more efficacious than small-molecule drugs

for the treatment of psoriasis based on efficacy data from registrational clinical trials.[11-18] The tutorial described in this study could be applied to estimate Emax values for these drugs, provided that appropriate data are available.

4. Discussion

The choice between the grouped and ungrouped approaches for analyzing binary dose–response data depends on the nature of the available dataset and the objectives of the analysis. Both methods yield identical parameter estimates when applied correctly, as demonstrated in this study, although they differ in practical considerations.

Grouped data methods are particularly advantageous when only summary-level information is available, such as in meta-analyses or when extracting data from published literature or regulatory documents. These approaches reduce data-handling complexity by condensing individual-level observations into aggregated counts or proportions, thereby simplifying model implementation.

On the other hand, ungrouped data methods should be prioritized when individual-level data are available, as they allow greater flexibility in modeling, including the incorporation of covariates, random effects, and hierarchical structures. These features are essential for addressing variability across subjects and for conducting more detailed analyses, such as mixed-effects modeling or time-to-event evaluations.

GLMs are straightforward to implement and interpret, making them suitable for scenarios where the dose–response relationship is approximately linear on the logit scale. GNMs, which incorporate Emax or similar functions, offer greater flexibility for modeling pharmacologically plausible relationships. While GNMs require more complex model specification and interpretation, they provide parameters (e.g., Emax, ED50) that are directly relevant to pharmacology and drug development decision-making. In practice, GLMs can be used for rapid assessments or when the primary interest lies in odds ratios and the data exhibit near-linear behavior on the logit scale. When modeling dose–response relationships that exhibit saturation effects, or when pharmacologic interpretability of parameters is critical, GNMs should be preferentially employed.

Furthermore, the ED50 and Emax parameters derived from the GNMs have direct clinical relevance. ED50 indicates the dose required to achieve half of the maximal therapeutic effect, guiding dose selection and optimization strategies. A lower ED50 may indicate a more potent drug, which can influence candidate prioritization and cost considerations. Similarly, Emax reflects the ceiling of efficacy achievable with a drug, which is critical for comparing therapeutic classes and establishing realistic expectations for treatment outcomes. Understanding these parameters within the context of grouped versus ungrouped approaches ensures that pharmacometric analyses remain clinically meaningful and aligned with patient-centric goals.

Ultimately, bridging statistical and pharmacometric methodologies enhances the robustness of dose–response evaluations and facilitates informed decision-making in drug development.

5. Conclusion

This simple and practical approach is presented not only to promote the use of grouped data for analyzing publicly available information on categorical endpoints but also to highlight the value of utilizing statistical methods within pharmacometrics and to further advocate for bridging these two disciplines.[19] This study illustrates different approaches by which grouped data can be analyzed using R—for example, as group-level outcomes represented either as proportions of successes or as a two-column matrix of grouped data showing the numbers of successes and failures. In addition, dose–response data can be replaced with quantile-based exposure-response data obtained from published studies, where the proportions of responders are reported per quartile or tertile of exposure data, or with information extracted from digitized clinical datasets. Such data may be valuable for deriving EC50 estimates for assets with publicly available information reported in scientific publications or on summary basis of approval documents from regulatory agencies.

Acknowledgments

None.

Funding

None.

Conflict of interest

Jenny-Hoa Q. Nguyen, Fudan Zheng, Yuan Xiong, and Mahesh N. Samtani are employees and stockholders of Johnson & Johnson (JNJ). This work was conducted as part of fulfilling responsibilities within the company.

Author contributions

Conceptualization: Mahesh N. Samtani
Data curation: Jenny-Hoa Q. Nguyen
Formal analysis: Jenny-Hoa Q. Nguyen
Methodology: Jenny-Hoa Q. Nguyen, Fudan Zheng
Visualization: Jenny-Hoa Q. Nguyen
Writing–original draft: Jenny-Hoa Q. Nguyen

Writing–review & editing: Fudan Zheng, Yuan Xiong, Mahesh N. Samtani

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data and relevant R code used in this publication are provided in Appendix 1.

