Poznan University of Medical Sciences
Poland

JMS Journal of Medical Science

previously Nowiny Lekarskie

Founded in 1889

2025 December
Vol. 94, No. 4

QUARTERLY

Journal Impact Factor 2024:
IF = 1.6

Indexed in:
Web of Science, DOAJ, Crossref,
Google Scholar, Polish Medical Bibliography,
Ministry of Education and Science

www.jms.ump.edu.pl

eISSN 2353-9801
ISSN 2353-9798
doi: 10.20883/ISSN.2353-9798

EDITOR-IN-CHIEF
Jarosław Walkowiak

ASSOCIATE EDITORS
Agnieszka Bienert
Ewa Mojs
Nadia Sawicka-Gutaj

EDITORIAL BOARD
David ADAMKIN, USA
Adrian BARANCHUK, Canada
Ewelina CHAWŁOWSKA, Poland
Judyta CIELECKA-PIONTEK, Poland
Jan DOMARADZKI, Poland
Piotr EDER, Poland
Michael GEKLE, Germany
Karl-Heinz HERZIG, Finland
Mihai IONAC, Romania
Paweł P. JAGODZIŃSKI, Poland
Jerzy JANKUN, USA
Lucian Petru JIGA, Germany
Berthold KOLETZKO, Germany
Stan KUTCHER, Canada
Talgat NURGOZHIN, Kazakhstan
Richard PRŮŠA, Czech Republic
Mitsuko SEKI, Japan
Puneet SINDHWANI, USA
Tomasz SZCZAPA, Poland
Jerzy P. SZAFLARSKI, USA
Dick TIBBOEL, Netherlands
Tomasz TORLIŃSKI, United Kingdom
Joanna TWAROWSKA-HAUSER, Poland
Dariusz WALKOWIAK, Poland
Przemysław ZALEWSKI, Poland

EXECUTIVE BOARD
Language editor
Sylwia Wiśniewska-Leśków, Poland

Statistical editor
Michał Michalak, Poland

Technical Editor and Typesetting
Bartłomiej Wąsiel, Poland

Publishing Manager: Grażyna Dromirecka

WYDAWNICTWO NAUKOWE
UNIWERSYTETU MEDYCZNEGO
IM. KAROLA MARCINKOWSKIEGO
W POZNANIU

60-812 Poznań, ul. Bukowska 70
tel./fax: +48 618547151
www.wydawnictwo.ump.edu.pl

Ark. wyd. 14,4. Ark. druk. 14,5
Zam. nr 9/26.

SECRETARIAT ADDRESS
27/33 Szpitalna Street, 60-572 Poznań, Poland
phone: +48 618491432, fax: +48 618472685
e-mail: jms@ump.edu.pl
www.jms.ump.edu.pl

DISTRIBUTION AND SUBSCRIPTIONS
70 Bukowska Street, 60-812 Poznań, Poland
phone/fax: +48 618547414
e-mail: sprzedazwydawnictw@ump.edu.pl

PUBLISHER
Poznan University of Medical Sciences
10 Fredry Street, 61-701 Poznań, Poland
phone: +48 618546000, fax: +48 618520455
www.ump.edu.pl

© 2025 by respective Author(s). Production and hosting by Journal of Medical Science (JMS)

This is an open access journal distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse

eISSN 2353-9801
ISSN 2353-9798
doi: 10.20883/ISSN.2353-9798

CONTENTS

ORIGINAL PAPERS

REVIEW PAPER

THE RATIONALE, DESIGN AND METHODS OF NEW STUDIES

Ethical guidelines

The Journal of Medical Science applies the ethical principles and procedures recommended by COPE (Committee on Conduct Ethics), contained in the Code of Conduct and Best Practice Guidelines for Journal Editors, Peer Reviewers and Authors available on the COPE website: https://publicationethics.org/resources/guidelines

ORIGINAL PAPER

JMS Journal of Medical Science

Evaluation of the Anticancer Potential of *S. euphratica*, *S. staminea*, and *S. trichoclada* in Colon Cancer: Monotherapy and Combination with Cisplatin

Aydın Sever
Department of Pharmacy Services, Vocational School of Health Services, Bingol University, Bingol, Türkiye
https://orcid.org/0000-0002-6727-1556
Corresponding author: asever@bingol.edu.tr

Mehmet Kadir Erdoğan
Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Bingol University, Bingol, Türkiye
https://orcid.org/0000-0002-1579-5737

Ramazan Gündoğdu
Department of Pharmacy Services, Vocational School of Health Services, Bingol University, Bingol, Türkiye
https://orcid.org/0000-0001-5230-2121

İbrahim Halil Geçibesler
Department of Occupational Health and Safety, Faculty of Health Science, Bingol University, Bingol, Türkiye
https://orcid.org/0000-0002-4473-2671

Lütfi Behçet
Department of Molecular Biology and Genetics, Faculty of Arts and Sciences, Bingol University, Bingol, Türkiye
https://orcid.org/0000-0001-8334-7816

https://doi.org/10.20883/medical.e1469

Keywords: phytotherapy, salvia, cancer

Received 2025-10-27
Accepted 2025-12-09
Published 2025-12-29

How to Cite: Sever A, Mehmet KE, Ramazan G, İbrahim HG, Lütfi B. Evaluation of the Anticancer Potential of *S. euphratica*, *S. staminea*, and *S. trichoclada* in Colon Cancer: Monotherapy and Combination with Cisplatin. Journal of Medical Science. 2025 December;94(4);e1469. https://doi.org/10.20883/medical.e1469

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Introduction. The use of herbal products in the treatment of colon cancer appears promising. Numerous *Salvia* species have been identified to exhibit cytotoxic properties. In this study, ethanolic extracts of three *Salvia* species collected from the Bingöl region (*Salvia euphratica, Salvia staminea, and Salvia trichoclada*), which have not been sufficiently investigated for their anticancer potential, were applied to the HT-29 colon cancer cell line both as monotherapies and in combination with cisplatin.

Material and methods. A crystal violet assay was performed to assess cell viability. Changes in cellular functions were evaluated using colony-formation and 3D soft agar assays for colony-forming ability, wound-healing assays for migration, and Western blot analyses for protein expression levels.

Results. The IC_{50} values were determined as 38.16, 40.85, and 43.40 µg/ml, respectively. In synergy assays with cisplatin, the combination indices were 9.54, 9.37, and 9.61, respectively. It was observed that the *Salvia* species alone reduced colony-forming ability in both 2D and 3D cultures, whereas this effect diminished when combined with cisplatin. The wound closure percentage decreased under combination treatment. Furthermore, although mTOR and AKT protein levels were reduced in combination treatments compared with monotherapy, p53 and c-PARP levels were elevated.

Conclusions. Although the examined *Salvia* species exhibited strong cytotoxic effects when used alone, these effects were attenuated when combined with cisplatin. Further investigations are needed to elucidate the underlying mechanisms of this phenomenon.

Introduction

Cancer is a disease in which pathophysiological changes occurring in the cell division process and DNA repair mechanisms are ineffective, causing the death of millions of individuals worldwide every year [1]. Colorectal cancer (CRC) is the third most commonly diagnosed cancer type in both men and women [2]. Of the 19.3 million new cancer cases observed in 2020, approximately 2 million were CRC. In the same year, nearly 900 thousand cancer deaths were attributed to the CRC group [1,3]. According to estimates for 2030, 1.1 million deaths and 2.2 million new cases related to CRC are expected [4]. Another study reports an estimated 3.2 million CRC-related deaths for the year 2040 [1]. Population growth and the proportion of older adults will likely increase CRC incidence [5].

Among the agents used in CRC chemotherapy, Cisplatin (Cis) forms cross-links by binding to DNA. These bonds damage DNA, leading to apoptosis in cells [6]. Although Cis is widely used for therapeutic purposes in many cancer types, it causes many undesirable outcomes and significant side effects, similar to other chemotherapeutic agents [7].

It has been reported that, compared with current colon cancer therapies, plant-based treatments cause fewer side effects and should be considered an alternative [8]. Studies have shown that some plant species support CRC treatment by inhibiting cell proliferation and inducing cell death [9]. The use of natural products together with synthetic drugs in CRC treatment has become a standard therapeutic approach in recent years. The protective and cytotoxic effects of herbal products on cancer cells are essential in this context [10]. It has been reported that the efficacy of chemotherapeutic agents used in CRC treatment is altered when they are combined with various plants [11]. *Salvia* species are used in multiple regions of the world in the prevention and treatment of many diseases [12]. Several studies have shown that *Salvia* species possess anticancer properties across various cell lines [13–15]. Previous phytochemical studies on these species have revealed that they are rich in rosmarinic acid, ferulic/vanillic acids, linalyl acetate, linalool, caryophyllene oxide, spathulenol, β-pinene, and related terpenoids [16–18]. In this study, the ethanol extracts of three different *Salvia* species collected from Bingöl province, which have not been sufficiently investigated in terms of anticancer properties *(Salvia euphratica* (SEE), *Salvia staminea* (SSE), *Salvia trichoclada* (STE)), were examined for their anticancer activity in a colon cancer cell line (HT-29) both as monotherapy and in combination with Cis.

Material and methods

Chemicals and reagents

Phosphate-buffered saline (PBS), Dulbecco's modified Eagle medium (DMEM), penicillin/streptomycin, trypsin-EDTA, and fetal bovine serum (FBS) were obtained from Gibco (Grand Island, NY, USA). Trypan blue, dimethyl sulfoxide (DMSO), methanol, acetic acid, and ethanol were obtained from Sigma Aldrich (St. Louis, MO, USA). Antibodies were obtained from Santa Cruz (Dallas, TX, USA). TEMED, Tris, KCl, NaCl, HCl, NaF, Na_3VO_4, Coomassie blue-G250, NaN_3, luminol, nonfat dry milk, Ponceau S, H_3PO_4, p-coumaric acid, PMSF, DTT, benzamidine, ethanol, glycerin, glycerol, sodium dodecyl sulfate (Sigma-Aldrich, Burlington, MA, USA); Tris HCl, DMSO, Tween-20, bromophenol blue, β-mercaptoethanol, NP-40, EDTA, EGTA, glycerophosphate, H_2O_2, NaOH, glycine (Merck, Darmstadt, Germany); 0.45 µm PVDF membrane (Millipore, Darmstadt, Germany); 25 and 75 cm² cell culture flasks (Sarstedt, Numbrecht, Germany); microcentrifuge tubes, 15 and 50 mL Eppendorf tubes (Isolab, Eschau, Germany); 6- and 96-well microplates (Corning, NY, USA); 5-, 10-, and 25-mL sterile pipettes and plastic pipette tips, 3- and 6-cm Petri dishes (Costar, Washington, DC, USA).

Plant extraction

Salvia species were collected during the flowering period from Bingöl, Türkiye. The identification was carried out by Prof. Dr Lütfi Behçet in accordance with the Flora of Turkey [19]. Specimens of the *Salvia* species were recorded in the Bingöl University Herbarium as *S. euphratica* (BIN 11635), *S. staminea* (BIN 11637), and *S. trichoclada* (BIN 11636). The aerial parts of the plants were ground using a laboratory-type grinder. From

each sample, 100 g were taken and extracted with ethanol in round-bottom flasks according to the method described by Erdoğan et al [20].

Cell culture

The HT-29 colon cancer cells used in the study were obtained from the Cancer Research Laboratory of Bingöl University. RPE-1 cells were used as the healthy cell line. All cell culture procedures were carried out in accordance with GCCP (Good Cell Culture Practice) aseptic standards. The cells were cultured in DMEM medium supplemented with 10% fetal bovine serum (FBS), 64 µg/mL penicillin, and 100 µg/mL streptomycin in a humidified cell culture incubator at 37°C containing 5–6% CO_2. Cells have been routinely screened for mycoplasma and other contaminants.

Cell viability

For the cell viability assay, the protocol described by Parker et al. was optimised and applied [21]. In summary, cells were seeded into 24-well plates at a density of 50,000 cells per well. After confirming cell attachment, 72-hour treatment was applied. Following the treatment, the DMEM medium was removed, and the surface was washed twice with PBS. Then, 1 mL of methanol-acetic acid (3:1) solution was added to each well, and the cells were fixed at room temperature for 5 minutes. The fixation solution was removed, and the wells were treated with 1 mL of 0.5% crystal violet dye. After 15 minutes of incubation at room temperature, the stained wells were carefully washed, and 1 mL of 10% acetic acid solution (prepared by mixing 1 mL of 100% acetic acid with 9 mL of ddH_2O) was added to each well and placed on a shaker for 5 minutes. Subsequently, the acetic acid from the wells was transferred to spectrophotometer cuvettes, and absorbance was measured at 595 nm to analyse cell viability. The data were then evaluated statistically.

Determining Synergy

To assess synergism between sage species and cis treatment in the HT-29 colon cancer cell line, the combination index (CI) was calculated. According to the calculation of the drug-sage extract interaction, CI > 1 indicates antagonism, CI = 1 indicates an additive effect, and CI < 1 indicates synergism [22]. The combination index was calculated using the formula reported in the literature [23]. The formula used in our analysis has now been clearly stated in the Methods section:

$$CI = \frac{(A_{CIS})_{50}}{(B_{SAL})_{50}} + \frac{(A_{SAL})_{50}}{(B_{CIS})_{50}}$$

In this equation:

- $(A_{CIS})_{50}$: cisplatin concentration required to reduce viability to 50% when *Salvia* extract is applied at $IC_{50}/2$,
- $(B_{SAL})_{50}$: IC_{50} of *Salvia* extract in monotherapy,
- $(A_{SAL})_{50}$: *Salvia* concentration needed to reduce viability to 50% when cisplatin is applied at $IC_{50}/2$,
- $(B_{CIS})_{50}$: IC_{50} of cisplatin in monotherapy.

Colony formation

The colony formation assay was performed to assess the effects of various treatments on cell colony-forming ability. For this assay, cells were seeded into 6-well plates at a density of 5 × 10^2 cells per well. The following day, after confirming cell attachment, the cells were treated with *Salvia* extracts and Cis for 72 hours. After treatment, the medium was removed, and the cells were incubated in fresh medium for 15 days (37 °C, 5% CO_2, 95% humidity). After this period, the medium was removed, the cells were washed with PBS, fixed in methanol: acetic acid (3:1) for 5 minutes, and stained with crystal violet for 15 minutes. The wells were then carefully washed, and the colonies were counted under a microscope. A formation of at least fifty cells was considered a colony [24].

Wound healing

The in vitro wound-healing assay was performed as modified from the method described by Zomer et al. [25]. Cells were seeded into 6-well plates, and the next day, a scratch was created using a sterile 100 µL pipette tip. The detached cells were removed by washing twice with PBS. Afterwards, photographs of the wounds were taken. The treatment group was then exposed to the respective plant extracts and chemotherapeutic agents. At 0, 6, 12, and 24 hours, the wound areas were photographed under a microscope and measured using ImageJ software. The percent-

age of wound closure for each group was calculated by normalising images acquired at 0, 6, 12, and 24 hours.

3D Soft Agar

The effect of the applied treatments on the colony-forming ability of the cells was evaluated using the soft agar assay [26]. To each well of the 6-well plates, 1 mL of agar-medium mixture was added, and the plates were incubated in a culture cabinet for 30 minutes to allow the bottom agar layer to solidify. A 0.5% agarose solution containing 6×10^3 cells and the respective treatment agents in cell culture medium (DMEM, 20% FBS) was mixed at a 1:1 ratio (final agarose concentration 0.5%) and incubated in a 36°C water bath for 15 minutes. Then, 1 mL of the agarose-cell-medium mixture (6×10^3 cells) was carefully overlaid on the solidified 0.5% agarose bottom layer. The cells were incubated for 7–21 days in a humidified cell culture incubator at 37°C with 5% CO_2, and every 3–4 days, fresh medium containing treatment agents was added onto the top agar layer. At the end of the experiment, colony formation was examined under a microscope, and the formed colonies were stained with 0.005% crystal violet prepared in methanol for 15 minutes. The experiment was performed with at least two replicates for each group, including the control.

Western blotting

Western blotting was used to assess changes in protein expression in cells treated with the applied treatments [24]. For protein sample analysis, a vertical SDS-PAGE tank system (Bio-Rad Mini-PROTEAN Tetra Cell) was used. Cells seeded in 6-cm wells were washed twice with cold PBS (without Ca^{2+}/Mg^{2+}) to remove them from the plates after 72 hours of treatment. The collected cells were suspended in PBS, centrifuged at 1100 RPM for 5 minutes at room temperature, and the supernatant was partially aspirated. The obtained pellet was resuspended in an adequate amount of lysis buffer (100 µL for 0.5×10^6 cells), and effective homogenization and protein extraction were achieved using a 1 mL pipette. The samples were incubated on ice for approximately 1 hour, then centrifuged at 15,000 RPM for 10 minutes at 4°C. The resulting supernatants were transferred into pre-prepared, labelled microcentrifuge tubes. Total protein concentration was determined using the Bradford method. According to the calculations, equal amounts of protein (30 µg per lane) were loaded into each well of the SDS-PAGE gel and electrophoresed. Protein concentrations were determined manually using the Bradford reagent with Coomassie Brilliant Blue G-250, and equal loading was confirmed with Tubulin as a loading control. To prevent protein degradation, the lysis buffer was supplemented with protease inhibitors and reducing agents, specifically 1 mM DTT, one mM Benzamidine, 0.5 mM PMSF, and 1 mM Leupeptin. The proteins were then transferred onto PVDF (polyvinylidene fluoride) membranes and incubated for 1 hour in blocking buffer (5% milk powder in TBS-T). After blocking, the membranes were incubated overnight at 4°C with the appropriate primary antibody (diluted in 5% milk powder) with rotation. The membranes were washed with TBS-T (3 × 10 minutes) and then incubated for 2 hours at room temperature with the relevant HRP-conjugated secondary antibody (diluted in 5% milk powder). The membranes were rewashed with TBS-T (3 × 10 minutes). Detection was performed using a chemiluminescence-based ECL kit. The signals on the membrane were transferred to film using an X-ray imaging system. Finally, densitometric analysis of the band intensities was performed using ImageJ software (National Institutes of Health, USA).

DNA Agarose Gel Assay

One method for visualising DNA damage and breaks is the DNA agarose gel assay. DNA fragments placed in the gel under an electric field are separated according to their sizes [27]. A. A sufficient amount of agarose and Tris-acetate (TAE) buffer was mixed and gently shaken. It was then heated in a microwave until it began to boil. At intervals, the flask was removed from the microwave and mixed. After complete dissolution, Ethidium Bromide (EtBr) was added, and the solution was shaken to ensure uniform dispersion. The cooled solution was slowly poured into a tray. It was left to polymerise. Once polymerisation was complete, the gel was placed in the tank, and TAE was added until it covered the gel. Samples were mixed with five µL of loading buffer and loaded into the gel. A marker was also loaded into one well. Electrophoresis was performed at 80 V for 20 minutes.

Statistical analysis

All experiments were performed in triplicate unless stated otherwise. Graphical and statistical analyses were conducted using GraphPad Prism (GraphPad Software, USA). The obtained data were analysed using Tukey's multiple comparisons test and One-Way ANOVA. A p-value of < 0.05 was considered statistically significant.

Results and Discussion

Since ancient times, plants have been the focus of attention in healthcare. Herbal products are used in the treatment of numerous diseases within traditional medicine. They are sometimes used in the diet and sometimes as herbal medication [28,29]. Colon cancer is a significant health issue affecting human health, as it is the third most common cause of cancer-related deaths in both men and women [30]. Cisplatin (Cis), a drug commonly used in the chemotherapy of many cancers, including CRC, exerts its effects by inducing oxidative stress and DNA damage in cells [31]. In addition to chemotherapeutic agents, plant-derived products that have long been used in cancer treatment are available. Plants can not only prevent cancer but also enhance the efficacy of other cancer treatments [32]. It is predicted that the discovery of new plants may yield greater benefits in cancer treatment [33].

Cell Viability Findings

Salvia species have been reported to possess anticancer properties across multiple cancer types, including CRC, owing to their bioactive constituents (see **Figure 1**). *Salvia* and its derived compounds have been reported to exhibit anticancer effects in various cancer cell lines [34–37]. High *levels* of phenolic compounds and antioxidant properties characterise the Salvia species used in this study [16–18]. In our study, consistent with the literature, the *Salvia* species tested exhibited cytotoxic effects on the HT-29 cell line, with vigorous cytotoxic activity observed at 40 µg/mL and above. The obtained IC_{50} and $IC_{50}/2$

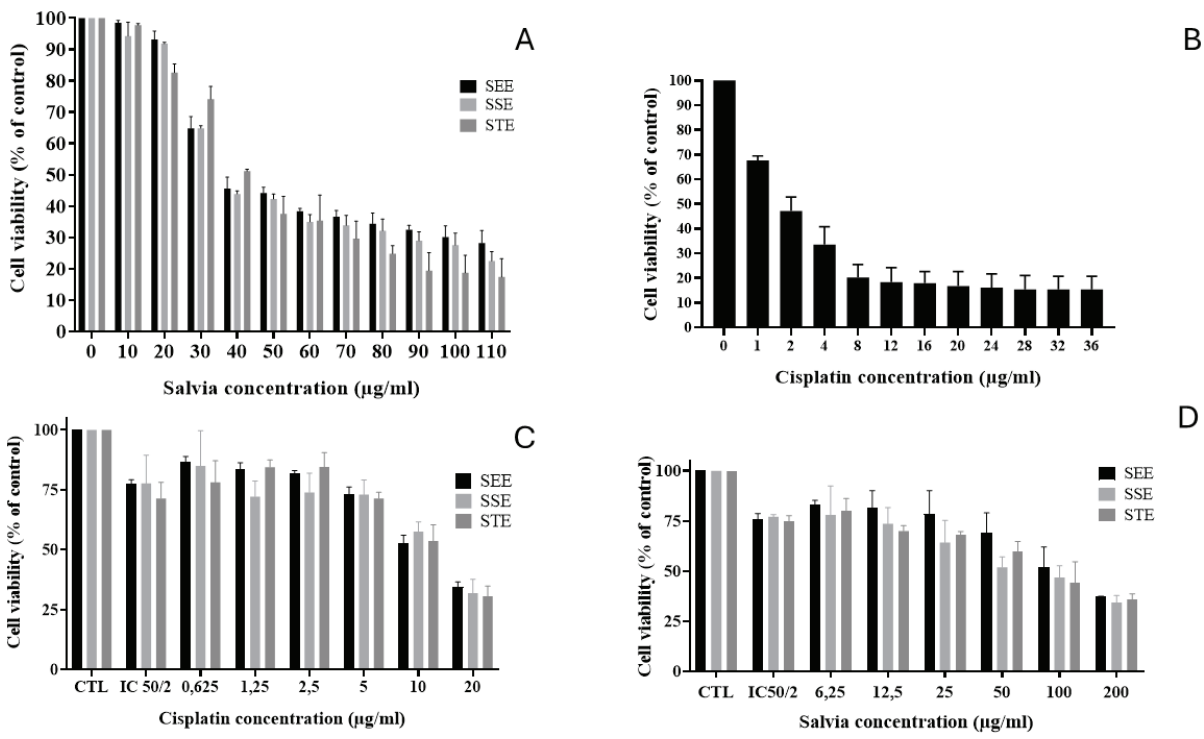


Figure 1. Viability graph. Cell viability of HT-29 cells treated with *Salvia* species for 72 hours at increasing concentrations (A), cell viability of HT-29 cells treated with cisplatin for 72 hours at increasing concentrations (B), viability assay with increasing cisplatin concentrations while keeping the *Salvia* dose constant (C), viability assay with increasing *Salvia* concentrations while keeping the cisplatin dose constant (D). (n=3, SD, ANOVA, p<0.05). Source: Author's own data.

doses are presented in **Table 1**. To assess the selectivity of the cytotoxic effects, the extracts were also tested on the healthy human retinal pigment epithelial (RPE-1) cell line. The IC_{50} values for SEE, SSE, and STE in RPE-1 cells were 89.67, 95.15, and 89.58 µg/ml, respectively. These values are approximately 2-2.5 times higher than those observed in HT-29 cancer cells, indicating a favourable selectivity index towards the colon cancer cells.

Cis is an important chemotherapeutic agent used in the treatment of many types of cancer. However, its use is associated with significant side effects. Using Cis in combination with plants is a critical approach to reducing these side effects [38]. Combining plants with chemotherapeutic agents holds promise for mitigating the adverse effects of chemotherapy. The use of natural compounds to reduce the side effects of current CRC therapies has been reported to have multiple beneficial effects [39]. However, in some cases, natural products can be ineffective in cancer treatment; in certain situations, they may even reduce the efficacy of the drug used [40,41]. Studies indicate that components exhibiting anti-inflammatory properties reduce the activity of cis [42]. It is emphasised that, owing to their high antioxidant capacity, *Salvia* species may reduce the efficacy of chemotherapeutic agents that act through oxidative stress. In a previous study, rosmarinic acid derived from *Salvia* species was found to diminish the effect of Cis [43]. In our study, we observed that the cytotoxic effect decreased when Cis and *Salvia* species were combined. In the synergism experiments, IC_{50} values increased. This phenomenon may be related to the reduction of Cis's oxidative effects, one of its essential mechanisms of action, by *Salvia* species with high antioxidant properties. The IC_{50} values and combination index (CI) results obtained from the synergism experiments are summarised in **Table 2**.

Colony Formation Findings

The colony formation assay is a technique used to assess the antiproliferative effect of treatments applied in cell culture [44]. Various plants have been shown to reduce the colony-forming ability of cancer cells. *Salvia* species and their derived compounds have been shown to inhibit colony formation in many cell lines [34,45–48]. Consistent with the literature, our study showed that Salvia *treatment* reduced the colony-forming ability of the HT-29 cell line. Although plant-derived products often show synergistic effects with many chemotherapeutic agents, antagonistic effects can also occur in some cases [49]. The combined use of Cis and *Salvia* species was observed to reduce colony-forming ability less than monotherapy ($p < 0.05$) (see **Figure 2**).

Table 1. $IC_{50}/2$ and IC_{50} values obtained after the viability assays. Source: Author's own data.

	$IC_{50}/2$	IC_{50}
CİS	1,43 µg/ml	2,92 µg/ml
SEE	18,21 µg/ml	38,16 µg/ml
SSE	19,89 µg/ml	40,85 µg/ml
STE	23,86 µg/ml	43,40 µg/ml

Table 2. IC_{50} values and combination indexes obtained in the synergism experiments. Source: Author's own data.

	IC_{50}	Combination Index (CI)
SEE $IC_{50}/2$+Cis	16,1	9,54
Cis $IC_{50}/2$+SEE	154	
SS $IC_{50}/2$+Cis	17,1	9,37
Cis $IC_{50}/2$+SSE	144	
ST $IC_{50}/2$+Cis	18	9,61
Cis $IC_{50}/2$+STE	150	

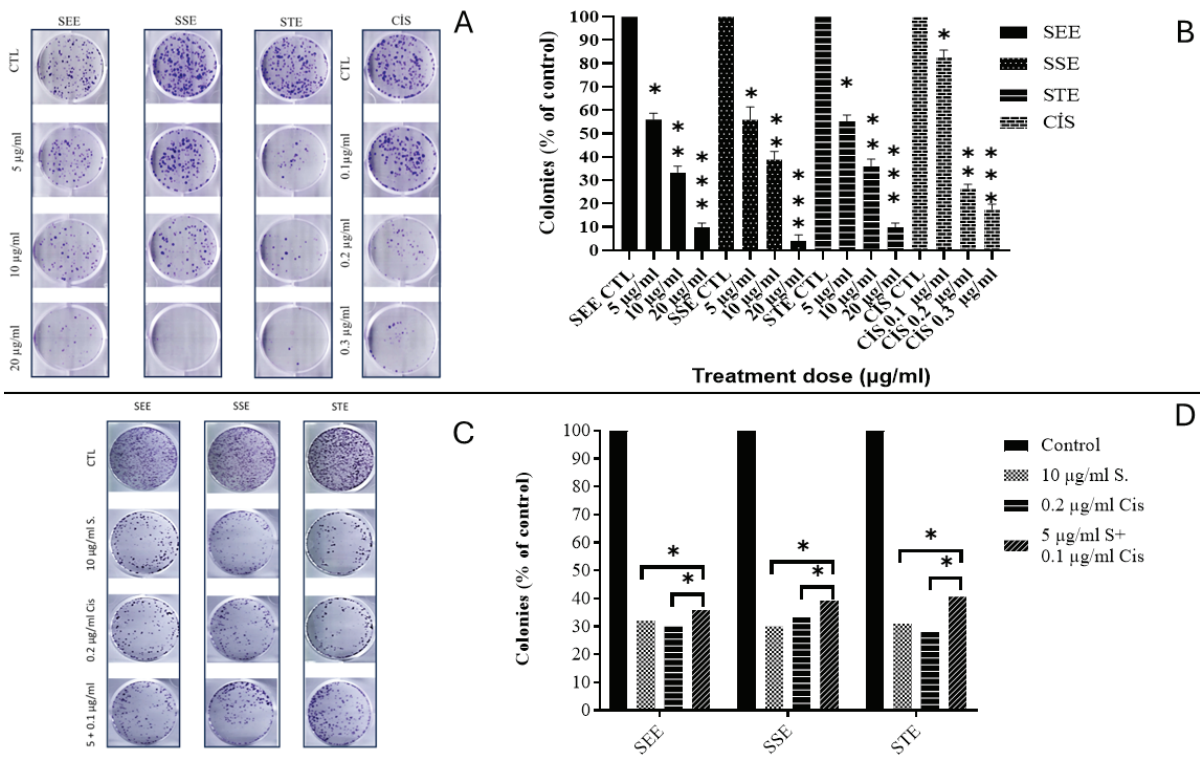


Figure 2. After seeding HT-29 cells into 6-well plates, they were treated with the determined IC_{50} and $IC_{50}/2$ doses of Cis and *Salvia* species. For monotherapy, IC_{50} doses were used, while for combination treatments, $IC_{50}/2$ + $IC_{50}/2$ doses were applied (n = 3, SD, ANOVA, $p < 0.05$). Source: Author's own data. * $p < 0.05$

Soft Agar Findings

Given the difficulty of monitoring tumour development in vivo, the soft agar technique remains an essential method for measuring cancer cell proliferation in a semi-solid environment [50]. Studies have shown that various *Salvia* species reduce the colony-forming ability of cancer cell lines in soft agar [48,51–54]. Consistent with the literature, our study demonstrated that the *Salvia* species used decreased the colony-forming ability of cells in soft agar when used as monotherapy, but that combining it with Cis reduced the effectiveness of Cis (see **Figure 3**).

Wound Healing Findings

The wound-healing assay is a standard in vitro technique used to analyse 2D cell migration. Cells migrate into a gap created mechanically, thermally, or chemically, and the percentage of gap closure is examined under a microscope [55]. Treatments with *Salvia* species have been shown to reduce migration ability in various cell lines [37,56,57]. In our study, treatment with *Salvia* species reduced the migration ability of HT-29 cells. As observed in the viability and colony formation assays, the effectiveness of the combination therapy was reduced (see **Figure 4**).

Normalised percentage graph of wound closure areas captured at 6, 12, and 24 hours (*$p < 0.05$; compared to control, **$p < 0.01$; compared to control, $^{\#}p < 0.05$; compared to the group treated with Cis only, $^{\#\#}p < 0.01$; compared to the group treated with Cis only).

Western blot findings

It is well established that many cancer therapies increase p53 levels. In colon cancer cell lines, 5-FU and various plant extracts have been shown to elevate p53 levels [58]. Similarly, various *Salvia* species have been reported to increase p53 protein levels in cells [59,60]. In our study, p53 levels were significantly higher in all groups treated with *Salvia* species and cisplatin for 72 hours than in the control group (see **Figure 5**). The p-values were $p < 0.001$ for the cisplatin group and $p < 0.05$ for the *Salvia* groups. In combination therapy groups, p53 protein levels were significantly lower compared to the cisplatin group ($p < 0.001$). Given the role of p53 in apoptotic processes and cellular stress, these

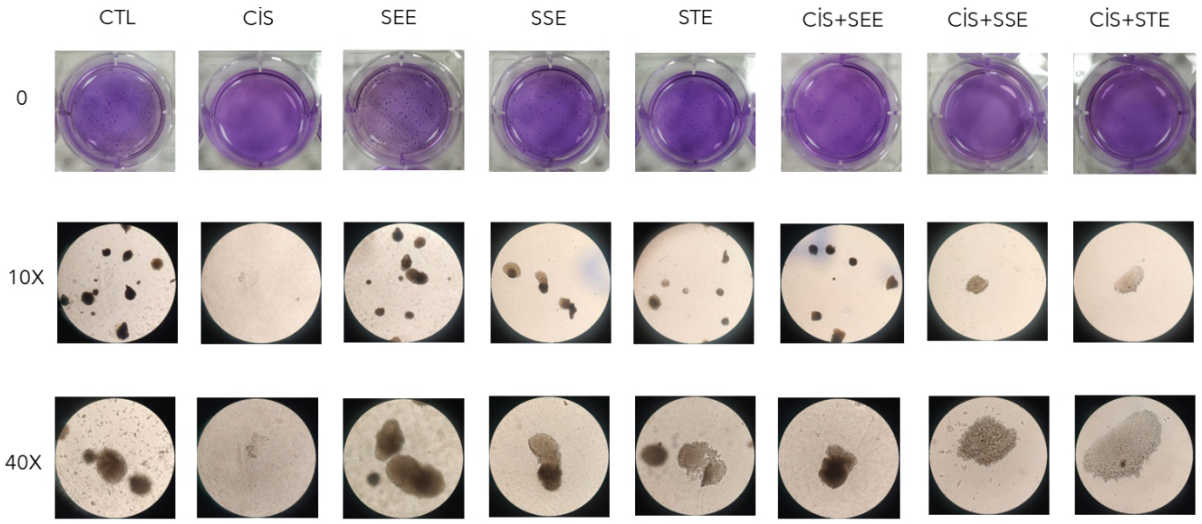


Figure 3. Microscopic images of the Soft Agar assay performed with HT-29 cells. Cells seeded in 6-cm well plates were treated with IC50 doses for monotherapy and IC25 + IC25 doses for combination therapies. During the 21-day experimental period, DMEM supplemented with 20% FBS, without treatment agents, was added to the plates at 5-day intervals. At the end of day 21, the colonies were stained as described in the Materials and Methods section to visualise their formation. Images of the plates were captured using an external camera at 10× and 40× magnification. (n=3). Source: Author's own data.