References

1. Venables WN, Ripley BD. Generalized linear models. In: *Modern Applied Statistics with S*. 4th ed. New York: Springer; 2002. p. 330-370.
2. Faraway JJ. Generalized linear models. In: *Extending the Linear Model with R*. 2nd ed. New York: Chapman and Hall; 2016. p. 151-174.
3. Lane PW. Generalized nonlinear models. In: Prat A. editors. *COMPSTAT*. Germany: Physica-Verlag HD; 1996.

   doi: 10.1007/978-3-642-46992-3_42
4. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria; 2022. Available from: https://www.com/r-project.org [Last accessed on 2025 Aug 20].
5. Turner H, Firth D. *gnm: Generalized Nonlinear Models*. R Package Version 1.1-5. 2023. Available from: https://www.com/cran.r-project.org/package=gnm [Last accessed on 2025 Aug 20].
6. Brooks ME, Bolker B, Kristensen K, *et al*. *glmmTMB: Generalized Linear Mixed Models using Template Model Builder*. R Package Version 1.1.12; 2025. Available from: https://www.com/cran.r-project.org/package=glmmtmb [Last accessed on 2025 Aug 20].
7. Pinherio J, Bates D, DebRoy S, *et al*. *Nlme: Linear and Nonlinear Mixed Effects Models*. R Package Version 3.1-168; 2025. Available from: https://www.com/cran.r-project.org/package=nlme [Last accessed on 2025 Aug 20].
8. Checchio T, Ahadieh S, Gupta P, *et al*. Quantitative evaluations of time-course and treatment effects of systemic agents for psoriasis: A model-based meta-analysis. *Clin Pharmacol Ther*. 2017;102(6):1006-1016.

   doi: 10.1002/cpt.732
9. Casella G, Berger RL. Binomial distribution. In: *Statistical Inference*. 2nd ed. Pacific Grove: Duxbury Press; 2002. p. 89-91.
10. Hosmer DW Jr., Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. 3rd ed. New York: Wiley; 2013.
11. Armstrong AW, Puig L, Joshi A, *et al*. Comparison of biologics and oral treatments for plaque psoriasis: A meta-analysis. *JAMA Dermatol*. 2020;156(3):258-269.

    doi: 10.1001/jamadermatol.2019.4029
12. Thaçi D, Blauvelt A, Reich K, *et al*. Secukinumab is superior to ustekinumab in clearing skin of subjects with moderate to severe plaque psoriasis: CLEAR, a randomized controlled trial. *J Am Acad Dermatol*. 2015;73(3):400-409.

    doi: 10.1016/j.jaad.2015.05.013
13. Reich K, Armstrong AW, Langley RG, *et al*. Guselkumab versus secukinumab for the treatment of moderate-to-severe psoriasis (ECLIPSE): Results from a phase 3, randomised controlled trial. *Lancet*. 2019;394(10201):831-839.

    doi: 10.1016/S0140-6736(19)31773-8
14. Gordon KB, Strober B, Lebwohl M, *et al*. Efficacy and safety of risankizumab in moderate-to-severe plaque psoriasis (UltIMMa-1 and UltIMMa-2): Results from two double-blind, randomised, placebo-controlled and ustekinumab-controlledphase3trials.*Lancet*.2018;392(10148):650-661.

    doi: 10.1016/S0140-6736(18)31713-6
15. Reich K, Pinter A, Lacour JP, *et al*. Comparison of ixekizumab with ustekinumab in moderate-to-severe psoriasis: 24-week results from IXORA-S, a phase III study. *Br J Dermatol*. 2017;177(4):1014-1023.

    doi: 10.1111/bjd.15666
16. Blauvelt A, Papp KA, Griffiths CEM, *et al*. Efficacy and safety of guselkumab, an anti-interleukin-23 monoclonal antibody, compared with adalimumab for the continuous treatment of patients with moderate to severe psoriasis: Results from the phase III, double-blinded, place- and active comparator-controlled VOYAGE 1 trial. *J Am Acad Dermatol*. 2017;76(3):405-417.

    doi: 10.1016/j.jaad.2016.11.041
17. Lebwohl M, Strober B, Meneter A, *et al*. Phase 3 studies comparing brodalumab with ustekinumab in psoriasis. *N Engl J Med*. 2015;373(13):1318-1328.

    doi: 10.1056/NEJMoa1503824
18. Reich K, Gooderham M, Green L, *et al*. The efficacy and safety of apremilast, etanercept and placebo in patients with moderate-to-severe plaque psoriasis: 52-week results from a phase IIIb, randomized, placebo-controlled trial (LIBERATE). *J Eur Acad Dermatol Venereol*. 2017;31(3):507-517.

    doi: 10.1111/jdv.14015
19. Ryeznik Y, Sverdlov O, Svensson EM, Montepiedra G, Hooker AC, Wong WK. Pharmacometrics meets statistics-A synergy for modern drug development. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1134-1149.