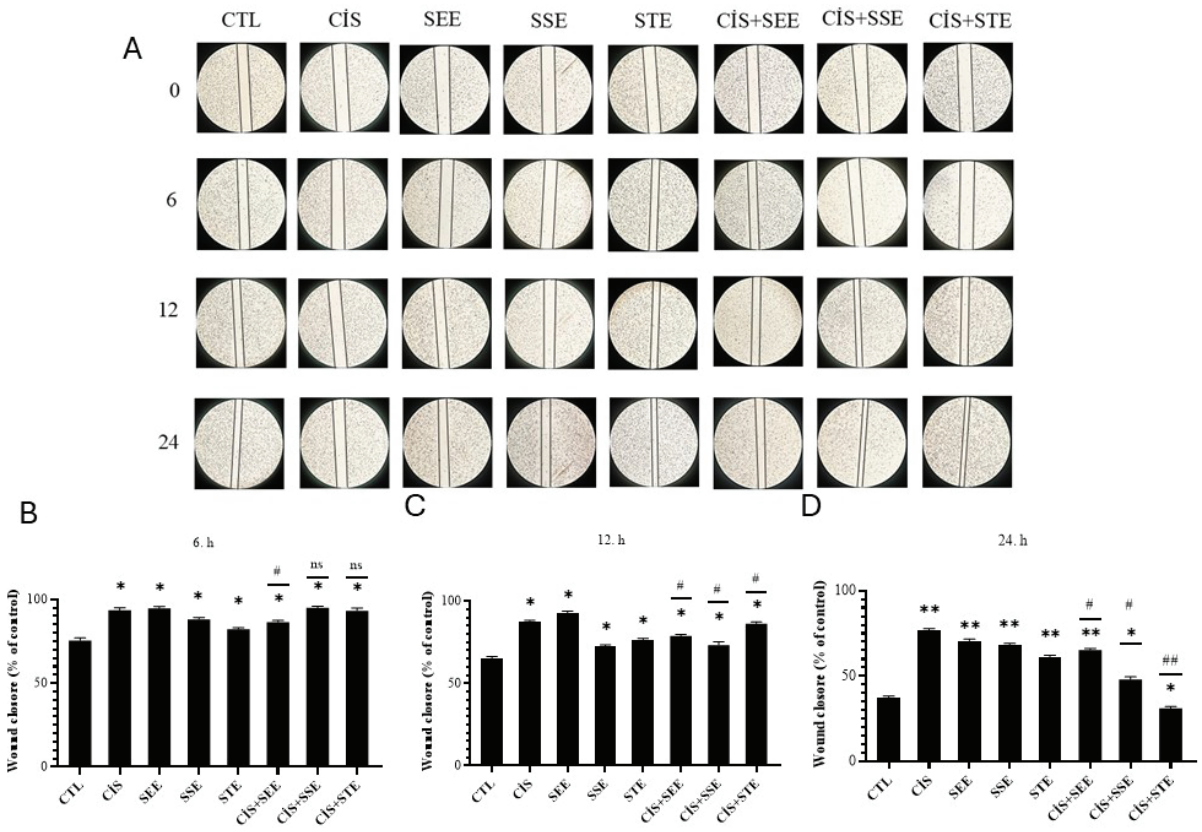


Figure 4. Microscopic images of the wound areas. HT-29 cells were seeded into 6-well plates. After ensuring cell attachment and sufficient confluency, scratches were created using a sterile pipette tip. Treatments were applied using the IC_{50} and IC_{50}/2 doses obtained from the viability assays. Images were captured under a microscope at 0, 6, 12, and 24 hours. (n=3, SD, ANOVA, p<0.05). Source: Author's own data. *p< 0.05, **p< 0.01, $^{\#}$p<0.01, $^{\#\#}$p<0.01

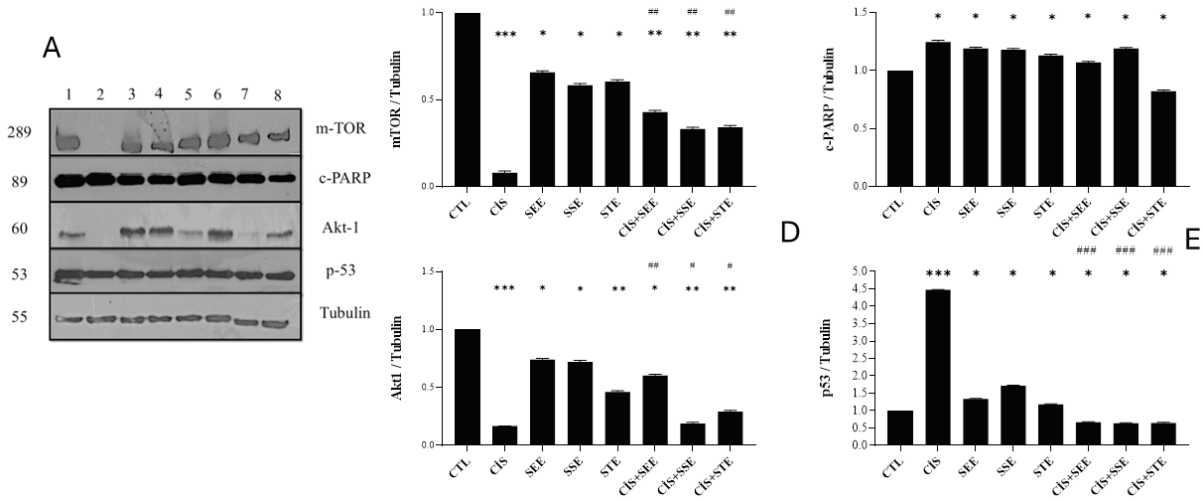


Figure 5. Band images of protein levels in HT-29 cells after treatment (A), quantitative graphs showing the expression levels of mTOR (B), c-PARP (C), Akt1 (D), and p53 (E). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: $p > 0.05$, compared to control. # $p < 0.05$, ## $p < 0.01$, ### $p < 0.001$, compared to cisplatin-treated cells. (n=3, SD, ANOVA, $p<0.05$) Source: Author's own data.

results are consistent with the literature, which shows that *Salvia* species and cisplatin induce stress and apoptosis when used alone in colon cancer cell lines. However, monotherapy appears to induce more stress and apoptosis compared to combination therapy.

c-PARP, a form of PARP-1, is associated with the late stage of apoptotic cell death [61]. As a biomarker in apoptosis, c-PARP levels have been found to increase in parallel with apoptotic processes [62]. Many chemotherapeutic agents and plant-derived products, including *Salvia* species, are known to elevate c-PARP levels [59,63]. Consistent with this information, c-PARP levels increased with monotherapies in our study ($p < 0.05$). In combination therapies, a significant decrease was observed relative to cisplatin treatment, indicating that DNA damage induced by monotherapy is reduced in combination therapy.

Akt, a central molecule in the PI3K-Akt signalling pathway, has been associated with multiple cancer types when abnormally activated. Cancer cell migration is closely linked to Akt, which transmits signals from growth factors and oncogenes to downstream targets [64,65]. Chemotherapeutic agents, *Salvia* species, and their derived compounds have been reported to reduce Akt1 protein levels in cancer cells [66–68]. In our study, monotherapy significantly decreased Akt1 levels, consistent with the literature. However, combination therapy showed a significant increase compared to cisplatin treatment.

mTOR is a protein that regulates cell growth, mRNA translation, ribosome biogenesis, autophagy, and metabolism, and it is hyperactivated in many cancer types [69]. Both chemotherapy and various plants, including *Salvia*, are known to reduce mTOR levels in cancer cells [66,70]. In our study, treatment with cisplatin and *Salvia* significantly decreased mTOR levels. However, combination therapy showed an increase relative to monotherapy, indicating reduced cytotoxicity in the combination.

DNA agarose gel findings

Kemoterapik Chemotherapeutic agents and substances with anticancer properties target cancer cells by inducing DNA damage in rapidly dividing cells, either directly or indirectly [71]. The cellular response to the resulting DNA damage is closely related to both cancer progression and the outcomes observed after chemotherapy. Modulating the DNA damage response system is considered an essential target for enhancing treatment efficacy in cancer therapy [72]. Various *Salvia* species are known to induce DNA damage in cancer cell lines [36,73,74]. In our study, DNA fragmentation was markedly more pronounced in monotherapies with cisplatin and *Salvia* species than

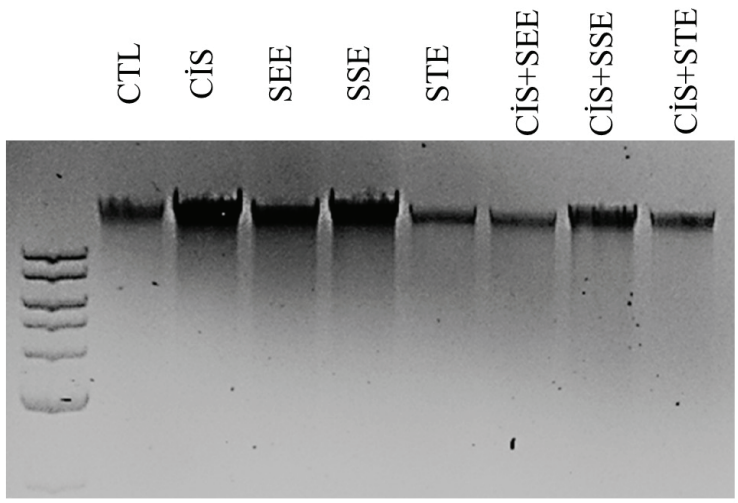


Figure 6. DNA Agarose Gel Image of HT-29 cells treated with Cis, SEE, SSE, STE, Cis+SEE, Cis+SSE, and Cis+STE. HT-29 cells were treated for 72 hours with IC_{50} and $IC_{50}/2$ doses and then subjected to an electric field using a vertical system. (n = 3). Source: Author's own data.

in the control group (see **Figure 6**). However, in combination therapy, fragmentation was less pronounced compared to the cisplatin group. This indicates that monotherapy alone induces greater DNA damage. The reduced DNA damage observed in combination treatments suggests that the plants may decrease the efficacy of cisplatin.

Conclusion

In conclusion, the studied *Salvia* species exhibited cytotoxic effects on HT-29 cell lines. Following treatment with *Salvia* species, colony formation and migration abilities of the cells decreased; levels of proteins associated with apoptosis, autophagy, and DNA damage (p53, c-PARP) increased; and the level of mTOR, which is related to cell proliferation, decreased. In combination treatments, results were less effective than those of monotherapy, indicating that the combination reduced the efficacy of cisplatin. Based on these findings, an antagonistic interaction between the studied *Salvia* species and cisplatin was observed. These results can be further supported by future *in vitro* and *in vivo* studies, which will provide clearer insights into the use of these plants in cancer therapy.

Disclosures

Conflict of interest

The authors declare no conflicts of interest.

Authors' Contributions

Lütfi Behçet collected the plants. İbrahim Halil Geçibesler performed the plant extractions. Mehmet Kadir Erdoğan conducted the cell viability and colony formation assays. Ramazan Gündoğdu performed the wound healing assays. Aydın Sever carried out the soft agar, Western blot, and DNA agarose experiments. Each author wrote the sections corresponding to their respective experimental work. All authors approve the final version of the manuscript.

Data Availability Statement (DAS)

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Statement

Ethical approval was not required for this study as it did not involve human participants or animals.

Funding Statement

This study was supported by Bingöl University, grant number BAP-FEF.2023.005.

References

1. Chhikara BS, Parang K. Global Cancer Statistics 2022: the trends projection analysis. Chem Biol Lett. 2023;10(1):451. Available from: https://pubs.thescienceincein.org/journal/index.php/cbl/article/view/451
2. Thanikachalam K, Khan G. Colorectal cancer and nutrition. Nutrients. 2019;11(1):164. https://doi.org/10.3390/nu11010164
3. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi DJ, John A, Lim YC, Kibria KMK, Mohiuddin AKM, Ming LG. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. Cancers (Basel). 2022;14(7):1732. https://doi.org/10.3390/cancers14071732
4. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. CA Cancer J Clin. 2023;73(3):233–54. https://doi.org/10.3322/caac.21772.
5. Araghi M, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, Arnold M. Global trends in colorectal cancer mortality: projections to the year 2035. Int J Cancer. 2019;144(12):2992–3000. https://doi.org/10.1002/ijc.32055
6. Ghosh S. Cisplatin: the first metal based anticancer drug. Bioorg Chem. 2019;88:102925. https://doi.org/10.1016/j.bioorg.2019.102925
7. Makovec T. Cisplatin and beyond: molecular mechanisms of action and drug resistance development in cancer chemotherapy. Radiol Oncol. 2019;53(2):148. https://doi.org/10.2478/raon-2019-0018
8. Asadi-Samani M, Kooti W, Aslani E, Shirzad H. A systematic review of ıran's medicinal plants with anticancer effects. J Evidence-Based Complement Altern Med. 2016;21(2):143–53. https://doi.org/10.1177/2156587215600873
9. Xavier CPR, Pereira-Wilson C. Medicinal plants of the genuses Salvia and Hypericum are sources of anticolon cancer compounds: Effects on PI3K/Akt and MAP kinases pathways. PharmaNutrition. 2016;4(2):112–22. https://doi.org/10.1016/j.phanu.2015.11.002
10. Hien NM, Nguyen TYN, Le NTH, Le TNT, Chau NTN, Le TMH, Nguyen BQH. Medicinal plants as a potential resource for the discovery of novel structures towards cancer drug resistance treatment. Heliyon. 2024. https://doi.org/10.1016/j.heliyon.2024.e39229
11. Jung T, Cheon C. Synergistic and additive effects of herbal medicines in combination with chemotherapeutics: a scoping review. Integr Cancer Ther. 2024;23:15347354241259416. https://doi.org/10.1177/15347354241259416
12. Al-Jaber HI. Essential oil composition, total phenolic and total flavonoid content and in-vitro antioxidant properties of hydro-alcoholic and water extracts of salvia deserti growing wild in jordan. Jordan J Chem. 2017;12(1):11–9. wosuid: wos:000410352300002
13. Choudhary MI, Hussain A, Ali Z, Adhikari A, Sattar SA, Ayatollahi SAM, Rahman A, Majid A. Diterpenoids including a novel dimeric conjugate from Salvia leriaefolia. Planta Med. 2012;78(3):269–75. https://doi.org/10.1055/s-0031-1280454
14. Kamatou GPP, Makunga NP, Ramogola WPN, Viljoen AM. South African Salvia species: A review of biological activities and phytochemistry. J Ethnopharmacol. 2008 Oct 28;119(3):664–72. https://doi.org/10.1016/j.jep.2008.06.030
15. Moridi Farimani M, Miran M. Labdane diterpenoids from Salvia reuterana. Phytochemistry. 2014;108:264–9. https://doi.org/10.1016/j.phytochem.2014.08.024
16. Zengin G, Llorent-Martínez EJ, Fernández-de Córdova ML, Bahadori MB, Mocan A, Locatelli M, Aktumsek A. Chemical composition and biological activities of extracts from three Salvia species: S. blepharochlaena, S. euphratica var. leiocalycina, and S. verticillata subsp. amasiaca. Ind Crops Prod. 2018;111:11–21. doi:10.1016/j.indcrop.2017.09.065
17. Çolak NU, Yıldırım S, Bozdeveci A, Yaylı N, Çoşkunçelebi K, Fandaklı S, Yaşar A. Essential oil composition, antimicrobial and antioxidant activities of Salvia staminea. Rec. Nat. Prod. 2018. https://doi.org/10.25135/rnp.08.17.03.013
18. Kilic O. Chemical composition of four Salvia L. species from Turkey: A chemotaxonomic approach. J Essent Oil Bear Plants. 2016;19(1):229–35. https://doi.org/10.1080/0972060X.2014.958560
19. Hedge IC. Salvia l. Flora Turkey East Aegean islands. 1982;7:400–61.
20. Erdogan MK, Toy Y, Gundogdu R, Gecibesler IH, Sever A, Yapar Y, Behçet L, Zengin G. Assessment of cytotoxic, apoptotic, enzyme inhibitory, and antioxidant properties, and phytochemical characterization of ethanolic extract from Cionura erecta. Food Biosci. 2025;65:106082. https://doi.org/10.1016/j.fbio.2025.106082
21. Parker C, Chambers AC, Flanagan DJ, Ho JWY, Collard TJ, Ngo G, et al. BCL-3 loss sensitises colorectal cancer cells to DNA damage by targeting homologous recombination. DNA Repair (Amst). 2022;115:103331. https://doi.org/10.1016/j.dnarep.2022.103331
22. Chou TC. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. Pharmacol Rev. 2006;58(3):621–81. https://doi.org/10.1124/pr.58.3.10
23. Li J, Wang R, Kong Y, Broman MM, Carlock C, Chen L, Li Z, Farah E, Ratliff TL, Liu X. Targeting Plk1 to enhance efficacy of olaparib in castration-resistant prostate cancer. Mol Cancer Ther. 2017;16(3):469–79. https://doi.org/10.1158/1535-7163.MCT-16-0361
24. Gomez V, Gundogdu R, Gomez M, Hoa L, Panchal N, O'Driscoll M, Hergovich A. Regulation of DNA damage responses and cell cycle progression by hMOB2. Cell Signal. 2015;27(2):326–39. https://doi.org/10.1016/j.cellsig.2014.11.016
25. Zomer HD, Varela GK dos S, Delben PB, Heck D, Jeremias T da S, Trentin AG. In vitro comparative study of human mesenchymal stromal cells from dermis and adipose tissue for application in skin wound healing. J Tissue Eng Regen Med. 2019;13(5):729–41. https://doi.org/10.1002/term.2820
26. Bettoun A, Joffre C, Zago G, Surdez D, Vallerand D, Gundogdu R, Sharif AAD, Gomez M, Cascone I, Meunier B, White MA, Codogno P, Parrini MC, Camonis JH, Hergovich A. Mitochondrial clearance by the STK38 kinase supports oncogenic Ras-induced cell

transformation. Oncotarget. 2016 ;7(28):44142–60. https://doi.org/10.18632/oncotarget.9875
27. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. JoVE (Journal Vis Exp. 2012;(62):e3923. https://doi.org/10.3791/3923
28. Balewska A, Szczechla M. Plants: past and present in the battle against diabetes. J Med Sci. 2024;93(1):e896–e896. https://doi.org/10.20883/medical.e896
29. Monika K, Anna B, Jana K, Daniel P, Bogusław C. Products of plant origin–benefits and the potential risk for the consumer. J Med Sci. 2017;86(1). https://doi.org/10.20883/jms.2017.143
30. Okolie NP, Olude OM. Isolation of some bioactive compounds in the methanol extract of ficus exasperata leaves and the effect of the extract on ınflammatory markers in 1, 2 dimethylhydrazine ınduced colorectal cancer in rats. J Med Sci. 2025;94(1):e1232–e1232. https://doi.org/10.20883/medical.e1232
31. Tang C, Livingston MJ, Safirstein R, Dong Z. Cisplatin nephrotoxicity: new insights and therapeutic implications. Nat Rev Nephrol 2022 191 [Internet]. 2022 [cited 2023 Jan 3];19(1):53–72. https://doi.org/10.1038/s41581-022-00631-7
32. Ahmed M, Khan MI, Khan MR, Muhammad N, Khan AU, Khan RA. Role of medicinal plants in oxidative stress and cancer. Open Access Sci Reports. 2013;2(2):2–4. doi:10.4172/scientificreports.641
33. David B, Wolfender JL, Dias DA. The pharmaceutical industry and natural products: historical status and new trends. Phytochem Rev. 2015;14(2):299–315. https://doi.org/10.1007/s11101-014-9367-z
34. Bai J, Qin Q, Li S, Cui X, Zhong Y, Yang L, An L, Deng D, Zhao J, Zhang R, Bai S. Salvia miltiorrhiza inhibited lung cancer through aerobic glycolysis suppression. J Ethnopharmacol. 2024;331. https://doi.org/10.1016/j.jep.2024.118281
35. Nasrabadi NS, Vedad A, Asadi K, Poorbagher MRM, Tabrizi NA, Dorooki K, Sabouni RS, Moghadam MB, Shafaei N, Karimi E, Oskoueian E. Nanoliposome-loaded phenolics from Salvia leriifolia Benth and its anticancer effects against induced colorectal cancer in mice. Biotechnol Appl Biochem. 2024;(71):641–50. https://doi.org/10.1002/bab.2564
36. Nicolella HD, Fernandes G, Ozelin SD, Rinaldi-Neto F, Ribeiro AB, Furtado RA, Senedese JM, Esperandim TR, Veneziani RCS, Tavares DC Manool, a diterpene from Salvia officinalis, exerts preventive effects on chromosomal damage and preneoplastic lesions. Mutagenesis. 2021;36(2):177–85. https://doi.org/10.1093/mutage/geab001
37. Keshavarz M, Bidmeshkipour A, Mostafaie A, Mansouri K, Mohammadi-Motlagh H. Anti tumor activity of Salvia officinalis is due to its anti-angiogenic, anti-migratory and anti-proliferative effects. Cell J. 2011;12(4):477–82.
38. Dasari S, Njiki S, Mbemi A, Yedjou CG, Tchounwou PB. Pharmacological effects of cisplatin combination with natural products in cancer chemotherapy. Int J Mol Sci. 2022;23(3):1532. https://doi.org/10.3390/ijms23031532
39. Rejhová A, Opattová A, Čumová A, Slíva D, Vodička P. Natural compounds and combination therapy in colorectal cancer treatment. Eur J Med Chem. 2018;144:582–94. https://doi.org/10.1016/j.ejmech.2017.12.039
40. Ju YH, Doerge DR, Woodling KA, Hartman JA, Kwak J, Helferich WG. Dietary genistein negates the inhibitory effect of letrozole on the growth of aromatase-expressing estrogen-dependent human breast cancer cells (MCF-7Ca) in vivo. Carcinogenesis. 2008;29(11):2162–8. https://doi.org/10.1093/carcin/bgn161
41. Heaney ML, Gardner JR, Karasavvas N, Golde DW, Scheinberg DA, Smith EA, O'Connor OA. Vitamin C antagonizes the cytotoxic effects of antineoplastic drugs. Cancer Res. 2008;68(19):8031–8. https://doi.org/10.1158/0008-5472.CAN-08-1490
42. Dickey DT, Wu YJ, Muldoon LL, Neuwelt EA. Protection against cisplatin-induced toxicities by N-acetylcysteine and sodium thiosulfate as assessed at the molecular, cellular, and in vivo levels. J Pharmacol Exp Ther. 2005;314(3):1052–8. https://doi.org/10.1124/jpet.105.087601
43. Vatan, Ö. (2019). Rosmarinik asidin cisplatine karşı antisitotoksik ve antigenotoksik etkisinin a549 ve beas-2b hücre hatlarında araştırılması. Uludağ Üniversitesi Tıp Fakültesi Dergisi, 45(3), 263-270. doi:10.32708/uutfd.613912
44. Guzmán C, Bagga M, Kaur A, Westermarck J, Abankwa D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. PLoS One. 2014;9(3):e92444. https://doi.org/10.1371/journal.pone.0092444
45. Dai Y, Sun X, Li B, Ma H, Wu P, Zhang Y, Zhu M, Li HM, Qin M, Wu CZ. The effect of hispidulin, a flavonoid from salvia plebeia, on human nasopharyngeal carcinoma Cne-2z cell proliferation, migration, invasion, and apoptosis. Molecules. 2021;26(6):1604. https://doi.org/10.3390/molecules26061604
46. Chen B, Huang C, Zhang Y, Tang X, Li S, Wang Q, Lin Y. Salvia bowleyana Dunn root is a novel source of salvianolic acid B and displays antitumor effects against gastric cancer cells. Oncol Lett. 2020;20(1):817–27. https://doi.org/10.3892/ol.2020.11611
47. Jiang YL, Xun Y. Molecular mechanism of salvia miltiorrhiza in the treatment of colorectal cancer based on network pharmacology and molecular docking technology. Drug Des Devel Ther. 2024;425–41. https://doi.org/10.2147/DDDT.S443102
48. Zanrè V, Campagnari R, Cerulli A, Masullo M, Cardile A, Piacente S, Menegazzi M. Salviolone from Salvia Miltiorrhiza roots impairs cell cycle progression, colony formation, and metalloproteinase-2 activity in A375 melanoma cells: involvement of P21 (Cip1/Waf1) expression and STAT3 phosphorylation. Int J Mol Sci. 2022;23(3):1121. https://doi.org/10.3390/ijms23031121
49. Golden EB, Lam PY, Kardosh A, Gaffney KJ, Cadenas E, Louie SG, Petasis NA, Chen TC, Schönthal AH. Green tea polyphenols block the anticancer effects of bortezomib and other boronic acid–based proteasome inhibitors. Blood, J Am Soc Hematol. 2009;113(23):5927–37. https://doi.org/10.1182/blood-2008-07-171389
50. Horibata S, Vo T V., Subramanian V, Thompson PR, Coonrod SA. Utilization of the soft agar colony forma-

tion assay to identify inhibitors of tumorigenicity in breast cancer cells. J Vis Exp. 2015;2015(99):e52727. https://doi.org/10.3791/52727
51. Abi-Rizk A, Rayess Y El, Iriti M, Tabet E, Mezher R, Beyrouthy M El. Chemical composition, antitumor and antioxidant effects of four lebanese plants extracts on human pulmonary adenocarcinoma. Nat Prod Res. 2021;35(22):4861–4. https://doi.org/10.1080/14786419.2020.1737056
52. Allegri L, Domenis R, Navarra M, Celano M, Russo D, Capriglione F, Damante G, Baldan F. Dihydrotanshinone exerts antitumor effects and improves the effects of cisplatin in anaplastic thyroid cancer cells. Oncol Rep. 2021;46(3):1–16. https://doi.org/10.3892/or.2021.8155
53. Cao Y, Tang H, Wang G, Li P, Song Z, Li W, Sun X, Zhong X, Yu Q, Zhu S, Zhu L. Targeting survivin with Tanshinone IIA inhibits tumor growth and overcomes chemoresistance in colorectal cancer. Cell Death Discov. 2023;9(1):351. https://doi.org/10.1038/s41420-023-01622-8
54. Kashyap A, Umar SM, JR AD, Prasad CP. Dihydrotanshinone-I modulates epithelial mesenchymal transition (EMT) thereby impairing migration and clonogenicity of triple negative breast cancer cells. Asian Pacific J Cancer Prev APJCP. 2021;22(7):2177. https://doi.org/10.31557/APJCP.2021.22.7.2177
55. Jonkman JEN, Cathcart JA, Xu F, Bartolini ME, Amon JE, Stevens KM, Colarusso P. An introduction to the wound healing assay using live-cell microscopy. Cell Adh Migr. 2014;8(5):440–51. https://doi.org/10.4161/cam.36224
56. Atmaca H, Bozkurt E. Apoptotic and anti-angiogenic effects of Salvia triloba extract in prostate cancer cell lines. Tumor Biol. 2016;37(3):3639–46. https://doi.org/10.1007/s13277-015-4208-2
57. Zare H. Effects of salvia officinalis extract on the breast cancer cell line. SciMedicine J. 2019;1(1):25–9. https://doi.org/10.28991/SciMedJ-2019-0101-4
58. Erdoğan MK, Ağca CA, Aşkın H. İnsan kolorektal kanser hücrelerinde Pistacia eurycarpa ekstraktlarıyla kombine 5-florourasilin artırılmış antiproliferatif ve apoptotik etkileri. Biol Divers Conserv. 2019;12(1):27–38. doi:10.5505/biodicon.2019.57441
59. Muscella A, Stefàno E, De Bellis L, Nutricati E, Negro C, Marsigliante S. Antitumor and antimigration effects of Salvia clandestina L. extract on osteosarcoma cells. Ann N Y Acad Sci. 2021;1500(1):34–47. https://doi.org/10.1111/nyas.14601
60. Sinala S, Dewi STR, Pakadang SR, Sabir M, Kamal SE. Anti-cancer potential of Nggorang leaves extract (Salvia occidentalis Sw.) as a protein P53 supressor in T47D cells. Pharmacogn J. 2021;(13):1036–45. doi:10.5530/pj.2021.13.134
61. Sittihan S, Sopha P, Ruchirawat S. Synthesis and anticancer activity of pentafluorobenzenesulfonamide derivatives as caspase-dependent apoptosis-ınducing agents. ChemMedChem. 2022;17(3):e202100637. https://doi.org/10.1002/cmdc.202100637
62. Yosef R, Pilpel N, Papismadov N, Gal H, Ovadya Y, Vadai E, Miller S, Porat Z, Ben-Dor S, Krizhanovsky V. p21 maintains senescent cell viability under persistent DNA damage response by restraining JNK and caspase signaling. EMBO J [Internet]. 2017 [cited 2022 Nov 4];36(15):2280. https://doi.org/10.15252/embj.201695553
63. Xia R, Sheng X, Xu X, Yu C, Lu H. Hesperidin induces apoptosis and GO/G1 arrest in human non-small cell lung cancer A549 cells. Int J Mol Med. 2018 Jan 1;41(1):464–72. https://doi.org/10.3892/ijmm.2017.3250
64. Crowell JA, Steele VE, Fay JR. Targeting the AKT protein kinase for cancer chemoprevention. Mol Cancer Ther. 2007;6(8):2139–48. https://doi.org/10.1158/1535-7163.MCT-07-0120
65. Islam M, Jones S, Ellis I. Role of akt/protein kinase B in cancer metastasis. Biomedicines. 2023;11(11):3001. https://doi.org/10.3390/biomedicines11113001
66. Erdoğan MK, AĞCA CA, Askin H. Enhanced antiproliferative and apoptotic effects of 5-fluorouracil by combined with Pistacia eurycarpa extracts on human colorectal cancer cells. Biol Divers Conserv. 2019;12(1):27–38. https://doi.org/10.5505/biodicon.2019.57441
67. Jin U, Suh S, Chang HW, Son J, Lee SH, Son K, et al. Tanshinone IIA from Salvia miltiorrhiza BUNGE inhibits human aortic smooth muscle cell migration and MMP-9 activity through AKT signaling pathway. J Cell Biochem. 2008;104(1):15–26. https://doi.org/10.1002/jcb.21599
68. Jiang Y, Ji F, Liu Y, He M, Zhang Z, Yang J, Wang N, Zhong C, Jin Q, Ye X, Chen T. Cisplatin-induced autophagy protects breast cancer cells from apoptosis by regulating yes-associated protein. Oncol Rep. 2017;38(6):3668–76. https://doi.org/10.3892/or.2017.6035
69. Guertin DA, Sabatini DM. Defining the role of mTOR in cancer. Cancer Cell. 2007;12(1):9–22. https://doi.org/10.1016/j.ccr.2007.05.008
70. Yu J ru, Liu Y yue, Gao Y yang, Qian L hui, Qiu J lin, Wang PP, Zhan GJ. Diterpenoid tanshinones inhibit gastric cancer angiogenesis through the PI3K/Akt/mTOR signaling pathway. J Ethnopharmacol. 2024;324:117791. https://doi.org/10.1016/j.jep.2024.117791
71. Woods D, Turchi JJ. Chemotherapy induced DNA damage response: convergence of drugs and pathways. Cancer Biol Ther. 2013;14(5):379–89. https://doi.org/10.4161/cbt.23761
72. Goldstein M, Kastan MB. The DNA damage response: Implications for tumor responses to radiation and chemotherapy. Annu Rev Med. 2015;66(1):129–43. https://doi.org/10.1146/annurev-med-081313-121208
73. Mauri G, Arena S, Siena S, Bardelli A, Sartore-Bianchi A. The DNA damage response pathway as a land of therapeutic opportunities for colorectal cancer. Ann Oncol [Internet]. 2020;31(9):1135–47. doi:10.1016/j.annonc.2020.05.027
74. Telang N. Anti-proliferative and pro-apoptotic effects of rosemary and constituent terpenoids in a model for the HER-2-enriched molecular subtype of clinical breast cancer. Oncol Lett. 2018;16(4):5489–97. https://doi.org/10.3892/ol.2018.9238