    doi: 10.1002/psp4.12696

Appendix

R code

```
# --- [0] R libraries
library(tidyverse)
library(gnm)
# --- [1] Data Setup
# summary-level dataframe
df_input_summary <- structure(
list(
dose_group = structure(1:5, levels = c("0 mg", "1 mg", "5 mg", "10 mg", "20 mg"), class = "factor"),
dose_val = c(0L, 1L, 5L, 10L, 20L),
time = c(12L, 12L, 12L, 12L, 12L),
endpoint = c("Endpoint A", "Endpoint A", "Endpoint A", "Endpoint A", "Endpoint A"),
n_success = c(4L, 12L, 27L, 42L, 42L),
n_fail = c(46L, 38L, 23L, 8L, 8L),
n_subjs = c(50L, 50L, 50L, 50L, 50L),
prop_resp = c(0.08, 0.24, 0.54, 0.84, 0.84)
),
row.names = c(NA, -5L),
class = c("tbl_df", "tbl", "data.frame"))
# individual-level dataframe
df_input_individual <- structure(
list(
ID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36,
37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71,
72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92,93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104,
105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,
131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156,
157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182,
183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208,
209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234,
235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250),

dose_group = structure(c(1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L,
2L, 2L, 2L, 2L, 2L, 2L, 2L, 2L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L,
3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 3L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L,
4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L,
4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 4L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L,
5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L,5L, 5L), levels = c("0 mg", "1 mg",
"5 mg", "10 mg", "20 mg"), class = "factor"),

dose_val = c(0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L,0L, 0L, 0L, 0L, 0L,
0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L, 0L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L, 1L,
1L, 1L, 1L, 1L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L,
5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 5L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L,
10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L,
10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 10L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L,
20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L,
20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L, 20L),
```

```
time = c(12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L, 12L,
12L, 12L),

endpoint = c(“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”,
“Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint
A”, “Endpoint A”, “Endpoint A”, “Endpoint A”, “Endpoint A”),

response = c(1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0)

),
row.names = c(NA, -250L),
class = c(“tbl_df”, “tbl”, “data.frame”)
)
# --- [2] GLM example fit
```

```
glm.mod1.fig1A <- glm(
formula = prop_resp ~ dose_val,
weights = n_subjs,
family = binomial(link = “logit”),
data = df_input_summary
)
glm.mod1.fig1B <- glm(
formula = cbind(n_success, n_fail) ~ dose_val,
family = binomial(link = “logit”),
data = df_input_summary
)
glm.mod1.fig1C <- glm(
formula = response ~ dose_val,
family = binomial(link = “logit”),
data = df_input_summary
)
# --- [3] GNM example fit
# define fx_emax
fx_emax <- function(x){
list(
predictors = list(Emax = 1, ED50 = 1),
variables = list(substitute(x)),
term = function(predLabels, varLabels) {
sprintf(
“%s*%s/(%s + %s) ”,
predLabels[1], varLabels[1],
predLabels[2], varLabels[1]
)
}
)
}
class(fx_emax) <- “nonlin”
# fit gnm
gnm.mod2.fig1A <- gnm(
formula = prop_resp ~ fx_emax(dose_val),
weights = n_subjs,
family = binomial(link = “logit”),
data = df_input_summary
)
glm.mod2.fig1B <- gnm(
formula = cbind(n_success, n_fail) ~ fx_emax(dose_val),
family = binomial(link = “logit”),
data = df_input_summary
)
glm.mod2.fig1C <- gnm(
formula = response ~ fx_emax(dose_val),
family = binomial(link = “logit”),
data = df_input_summary
    )
```

Journal of Clinical and Translational Research

Journal of Clinical and Translational Research (JCTR) welcomes submissions from various research topics that are centered on solving clinically-driven issues to ultimately benefit patients.

You will benefit from the following key features of JCTR as our author:

- Open access
- Author-friendly guidelines: 'your paper, your way'
- Reputable editorial board
- No word count or reference restrictions
- Double-blind review process to minimize bias
- Rapid production and publication
- Broad scope, interdisciplinary research exchange platform

The research areas that JCTR covers include, but are not limited to:

Internal medicine (all branches)	Gastroenterology and hepatology
Vascular medicine and phlebology	Surgery and transplantation
Oncology	Hematology
Cardiology	Nephrology
Intensive care medicine	Dermatology
Ophthalmology	Endocrinology and metabolism
Neurology and neurosciences	Anesthesiology
Anatomy, physiology, and embryology	Radiology and nuclear medicine
Pathology	Clinical chemistry
Clinical physics	Genetics and epigenetics
Epidemiology	Global health
Medical devices	Nutrition
Pharmacology	Immunology
Microbiology	Virology
Parasitology	Biomedical engineering

Biomedical spectroscopy and spectrometry

Thanks for considering the Journal of Clinical and Translational Research.

Editorial team JCTR

https://accscience.com/journal/JCTR

Contact
www.accscience.com
9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619
Email: editorial@accscience.com
Phone: +65 8182 1586