ORIGINAL PAPER

JMS Journal of Medical Science

Deep neural networks to classify motor unit action potential signals acquired using needle electromyography

Isil Tatlidil
Karadeniz Technical University Faculty of Medicine, Department of Clinical Neurophysiology, Trabzon, Turkiye
https://orcid.org/0000-0002-9666-3647
Corresponding author: tatlidilisil173@hotmail.com

Murat Ekinci
Karadeniz Technical University Engineering Faculty, Department of Computer Engineering, Trabzon, Turkiye
https://orcid.org/0000-0001-9326-8425

Cavit Boz
Karadeniz Technical University Faculty of Medicine Department of Clinical Neurophysiology, Trabzon, Turkiye
https://orcid.org/0000-0003-0956-3304

https://doi.org/10.20883/medical.e1391

Keywords: deep learning, convolutional neural network, gate recurrent unit, needle electromyography, neuromuscular disorders

Received 2025-08-15
Accepted 2025-10-10
Published 2025-12-29

How to Cite: Tatlidil I, Ekinci M, Boz C. Deep neural networks to classify motor unit action potential signals acquired using needle electromyography. Journal of Medical Science. 2025 December;94(4);e1391. doi:10.20883/medical.e1391

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Aim. To classify motor unit action potential patterns using a deep learning technique with high accuracy.
Material and methods. A dataset was compiled from three main groups of motor unit action potential patterns, including myopathy, neuropathy, and normal, as assessed by a clinical neurophysiologist during routine clinical assessments. After preprocessing the raw signals in the dataset, a total of 3,152 signal segments from 96 muscles of 26 individuals were divided into training and test sets. Deep learning network models were developed in Python using the Keras API in Jupyter Notebook.
Results. Among the deep learning models, a hybrid deep neural network model with a one-dimensional convolution layer as an input layer and four layers of gate recurrent units (1DCNN-GRU) achieved the highest accuracy rates. Ten-fold cross-validation resulted in a mean accuracy rate of 98.13 ± 1.05%.
Conclusions. Both conventional machine learning models and deep learning models could classify needle EMG patterns that belonged to three neuromuscular disorder groups with high accuracy. However, more clinical studies with larger datasets are needed for validation. In contrast to conventional machine learning techniques, deep learning models could receive signals as input data and automatically extract the required features. Therefore, they could facilitate the real-time implementation of the pattern recognition tasks in the future.

Introduction

Assessing audio-visually, needle electromyography (EMG) is a crucial diagnostic tool used to differentiate between the three main types of neuromuscular conditions: healthy, neuropathic, or myopathic [1]. During needle EMG procedures, potentials originating from needle insertion into a muscle, muscle cell membrane potentials while resting, and motor unit action potentials during

muscle contraction are recorded [2,3]. Although there are many experienced electromyographers or neurophysiologists, needle EMG examinations are still highly subjective, and, in some cases, reaching a consensus for certain kinds of patterns is unlikely [1,4]. Machine learning methods, intensive learning, have the potential to enhance standardisation in pattern recognition and may offer valuable diagnostic support to less experienced electromyographers in the future [5].

Machine learning methods have been extensively applied in recent clinical research to address challenges related to classification, decision-making, and prognosis prediction within clinical practice. [5,6]. As a subtype of artificial intelligence, machine learning has yielded promising results in biomedical image and signal processing studies, paving the way for the development of clinical decision support systems [7]. Various machine learning techniques are available for use in medical studies, which can be grouped into classic or conventional machine learning and deep learning techniques based on the data processing steps [5,8]. Traditional machine learning techniques share standard pipelines for processing image or time series data. After preprocessing data, quantitative analysis is applied to extract features. The best meaningful features for classification are then required to be chosen as the input data for a machine learning tool, such as a decision tree, support vector machine, logistic regression, or a small-scale artificial neural network, such as a multi-layer perceptron [6,9]. Although conventional machine learning techniques have hand-crafted feature extraction and selection steps in their data processing pipeline, deep learning techniques such as convolutional neural networks (CNN), recurrent neural networks (RNN), autoencoders, long-short term memories (LSTM), and gate recurrent units (GRU) can automatically achieve these processing steps. Therefore, the image or signal presented in a time-series format can be provided as input data to the classification pipeline of an artificial neural network [1, 5, 8]. Direct use of signal data as input may streamline the automation process and enhance decision-making capabilities, enabling real-time implementation.

Despite several conventional machine learning studies showing promising results for classifying needle EMG signals, studies that implement deep learning for needle EMG signal classification are scarce [1,8]. Few studies have focused on classifying resting state membrane potentials [1]. Nodera et al. [10] classified resting-state potentials with 86% accuracy using Mel spectrograms as input data, and pre-trained models of convolutional neural networks for image recognition were employed. The accuracy reached up to 94% in the 19-layered Visual Geometry Group (VGG-19), a convolutional neural network model with 19 weighted layers, by using data augmentation techniques. Nam et al. [11] classified resting state potentials using Inception-v4 as a convolutional neural network, yielding results with 93% accuracy. The input data consisted of image files with .png extensions representing the resting state potentials. There are also a few studies with encouraging results that aim to classify motor unit action potentials during muscle contraction using deep learning techniques [1]. Sengur et al. [12] used a continuous wavelet transform (CWT) spectrogram and Pseudowigner-Wille distribution function for preprocessing. Two-dimensional spectrograms were used as input data for the two-dimensional convolutional neural network (2D-CNN) pipeline, which achieved 96.8% accuracy in classifying two groups: Amyotrophic Lateral Sclerosis (ALS) and normal. Yoo et al. [8] utilised nEMGnet, a one-dimensional residual convolutional neural network inspired by the Residual Neural Network (ResNet) and the Visual Geometry Group (VGG) neural network, for the classification of EMG segments, achieving 62.35 ± 4.60% accuracy in classifying normal, neurogenic, and myogenic segments. Then, the divide-and-vote algorithm was used to determine the disease labelling for each muscle and patient with 83% accuracy.

An RNN is a specialised form of deep neural network particularly well-suited for handling sequential data, such as time series [13]. They are mainly used to predict the following likely data. However, classifying signals or time series could also be possible by arranging the architecture [14]. LSTM and GRU are subtypes of RNNs that yield successful results in surface EMG studies [15, 16]. Our study aims to classify nEMG data with motor unit action potentials into three categories, standard, neuropathy, and myopathy, using deep learning techniques with high accuracy.

Material and methods

Data acquisition

We included EMG traces acquired using concentric needle EMG recordings on a Natus Neurology Nicolet EMG system with an AT2+6 amplifier in this study. We retrospectively reviewed the recorded needle EMG signals obtained in routine clinical practice from December 2022 to June 2023, together with the EMG reports of clinical neurophysiologists. We received a total of 155 EMG signal traces corresponding to 134 muscles in 33 patients. We applied labelling at the patient and muscle levels according to the EMG reports in the archive. We excluded cranio-bulbar muscles due to the shorter duration and smaller amplitude of motor unit action potentials compared with limb muscles [3,17]. Among 33 subjects, 15 were normal, 11 had pathologies causing neurogenic motor unit action potentials, and seven had myopathies with myogenic action potentials. Among the 15 normal subjects, we excluded one subject due to an increased percentage of polyphasic motor unit action potential findings and another two subjects because the study was performed only on cranio-bulbar muscles. In the neurogenic group, we excluded one patient because the study focused only on cranio-bulbar muscles, and two patients due to being in the acute to subacute phase of neuropathy, as no changes in motor unit action potential morphology related to reinnervation were observed. We also excluded muscles that did not represent labelled patterns.

Finally, we included EMG traces of 29 muscles from seven individuals with myopathy, 24 muscles from eight individuals with neuropathy, and 43 muscles from 11 normal individuals in the study. Detailed information about the subjects is presented in **Table 1A–1E**. The recording parameters included a 30-50 Hz high-pass filter, a 10 kHz low-pass filter, and a sampling rate of 48,000 per second. We extracted EMG data from the Synerqy software system in .txt format. We reformatted the .txt files for use with a program developed in WinForms Visual Studio C++ to review traces in 100-msec sweep times. We also used this pro-

Table 1A. Demographics of the cases according to records.

Demographics	Female	Male	Total	Age (Mean/Median/Range)
Myogenic	2	5	7	52.8/56/21-80
Neurogenic	2	6	8	52.6/52.5/38-72
Normal	5	6	11	45.4/54/12-76
Total	9	17	26	49.6/53.5/12-80

Table 1B. Sampled muscle records from the upper limb and shoulder

Muscles sampled	PM	Is	Ss	TB	D	BB	EDC	FDS	FCU	FDI	ADM	APB
Myogenic	4	-	2	-	4	5	4	-	-	1	-	-
Neurogenic	-	-	1	-	2	-	-	-	1	1	1	2
Normal	1	1	1	1	6	5	2	1	1	5	-	1
Total	5	1	4	1	12	10	6	1	2	7	1	3

PM = Pectoralis major, Ss = Supraspinatus, Is = Infraspinatus, Triceps Brachi = TB, D = Deltoid, BB = biceps brachi, EDC = Extensor digitorum Communis, FDS = Flexor Digitorum Superficialis, FCU = Flexor Carpi Ulnaris, FDI = First Dorsal interosseus, ADM = Abductor Digiti Minimi, APB = Abductor Pollicis Brevis

Table 1C. Sampled muscle records from the lower limb and hip.

Muscles sampled	Ip	GlMax	GlMed	TFL	RF	VM	AM	TA	GcM	TP	EHL
Myogenic	7	-	-	-	1	1	-	-	-	-	-
Neurogenic	-	2	2	1	-	1	1	5	2	1	1
Normal	1	-	-	-	1	4	-	5	6	1	-
Total	8	2	2	1	2	6	1	10	8	2	1

Ip = Iliopsoas, RF = Rectus Femoris, VM = Vastus Medialis, AM = Adductor Magnus, TA = Tibialis Anterior, GcM = Gastrocnemius Medialis, TP = Tibialis Posterior, EHL = Extensor Hallucis Longus, TFL = Tensor Fascia Lata, GlMax = Gluteus Maximus, GlMed = Gluteus Medius

Table 1D. Number of neurogenic conditions diagnosed.

Neurogenic Pathologies	n
Chronic L5-S1 radiculopathy	1
Subacute-Chronic L4-L5 radiculopathy	1
Chronic L4-L5 radiculopathy	1
Chronic L3-L4, L4-L5 radiculopathy	1
Chronic C5-C6 radiculopathy	2
Chronic median nerve injury	1
Chronic lower truncus injury	1

Table 1E. Number of myogenic conditions diagnosed.

Myogenic Pathologies	n
Myopathy alone	5
Myopathy with fibrillations	1
Myopathy with myotonia	1

gram for preprocessing and labelling. The study received approval from the ethics committee of Karadeniz Technical University.

Preprocessing

We down-sampled raw needle EMG signals to 9600 Hz to reduce computational complexity. For the down-sampling procedure, we used an algorithm that calculates the second derivative of each signal. The signal with the lowest second derivative value was chosen as the representative signal for each of the five consecutive signals in the trace. As a result, the local minimums and maximums of high-frequency patterns – containing details regarding small motor unit action potentials, turns, and phases, which are essential characteristics of myopathy – were retained during the down-sampling process (see **Figure 1**). We also filtered the signal traces using a 50-Hz high-pass filter to remove redundant low-frequency information digitally. Each nEMG data sample recorded from the labelled muscle has traces that are 4 to 20 seconds long. Two neurophysiologists (IT and CB) reviewed all nEMG traces and selected segments for labelling that were artefact-free and at least 0.5 seconds long. We sliced the extracted traces into 100-msec segments with 50-msec overlap. Following processing, the neurophysiologists conducted a thorough review of each signal segment. They removed any traces from the dataset that did not contain characteristic features of the designated neuromuscular disorder group. We also excluded the traces with far-field potentials. Most of the signal segments had 945-960 consecutive signal series. We determined the standard signal vector length to be 940 to ensure a consistent size for each signal segment. Eight signal segments, three of them belonging to the normal group, and five belonging to the neurogenic group, required zero padding at the end as a result of being shorter than 940. Finally, we selected 1221 normal, 718 neurogenic, and 1212 myogenic EMG signal segments as the training and test sets. Thus, we generated a dataset consisting of 3,152 signal rows with 940 columns.

Deep Learning Network Model and Training

We performed deep learning using Python 3.7.16 in Jupyter Notebook. We designed deep neural networks using LSTM, GRU, and CNN in various architectures with the Keras library and a TensorFlow 2.6.0 backend in Python. We utilised the classification model to categorise EMG signal segments into three groups: neurogenic, myogenic, and normal. Therefore, the output layer consisted of three neurons, representing three neuromuscular disease groups according to the one-against-all principle. Preprocessed EMG signal segments, recorded as a CSV table, were provided as input data. Initially, we used 80% of labelled signal segments as the training set and the remainder as the test set. Then, avoiding overfitting, we also applied ten-fold cross-valida-

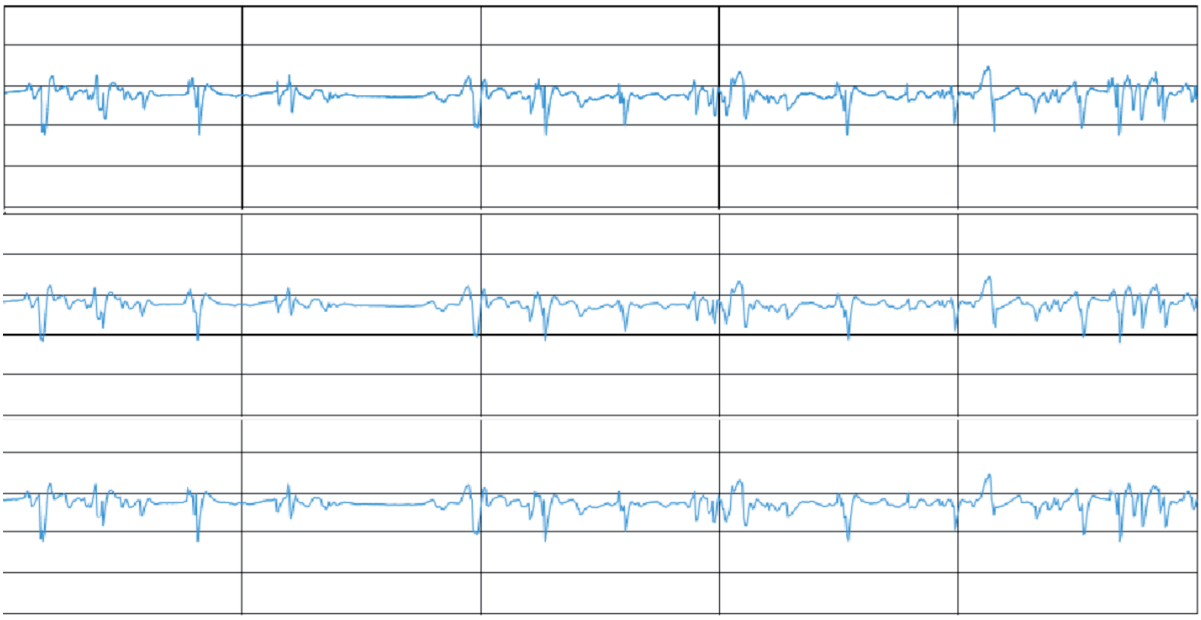


Figure 1. We down-sampled the signal segments to 9600 Hz to reduce computational complexity. However, during the down-sampling procedure, high-frequency information could be missed, which is particularly relevant for the recognition of myopathy forms. Each trace segment contains 100-msec-long signals. The raw trace, the first trace, contains high-frequency details that may be overlooked by traditional down-sampling methods, such as average pooling, which is seen in the second trace. By computing the second derivative at each signal point, peaks can be identified and selected as representative points during downsampling, as shown in the third trace. This approach enables greater preservation of high-frequency information compared with conventional techniques.

tion with randomly chosen training and test sets to assess the performance of the models. We performed the training and testing procedure on a CUDA-enabled NVIDIA GPU (GeForce RTX 3060) running on the Windows 11 operating system.

Statistics

We performed the statistics using the R Studio software package, version 4.4.1. For continuous variables, we conducted the Shapiro-Wilk test to check the normality of data distribution. We used Student's t-test for normally distributed continuous variables and the Mann-Whitney U test for non-normally distributed continuous variables to evaluate the differences between the groups. Statistical significance was considered significant if $p < 0.05$.

Results

We tested various deep learning architectures and optimised the best-performing models using the grid search algorithm. Finally, we obtained the mean accuracy rates of the ten-fold cross-validation results and assessed the models through statistical analysis.

We developed two main models. One was a residual convolutional network, similar to nEMGnet from Yoo et al., and the other was a multilayer RNN model that utilised either a unidirectional or bidirectional GRU or LSTM. We achieved the highest mean accuracy rate (98.13 ± 1.05%) using ten-fold cross-validation when testing the model 1D-CNN+4-layer-unidirectional GRU without max pooling. Statistical analysis revealed that the accuracy rate of the model 1D-CNN+4-layer-unidirectional GRU without maxpooling was significantly higher than 5-layer CNN ($p = 0.024$), 11-layer CNN without residual layers ($p = 0.005$), 4-layer unidirectional LSTM ($p < 0.001$), 4-layer-bidirectional-LSTM ($p < 0.001$), 4-layer-unidirectional GRU ($p = 0.012$), 3-layer-1D-CNN+2-layer-unidirectional GRU ($p = 0.028$), 1-layer CNN-2-layer-unidirectional GRU ($p = 0.003$), 2-layer CNN-1-layer-bidirectional GRU ($p = 0.01$), and 2-layer CNN-1-layer-unidirectional LSTM ($p < 0.001$). We summarised all results obtained with ten-fold cross-validation on the dataset, which was split randomly into a train set and a test set, and the

p-values for the models compared to 1-layer-1-D-CNN+4-layer-unidirectional GRU are presented in Table 2. For the best-performing model, 1D-CNN+4-layer-unidirectional GRU without max pooling, we present the normalised confusion matrix of the ten-fold cross-validation results, along with the mean and standard deviation, in **Figure 2**. The mean and standard deviation of precision, recall, and F1 scores are also presented in **Table 3**. The pipeline and the detailed archi-

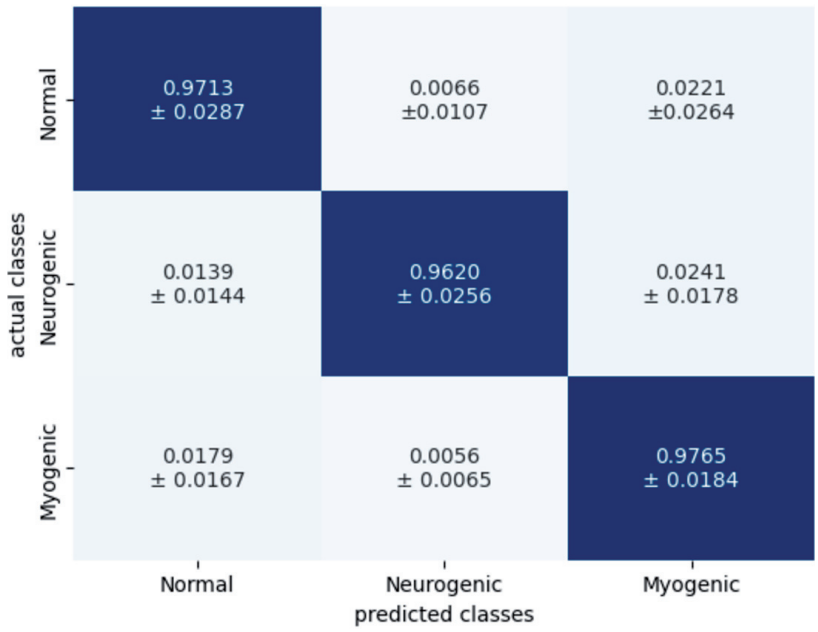


Figure 2. Confusion matrix of the optimised 1D-CNN-4-layer GRU model obtained through ten-fold cross-validation with normalised means and standard deviation.

Table 2. The optimised deep learning models and accuracy rates with ten-fold cross-validation.

Deep learning model	Epochs	10-fold cross-validation results (accuracy)	P*
5-layer CNN	50	96.55 ± 1.60 %	0.024
11-layer CNN without residual layers	50	95.71 ± 2.23 %	0.005
11-layer CNN with 3 residual layers	50	97.1 ± 1.05 %	0.052
4-layer-unidirectional-LSTM	70	95.23 ± 1.23%	<0.001
4-layer-bidirectional-LSTM	90	94.64 ± 1.87%	<0.001
4-layer -unidirectional -GRU	30	96.13 ± 1.82%	0.012
4-layer- bidirectional-GRU	35	97.04 ± 2.18 %	0.116
1-layer-1D-CNN+4-layer-unidirectional GRU	90	97.77 ± 1.08%	0.487
1 layer-1D-CNN+4-layer-unidirectional GRU (without maxpooling)	90	98.13 ± 1.05%	NA
1-layer-1D-CNN+4-layer unidirectional LSTM (without maxpooling)	250	95.52 ± 3.46%	0.052
1-layer-1D-CNN+4-layer-bidirectional GRU (without maxpooling)	90	97.81 ± 0.96%	0.513
1-layer-1D-CNN+3-layer-unidirectional GRU	80	96.47 ± 2.54%	0.093
2-layer-1D-CNN+2-layer-unidirectional GRU	70	97.82 ± 0.59%	0.445
2-layer-1D-CNN+3-layer-unidirectional GRU	90	97.77 ± 1.45%	0.878
3-layer-1D-CNN+3-layer-unidirectional GRU	70	96.98 ± 1.87%	0.145
3-layer-1D-CNN+2-layer-unidirectional GRU	90	97.02 ± 0.92%	0.028
1-layer CNN-2-layer-unidirectional GRU	70	96.50 ± 1.16%	**0.003**
2-layer CNN-1-layer-unidirectional GRU	70	97.92 ± 0.82%	0.661
2-layer CNN-1-layer-bidirectional GRU	70	96.27 ± 1.60%	0.010
2-layer CNN-1-layer-unidirectional LSTM	90	96.03 ± 1.08%	<0.001

* P-value for the models in comparison with 1-layer, -1D-CNN+4-layer-unidirectional GRU

Table 3. Precision, recall, and F1 scores for each group in the 1d-CNN-4-layer GRU model.

	Precision	Recall	F1-scores
Normal	0.9778 ± 0.0174	0.9845 ± 0.0139	0.9800 ± 0.0118
Neurogenic	0.9926 ± 0.0118	0.9761 ± 0.0162	0.9842 ± 0.0081
Myogenic	0.9787 ± 0.0148	0.9817 ± 0.0235	0.9800 ± 0.015
Accuracy	0.9813 ± 0.0105	0.9813 ± 0.0105	0.9813 ± 0.0105

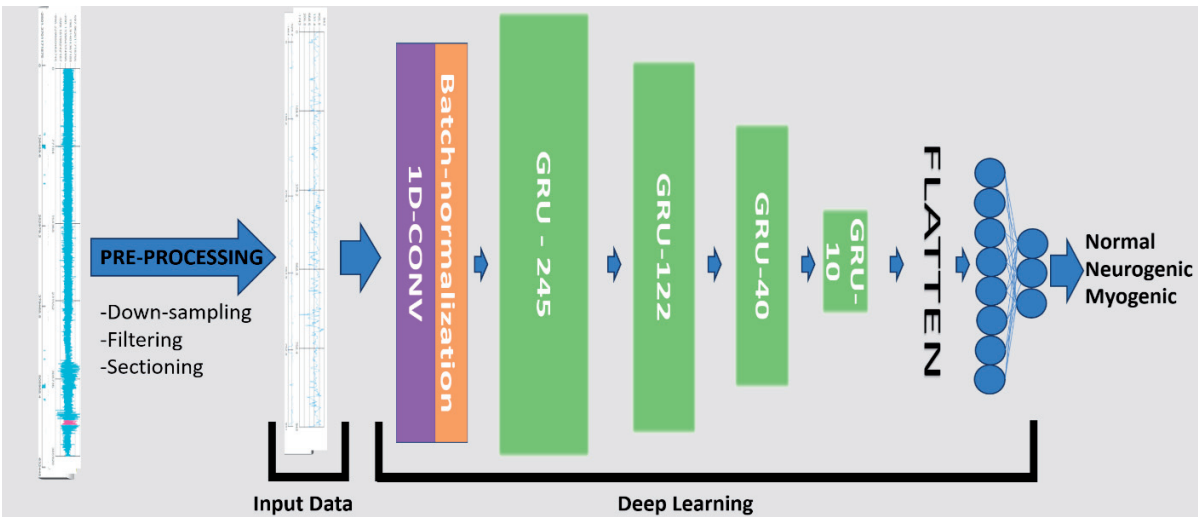


Figure 3. The deep learning pipeline achieved a mean accuracy rate of 98.13 ± 1.05% with ten-fold cross-validation. A similar architecture was also designed using LSTM and bilateral GRU instead of unilateral GRU.

tecture of the model are shown in **Figures 3** and **4**, respectively. The residual convolutional neural network pipeline and architecture are presented in **Figure 5**.

Discussion

Receiving needle EMG signal segments as input data, our study is the first to classify needle EMG data using a deep learning technique, achieving high accuracy. Both the residual convolutional neural networks and the 1D-CNN-4-layer GRU model without max pooling yielded high accuracy results. In the literature, although CNN is a successful deep learning tool receiving raw images as input, LSTM and GRU, as subtypes of RNN, show more promising results in time series data, such as electroencephalography [18]. Accordingly, our study mainly focused on RNN-based models to achieve high accuracy rates. Notably, GRU-based models showed slightly, and in some cases, statistically significantly better performance than LSTM-based models. Although the accuracy rates of LSTM-based models surpass those of GRU-based models in most deep learning studies in the literature, GRU-based models have also been reported to show superior success rates in some deep learning studies. The small size of the data set and the low dimension of the input data could account for the better accuracy results of GRU-based models [19,20]. The improved accuracy rates observed in unidirectional models may also be attributed to similar underlying factors.

Despite the success of RNN in time series data, residual CNN models also yielded promising results in our study. The residual CNN model architecture we designed was inspired by the architecture of nEMGnet, as described in the study by Yoo et al. [8]. Because the signal segment length for the input layer was 4-fold shorter than nEMGnet, the same number of layers would cause a negative output dimension. Therefore, the number of layers was fewer than that of nEMGnet. However, in our study, results for the accuracy of segment-based classification are significantly higher compared with the 62.35 ± 4.60% reported by Yoo et al. [8]. This may be due to both the character-

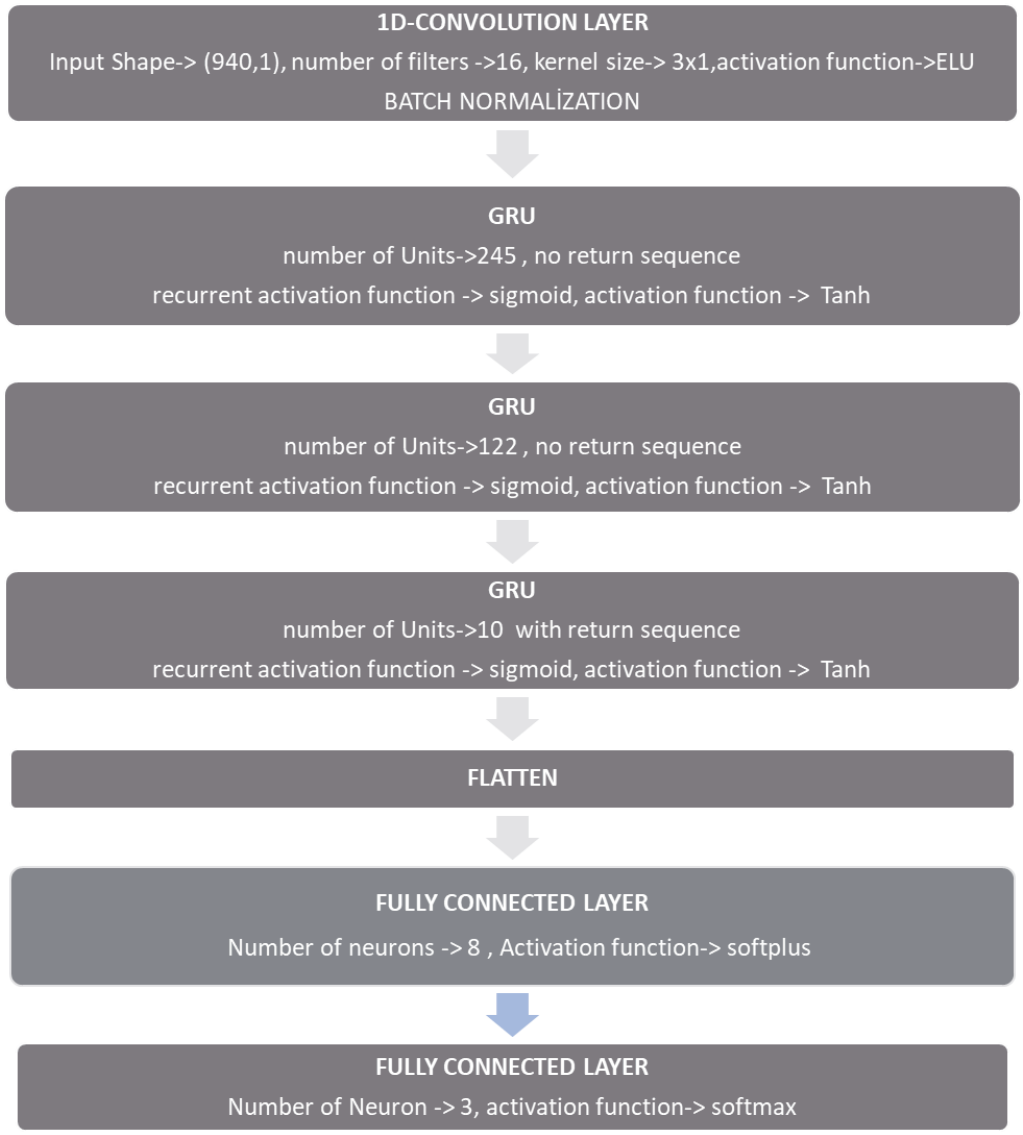


Figure 4. Detailed architecture of the 1D-CNN-4-layer GRU model deep learning model.

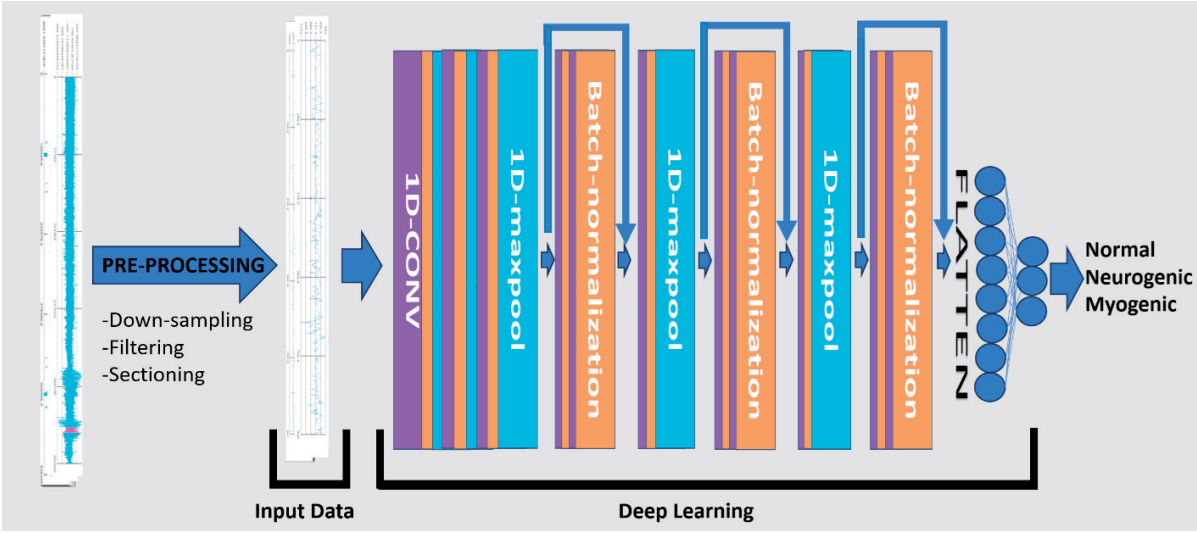


Figure 5. The residual convolutional neural network pipeline achieved an accuracy of 97.1 ± 1.05%.

istics of the dataset and nuances in the architectural design. Overall, adding residual layers could improve the performance of convolutional neural networks, as indicated by the statistical analysis of the study. Although the suggested model, model 1D-CNN+4-layer-unidirectional GRU without maxpooling, had significantly better performance than other CNN models, there was no statistically significant difference between the proposed model and the 11-layer CNN with residual layers.

On the other hand, Sengur et al. [12] achieved 96.8% accuracy in classifying normal and ALS groups using deep learning, which was higher than in other deep learning studies. However, the input data was not the EMG trace itself. EMG signal segments were converted into two-dimensional spectrograms through the CWT and Pseudowigner-Wille distribution function implementation.

Despite our study having some promising results, conventional machine learning techniques have yielded better accuracy. Extracting features from time-frequency analysis, Samantha et al. [21] achieved 99.5% accuracy in classifying ALS and normal groups using a genetic algorithm. Roy et al. [22] achieved 100% accuracy in discriminating between healthy and neuropathy conditions using a support vector machine and feature extraction methods based on time-frequency analysis. Dubey et al. [23] employed the Hilbert transform for feature extraction, achieving 99.5% accuracy with feedforward neural networks in classifying neuropathy, myopathy, and normal groups. Kamali et al. [24] achieved 100% accuracy using a random forest algorithm on samples from the tibialis anterior muscle for the classification of neuropathy, myopathy, and normal groups. They also employed time-frequency analysis for feature extraction. As seen in previous studies, feature extraction from time-frequency analysis yields the best results. Therefore, it may be pertinent to deepen the understanding of deep learning methodologies, as well as to design and refine their automatic feature extraction capabilities within the scope of time-frequency analysis. Further optimisation of deep learning models is necessary to achieve higher accuracy rates.

In our study, we created the dataset by retrospectively using recorded data. Similar to publicly available datasets, it consisted of needle EMG signals recorded from the muscles of 26 subjects: seven individuals with myopathy, eight individuals with neuropathy, and 11 healthy individuals. The number of subjects for each group was similar to the EMGLab dataset, a publicly available dataset used in most studies [1]. In contrast to the EMGLab dataset, our dataset did not contain patients with motor neuron disease. However, patients with radiculopathy and nerve palsies have typical clinical features and needle EMG patterns easily classified by an expert.

In contrast to the EMGLab dataset, since the study was retrospective, all patients had records from various muscles in our dataset. Additionally, depending on the patient's cooperation during the recorded part of the study, different recruitment patterns may be observed in the signal segments. On one hand, the involvement of various muscles and a wide variety of recruitment patterns could make the classification task more challenging. On the other hand, standardised recordings from the same muscle could help achieve more difficult tasks, such as disease subgroup classifications.

Despite deep neural networks being considered black boxes in terms of their feature extraction capabilities, we arranged the input parameters based on some electrophysiologic parameters to optimise the results. Considering that the sweep time of an EMG device screen is commonly adjusted to 5 msec/div or 10 msec/div, and a 100-msec trace provides an idea about motor unit action potential firing rates and morphology, we determined the length of EMG segments to be 100 msec. Most previous studies also used 100-msec segments [8]. The residual convolutional network architecture was similar to that of nEMGnet in the study by Yoo et al. [8]. However, we preferred smaller sizes of convolutional filter kernels to process high-frequency information. The design of the 1-D CNN-4-layer GRU model also contains architectural nuances. We estimated the number of units in the first RNN layer by considering specific parameters of the motor unit action potentials. The approximate motor unit action potential duration varies between 5 and 15 msec; it is shorter in myopathy (1-5 msec) but longer in neuropathy (15-25 msec) [3]. As 245 units of RNN capture approximately 25 milliseconds, it could be adequate for processing all types of motor unit action potential morphologies.

Additionally, because high-frequency information plays a crucial role in the discrimination

of myopathy, neuropathy, and standard motor unit action potentials, and GRU layers are specialised in learning time series patterns, the model was further optimised by omitting the maxpooling layer, which might cause losses in the high-frequency information captured by the GRU units. In most studies, a flattened layer is used before GRU units. However, we did not use this due to its low accuracy rates in tests. Because the convolution layer produces output data as several time series equal to the number of convolution filters, flattening it may cause some information to be lost in the time series processed by the GRU layer, which could lead to low accuracy rates.

The retrospective design of the study did not permit a standardised assessment of motor unit potential recruitment patterns. This limitation also led to the exclusion of acute neuropathies from the study, conditions that require analysis of recruitment patterns for an accurate understanding [3]. When a muscle is slightly contracted, the smallest motor unit action potential first starts firing, and its frequency is approximately 5-8 Hz at the beginning. By increasing the force, the frequency increases by 3-5 Hz, then a second larger motor unit is recruited at <15 Hz [25]. By 50% of maximal voluntary contraction, almost all motor units are recruited in small limb muscles, and further voluntary contraction rates could increase the firing frequency of motor unit action potentials up to 50 Hz. In large limb muscles, motor units are recruited at least 90% of maximal voluntary contraction, while muscle firing frequency reaches up to 30-40 Hz [26]. Thus, a motor unit action potential can be fired one to five times in small muscles and one to four times in large muscles within a 100-msec period. Although this rate increases in myogenic muscle, the number of motor unit action potentials recruited reduces in neurogenic muscle [3]. Finally, a wide variety of voluntary motor unit contraction rates have been successfully classified using the proposed deep learning model with 245 hidden units, which can capture periods of approximately 25 milliseconds. Considering that patients may have poor cooperation in recruitment tasks in clinical practice, the high accuracy rate of the proposed model in classifying neuromuscular disorders offers hope for the real-time implementation of the model. Nevertheless, standardisation of recruitment patterns, particularly for training neural network models, could both increase accuracy and help classify neuropathic and myopathic pattern subtypes.

We included a wide variety of small and large muscles in the limbs, hips, and shoulders in our study. The age of the subjects also ranged from 12 to 80 years. Different limb muscles have slightly different ranges of amplitude, duration, or polyphasic rates [3]. Furthermore, with ageing, motor unit action potential morphology can also differ slightly in size, and action potentials with longer durations and higher amplitudes are observed in healthy older subjects [3], which may lead to lower accuracy rates in pattern recognition. Despite acquiring data recorded from various muscles from a highly heterogeneous patient population, we achieved high accuracy rates using the proposed model. Although this situation may be helpful for the real-life implementation of the model, standardising the muscles could aid in muscle-specific assessments and increase accuracy. Studying only with the tibialis anterior muscle, Roy et al. [22] and Kamali et al. [24] achieved 100% accuracy.

After labelling the muscle according to the EMG report results, both artefacts and EMG signal segments that did not represent typical patterns of labelled neuromuscular conditions were excluded. This approach may increase accuracy rates and could be considered a barrier to implementing the model for real-time use. However, the fact that most patients in the data set had acquired myopathy and radiculopathy, with patchy muscle involvement, particularly in some acquired myopathies and radiculopathies, could explain the standard motor unit action potential patterns in pathologic signal segments [3]. Labelling these segments as neuropathic or myopathic would reduce the quality of training and result in low accuracy rates. However, there may still be some subtle patterns of the pathology that could not be discerned by an expert but could be detected with deep learning systems [27].

This study has certain limitations. The small sample size is a primary constraint, which may increase the risk of overfitting. Additionally, the limited number of patients prevented the use of patient-based classification and the application of an external validation data set, as referenced by de Jonge et al. [1]. Furthermore, we were unable to use an independent external validation set. The results cannot be generalised to the entire

neuromuscular disorder classification because the patient population in our study was small and restricted to patients with radiculopathy, axonal nerve injury, and predominantly inflammatory myopathies. As a result, both the limited diversity of the neuromuscular disorders and the small size of the datasets reduce the generalizability of the results and increase the risk of overfitting. Moreover, the real-time implementation of automatic classification of neuromuscular disorders requires the discrimination of motor unit action potentials, artefacts, needle insertion potentials, and resting state potentials. In a retrospectively designed clinical study, Hubers et al. [17] classified signal segments with artefacts, resting-state potentials, and motor unit action potentials with 96% accuracy at the first step, and then the classification of motor unit action potentials was performed. Our study focused solely on classifying motor unit action potentials. We were unable to assess different levels of recruitment due to the small dataset size and the retrospective design of the study. For this reason, having only reduced recruitment as a pathologic finding, acute neuropathies were excluded. Therefore, assessing recruitment patterns in future studies would also improve real-life generalizability.

The current findings of this study are preliminary. Future work will focus on increasing the number of patients for patient-based classification and expanding the dataset to include a broader spectrum of pathologies, such as common genetic and inflammatory myopathies, motor neuron disease, and both acquired and genetic polyneuropathies. Additionally, plans include conducting multicenter studies to obtain external validation datasets and using labelling based on expert consensus, aiming to enhance the reliability of the results. Ultimately, after developing deep learning models for recognising motor unit action potential morphology, artefacts, resting-state potentials, and recruitment patterns, the objective is to create real-time systems with high accuracy.

Conclusion

Deep learning is actively used in most current studies for pattern recognition tasks. The automatic feature extraction ability of deep learning tools reduces the complexity of the machine learning procedure. Showing good performance on time-series data, both 1D-CNN-4-layer GRU and residual convolutional neural network models show promising results on needle-EMG data in terms of neuromuscular disease classification. However, more clinical studies with larger datasets are needed to validate the results for clinical application.

Disclosures

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Ethical Approval

This retrospective study was reviewed and approved by the Ethical Committee of Karadeniz Technical University on October 3, 2024 (Registry number: E-82554930-050.01.04-567570).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Sources

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

Declaration of Generative AI

The authors declare that they did not use generative AI in the production of this manuscript.

CRediT authorship contribution statement

Isil Tatlidil: Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Software, Visualisation, Writing – original draft. **Murat Ekinci**: Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. **Cavit Boz**: Data curation, Resources, Supervision, Validation, Writing – review & editing.

References

1. De Jonge S, Potters W V, Verhamme C. Artificial intelligence for automatic classification of needle EMG signals: A scoping review. Clin Neurophys-

iol. 2024 Mar;159:41–55. https://doi.org/10.1016/j.clinph.2023.12.134.
2. Rubin DI. Needle electromyography: Basic concepts. Handb Clin Neurol. 2019 Jan 1;160:243–56.
3. Preston DC, Shapiro BE. Electromyography and Neuromuscular Disorders. 4th ed. Philadelphia: Elsevier; 2021. 134–259 p.
4. Kendall R, Werner RA. Interrater reliability of the needle examination in lumbosacral radiculopathy. Muscle Nerve. 2006 Aug;34(2):238–41. https://doi.org/10.1002/mus.20554.
5. Shen C, Nguyen D, Zhou Z, Jiang SB, Dong B, Jia X. An introduction to deep learning in medical physics: advantages, potential, and challenges. Phys Med Biol. 2020 Mar 3;65(5):05TR01. https://doi.org/10.1088/1361-6560/ab6f51.
6. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol. 2020 Feb 27;9(2):14. https://doi.org/10.1167/tvst.9.2.14.
7. Caiafa CF, Sun Z, Tanaka T, Marti-Puig P, Solé-Casals J. Special Issue 'Machine Learning Methods for Biomedical Data Analysis'. Sensors (Basel). 2023 Nov 24;23(23). https://doi.org/10.3390/s23239377.
8. Yoo J, Yoo I, Youn I, Kim SM, Yu R, et al. Residual one-dimensional convolutional neural network for neuromuscular disorder classification from needle electromyography signals with explainability. Comput Methods Programs Biomed. 2022 Nov;226:107079. https://doi.org/10.1016/j.cmpb.2022.107079.
9. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. Comput Methods Programs Biomed. 2018 Jul;161:1–13. https://doi.org/10.1016/j.cmpb.2018.04.005.
10. Nodera H, Osaki Y, Yamazaki H, Mori A, Izumi Y, Kaji R. Deep learning for waveform identification of resting needle electromyography signals. Clin Neurophysiol. 2019 May;130(5):617–23. https://doi.org/10.1016/j.clinph.2019.01.024.
11. Nam S, Sohn MK, Kim HA, Kong HJ, Jung IY. Development of artificial intelligence to support needle electromyography diagnostic analysis. Healthc Inform Res. 2019;25(2):131. https://doi.org/10.4258/hir.2019.25.2.131.
12. Sengur A, Akbulut Y, Guo Y, Bajaj V. Classification of amyotrophic lateral sclerosis disease based on convolutional neural network and reinforcement sample learning algorithm. Health Inf Sci Syst. 2017 Dec 30;5(1):9. https://doi.org/10.1007/s13755-017-0029-6.
13. Zhang Z, He C, Yang K. A novel surface electromyographic signal-based hand gesture prediction using a recurrent neural network. Sensors. 2020 Jul 17;20(14):3994. https://doi.org/10.3390/s20143994.
14. Das S, Tariq A, Santos T, Kantareddy SS, Banerjee I. Recurrent neural networks (RNNs): Architectures, training tricks, and introduction to influential research. Newyork: Humana; 2023. https://doi.org/ 10.1007/978-1-0716-3195-9_4.
15. Keleş AD, Turksoy RT, Yucesoy CA. The use of non-normalized surface EMG and feature inputs for LSTM-based powered ankle prosthesis control algorithm development. Front Neurosci. 2023;17:1158280. https://doi.org/10.3389/fnins.2023.1158280.
16. Aviles M, Alvarez-Alvarado JM, Robles-Ocampo JB, Sevilla-Camacho PY, Rodríguez-Reséndiz J. Optimizing RNNs for EMG Signal Classification: A novel strategy using grey wolf optimization. bioengineering (Basel). 2024 Jan 13;11(1). https://doi.org/10.3390/bioengineering11010077.
17. Hubers D, Potters W, Paalvast O, de Jonge S, Doelkahar B, Tannemaat M et al. Artificial intelligence-based classification of motor unit action potentials in real-world needle EMG recordings. Clinical Neurophysiology. 2023 Dec;156:220–7. https://doi.org/10.1016/j.clinph.2023.10.008.
18. Khademi Z, Ebrahimi F, Kordy HM. A review of critical challenges in MI-BCI: From conventional to deep learning methods. J Neurosci Methods. 2023 Jan 1;383:109736. https://doi.org/10.1016/j.jneumeth.
19. Yang S, Yu X, Zhou Y. LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp review dataset as an example. In: 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). IEEE; 2020. p. 98–101.
20. Rivas F, Sierra-Garcia JE, Camara JM. Comparison of LSTM- and GRU-Type RNN networks for attention and meditation prediction on raw EEG data from low-cost headsets. Electronics (Basel). 2025 Feb 12;14(4):707. https://doi.org/10.3390/electronics14040707.
21. Samanta K, Chatterjee S, Bose R. Neuromuscular disease detection based on feature extraction from time–frequency images of EMG signals employing robust hyperbolic Stockwell transform. Int J Imaging Syst Technol. 2022 Jul 25;32(4):1251–62. https://doi.org/10.1002/ima.22709.
22. Roy SS, Dey D, Karmakar A, Roy AS, Ashutosh K, Choudhury NR. Detection of abnormal electromyograms employing DWT-based amplitude envelope analysis using Teager energy operator. Int J Biomed Eng Technol. 2022;40(3):224. https://doi.org/10.1504/ijbet.2022.126493
23. Dubey R, Kumar M, Upadhyay A, Pachori RB. Automated diagnosis of muscle diseases from EMG signals using an empirical mode decomposition-based method. Biomed Signal Process Control. 2022 Jan 1;71:103098. https://doi.org/10.1016/j.bspc.2021.103098
24. Kamali T, Stashuk DW. Electrophysiological muscle classification using multiple instance learning and unsupervised time and spectral domain analysis. IEEE Trans Biomed Eng. 2018 Nov;65(11):2494–502. https://doi.org/10.1109/TBME.2018.2802200.
25. Nandedkar SD, Barkhaus PE, Stålberg E V. Motor unit recruitment and firing rate at low force of contraction. Muscle Nerve. 2022 Dec;66(6):750–6. https://doi.org/10.1002/mus.27737.
26. Masakado Y. Motor unit firing behavior in man. Keio J Med. 1994 Sep;43(3):137–42. https://doi.org/10.2302/kjm.43.137.
27. Van Putten MJAM, Olbrich S, Arns M. Predicting sex from brain rhythms with deep learning. Sci Rep. 2018 Feb 15;8(1):3069. https://doi.org/10.1038/s41598-018-21495-7.

ORIGINAL PAPER

JMS Journal of Medical Science

Gold nanoparticles capped with sulfamethoxazole-ovotransferrin conjugate as a potential nanomedicine for the treatment of microbial infections

Hisham R. Ibrahim
Department of Biochemistry and Biotechnology, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan
https://orcid.org/0000-0002-7809-900X
Corresponding author: k2504042@kadai.jp

Takemichi Kubo
Department of Biochemistry and Biotechnology, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan
–

https://doi.org/10.20883/medical.e1369

Keywords: Ovotransferrin, gold nanoparticles, drug-delivery, sulfamethoxazole, antibiotic, antibacterial, intracellular infection

Received 2025-07-29
Accepted 2025-10-14
Published 2025-12-29

How to Cite: Ibrahim HR, Kubo T. Gold nanoparticles capped with sulfamethoxazole-ovotransferrin conjugate as a potential nanomedicine for the treatment of microbial infections. Journal of Medical Science. 2025 December;94(4);e1369. doi:10.20883/medical.e1369

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Introduction. Although sulfonamide antibiotics are potent antimicrobial agents against bacterial infections, their water insolubility and toxicity at high doses limit their therapeutic efficacy. This study explores a potent anti-infection drug-delivery system using gold nanoparticles (gNPs) capped with ovotransferrin (OTf) as a targeting carrier and solubilising agent for sulfamethoxazole (SMZ) antibiotic.

Materials and methods. The OTf was conjugated with sulfamethoxazole OTf(SMZ) at pH 9.0 for 24 h at 29°C. The conjugate (OTf(SMZ)) or free OTf was added to the gold chloride solution containing sodium citrate as a reducing agent, and then stirred for 24 h at 37°C. The gold nanoparticle (GNP) formulations were purified by gel filtration. The gNP formulations and their individual components (free OTf, SMZ, and gNP tested separately) were evaluated against several microbial strains, including drug-resistant Salmonella, and against bacteria that infect human cells intracellularly.

Results. The gold nanoparticle capped with OTf loaded with SMZ [OTf(SMZ)-gNP] showed superior microbicidal activity against several bacterial strains and the fungi *Candida albicans* compared to the activities of gNP capped with OTf alone [OTf-gNP] or the individual agents. The wild-type *Salmonella* enteritidis, which encodes the multidrug efflux channel TolC, becomes susceptible to both OTf(SMZ)-gNP and OTf-gNP nanoformulation but not to their separate components. However, the *tol*C-knockout *Salmonella* enteritidis mutant strain was susceptible only to the OTf(SMZ)-gNP, indicating the ability of this nanoformulation to deliver the antibiotic SMZ into bacterial cells through self-promoted uptake. The OTf(SMZ)-gNP efficiently killed pathogens intracellularly infecting human colon carcinoma cells.

Conclusions. The results demonstrate that OTf(SMZ)-gNP nanomaterials can mediate the endocytosis of SMZ, making them suitable for targeting bacterial infections, including those that have acquired antibiotic resistance. The study highlights the potential of OTf-capped nanomaterials that can be engineered to enhance the potency of hydrophobic antibiotics for treating infectious diseases.

Introduction

Although many antibiotics, such as sulfonamides, have been effective in treating bacterial infections, antibiotic resistance remains a public health concern. This concern thus urges the development of a new strategy to combat bacterial infections. Sulfamethoxazole (SMZ) is a sulfonamide antibiotic used to treat a variety of infections with a broad antimicrobial spectrum, where it is effective against both gram-negative and gram-positive bacteria [1]. The mechanism of SMZ action involves inhibiting the production of the dihydrofolate intermediate, which interferes with the normal bacterial synthesis of folic acid, an essential component for bacterial growth and replication. Sulfonamide drugs can potentially cause allergic and toxic effects, and the factors influencing the toxicity include the dosage of the drug and its low solubility in blood [4]. The relatively poor water solubility of SMZ may contribute to toxicity due to the high concentration of acetylated groups. Therefore, drug-targeting is a practical approach in reducing allergenicity and toxicity, as the dosage of the drug can be reduced, and its efficacy is locally enhanced around the site of infection.

In a previous study, SMZ was rendered completely water-soluble by loading it into the molecule of ovotransferrin (OTf), and the OTf(SMZ) conjugate proved to be more potent antimicrobial than the individual components (free OTf, SMZ, and gNP) when tested separately, even against the antibiotic-resistant *Salmonella enteritidis* strain [5]. Furthermore, OTf has been proven effective in targeting anticancer drugs to colon and breast cancer cells through the transferrin receptor (TfR) [6]. However, the amount of drug that could be internalised into bacterial or infected cells was limited by the number of OTf-TfR complexes. Consequently, it may become necessary to make adjustments in dosing regimens of the drug by using nanoparticles, which have been considered valid for high drug loading [7]. Metallic nanoparticles have been studied as antibiotic carriers, including gold, silver, copper, titanium and zinc metal nanoparticles, which can penetrate intracellularly and exert antimicrobial properties [8,9].

It has been reported that gold nanoparticles exhibit direct permeation across cell membranes compared to other materials [10. Nanoparticle translocation across the cell membrane has been reported to occur through two pathways: endocytosis and direct diffusion [8]. The nanoparticle binds receptors in the membrane via multivalent interactions and captures several tens of receptors at any point [11]. By doing so, the particle deforms the membrane underneath it, thereby improving the therapeutic index of the drug payload compared to free drugs. Direct administration of antibiotics faces various barriers which prevent effective localisation of the antibiotic and reduce its retention [5]. An approach to overcome poor antibiotic retention and promote its effective localisation to the pathogen is to integrate nanoparticles with cell-targeted drug delivery systems.

In this study, we investigate the development of an effective drug-targeting strategy to deliver the water-insoluble SMZ directly to pathogens or intracellularly infected mammalian cells. Specifically, this study aimed to develop and characterise gold nanoparticles coated with the OTf(SMZ) complex as a targeted antibiotic delivery system and to evaluate their efficacy against resistant bacterial strains and intracellular infections. The strategy involves loading SMZ into the OTf molecule and then tagging gold nanoparticles with the SMZ-OTf conjugate. Most pathogens, as well as human epithelial cells, are known to express the transferrin receptor (TfR) and selectively take up OTf [5]. Thus, the SMZ-OTf conjugate is directly localised to pathogens or to the cells in which they reside. Therefore, while nanoparticles promote direct diffusion, effective dosage, and retention of the drug, they may also help alleviate antibiotic resistance. Through this approach, the gold nanoparticle SMZ-OTf demonstrated the ability to specifically deliver a high dose of antibiotics to, thereby overwhelming the efflux pumps that expel antibiotics from drug-resistant wild-type *Salmonella enteritidis* cells.

Materials and methods

Materials

Ovotransferrin, purchased from Inovatech Bio-Products Inc. (Abbotsford, Canada), was recrystallised and purified chromatographically using a Sephadex G-50 column (G5080). Sulfame-

thoxazole, SMZ (31737), gold chloride trihydrate, $HAuCl_4 \cdot 3H_2O$ (520918), fetal bovine serum, FBS (TMS-013), and trypan blue, TB (T6146), were from Sigma (St. Louis, MO). McCoy 5a medium (16600082) was from Invitrogen, Japan. CellTiter-Blue cell viability assay (G8080) was from Promega, Japan. Brain Heart Infusion, BHI (05509), trypticase soy broth, TSB (05518), and nutrient agar (05514) were obtained from Nissui (Tokyo, Japan). Unless otherwise specified, all other reagents were of analytical grade.

Microorganisms and cell lines

Microorganisms used for antimicrobial assays, *Salmonella enteritidis* (IFO 3313) and *Escherichia coli* K-12 (IFO 3301), were obtained from the Institute of Fermentation, Osaka, Japan. *Staphylococcus aureus* (NBRC 14462) and *Corynebacterium minutissimum* (NBRC 15361) were from the Nite Biological Resource Centre (Tokyo, Japan). *Propionibacterium acnes* (ATCC 6919), *Candida albicans* (ATCC 2091), and *Staphylococcus epidermidis* (ATCC 12228) were from the American Type Culture Collection (Rockville, MD, USA). The wild strain of *Streptococcus zooepidemicus* was obtained from the Institute of Bacteriology of the Veterinary Hospital, Zürich (Switzerland). The virulent wild-type of *Salmonella enterica* serotype Enteritidis phage type 4 strain 147 (147^{str}), originally isolated from egg content [12]. The *tol*C-deficient mutant *Salmonella Enteritidis* 147^{str} strain (Δ*tol*C), prepared by a one-step inactivation method [13], was a gift from the Department of Pathology, Bacteriology, and Avian Diseases, Faculty of Veterinary Medicine, Ghent University, Belgium.

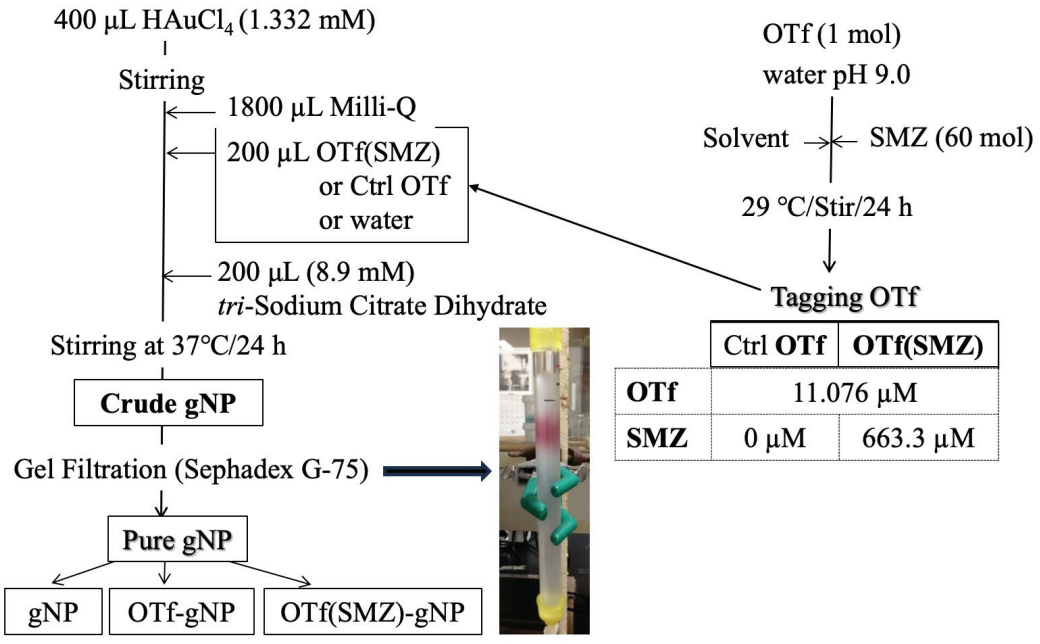
Cell lines

Human colon carcinoma cell line HCT-116 (ATCC# CCL-247) from American Type Culture Collection (Rockville, MD) were maintained in McCoy's 5a media with 10% fetal bovine serum (FBS) and antibiotics (penicillin and streptomycin) at 37˚C in a humidified incubator with 5% CO_2. The medium was changed every other day.

Methods

Preparation of SMZ-OTf gold nanoparticle conjugate

Conjugation of OTf with sulfamethoxazole OTf(SMZ) was prepared as previously described [5]. Briefly, a freshly prepared antibiotic stock solution in 50% ethanol was added to the OTf solution to achieve a 60-fold molar excess of antibiotic over OTf in distilled water at pH 9.0, with stirring for 24 h at 29°C. The concentration of OTf in the mixture was 0.864 mg/mL. The prepara-



Scheme 1. Preparation of the OTf-SMZ tagged gNP. Concentration of OTf(SMZ) and control OTf are adjusted to 0.864 mg/mL (OTf- based).

tions were referred to as OTf(SMZ) complex and control (Ctrl OTf). Ctrl OTf was treated similarly except without an antibiotic.

To a 400 µL gold chloride solution (1.332 mM), 200 µL of 11.076 µM OTf(SMZ) or Ctrl OTf were added to 1800 µL Milli-Q water. For gold nanoparticles alone (gNP), water pH 9.0 was added. After adding 200 µL of sodium citrate dihydrate (8.9 mM) as a reducing agent, the mixture was stirred for 24 h at 37°C. This crude GDP formulation was purified by passing it through a gel-filtration column on Sephadex G-75. The molar ratio of gold chloride to OTf was 120:1 (mol/mol) in the respective OTf-tagged gold nanoparticle formulation. The gold nanoparticle formulations tagged with the OTf(SMZ) conjugate, OTf, or lacking the tag molecule were referred to as OTf(SMZ)-gNP, OTf-gNP, and gNP, respectively. The preparation procedure is detailed in **Scheme 1**.

Loading efficiency of SMZ into OTf(SMZ)-tagged gNPs

The crude gNP formulations were subjected to centrifugation at 15,000 × g for 30 min, and the supernatant was used to evaluate the SMZ loading efficiency onto OTf(SMZ)-gNPs. The total amount of free SMZ in the supernatant was determined spectrophotometrically at a maximum wavelength of 262 nm and quantified with an absorption molar coefficient of 11400 mol^{-1} L cm^{-1} [14]. The following equation was applied to calculate the loading efficiency:

$$\text{Loading efficiency (\%)} = [(\text{total SMZ} - \text{free SMZ}) / \text{total SMZ}] \times 100$$

Ultraviolet-visible light absorption spectrum analysis

To measure the adsorption of substances on the surface of the gNP preparations, the surface plasmon resonance peak unique to gNPs was measured using a SmartSpec-3000 spectrophotometer (Bio-Rad, USA). 150 150-µl of each sample was added to a quartz cell with an optical path length of 1.0 cm, and measurements were taken in the wavelength range of 300–800 nm.

Binding of OTf to gNPs

The confirmation of OTf incorporation into gNPs was verified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins that bind to the nanoparticle surface were isolated using the method of Saptarshi et al. [15]. Briefly, OTf-gNP preparation was denatured with 2-mercaptoethanol and centrifuged to recover the total OTf into the supernatant (Total OTf). Another portion of each preparation was centrifuged at 10,000 rpm for 10 minutes to remove the unbound OTf. The precipitate was denatured with 5% 2-mercaptoethanol (equal to the original volume) and centrifuged again to recover the bound OTf into the supernatant (Bound OTf), and the precipitate was the gNP. The precipitated gNP was washed with 5% 2-mercaptoethanol and centrifuged again to confirm recovery of all bound OTf into the supernatant (Wash). The collected supernatants were run on a 4% separation gel and a 15% concentration gel, at currents of 10 mA and 25 mA, respectively. After electrophoresis, the gel was stained with Coomassie Brilliant Blue (CBB) and destained with a 20:10:70 (v/v/v) methanol: acetic acid: water solution.

Transmission (TEM) and scanning (SEM) electron microscopy analysis

The size and shape of the formulated nanoparticles were analysed using a JEM-3010 transmission electron microscope (JEOL, Japan). A 2-µL sample was dropped onto a mesh and dried in a desiccator at room temperature for 24 hours. The acceleration voltage was 200 kV, and observations were made at magnifications of 20,000x and 100,000x. Particle diameter was determined from the images by comparing the measured data of 10 randomly selected particles with the scale bar length value. The average particle diameter of the nanoparticles was statistically calculated based on image data obtained at high magnification, and all values are presented as the mean value ± standard deviation (SD; n = 10). The morphology of nanoparticles was characterised using a SU-70 scanning electron microscope (SEM) operated at 15 kV (Hitachi High-Tech, Japan).

Antibacterial assay

The antibacterial activity of the nanoparticles against *Staphylococcus aureus, Staphylococcus epidermidis*, *Streptococcus zooepidermicus*, *Corynebacterium minutissimum*, *Propionibacterium acnes*, *Escherichia coli*, *Salmonella enteritidis* (wild-type), and *tol*C-deficient mutant

Salmonella enteritidis (Δ*tol*C) was tested using a liquid medium method. The mid-logarithmic phase cells, grown in brain heart infusion (BHI) broth, were washed and resuspended (to give 10^6 ~ 10^7 CFU/mL) in 1% trypticase soya broth (TSB), pH 7.3. Each inoculum was mixed with an equal volume of nanoparticle preparations, SMZ alone, or OTf equivalent to their contents in the respective nanoparticle. After incubation at 37°C for 24 h, a 30-µl portion or dilutions in physiological saline were spotted in triplicate onto a nutrient agar plate. The colony-forming units (CFU) were obtained after incubating agar plates at 37°C for 18 hours.

For *Candida albicans*, the blastoconidia grown at 28˚C for 36 h in 1% TSB, pH 7.3, were washed, resuspended (to a 10^6 ~ 10^7 CFU/mL) in Sabouraud dextrose broth (SAB) and mixed with an equal volume of nanoparticles or SMZ or OTf equivalent to their concentration in each derivative. After a 24-hour incubation at 28˚C, the mixture was plated on SAB agar. The plates were incubated for 24 h at 30°C, and the colonies were counted. All results are expressed as log CFU/mL as a function of OTf concentration in the assay medium of three independent experiments.

Intracellularly infected human cells

To elucidate the ability of OTf(SMZ)-gNP to deliver SMZ to intracellularly residing bacteria, human colon epithelial carcinoma cells, HCT-116, were infected with *Staphylococcus aureus* or *Escherichia coli,* and the intracellular bacteria were counted on agar plates. Briefly, Bacteria were grown in TSB broth at 37˚C for 20 h. The bacteria were washed twice and suspended in PBS buffer (pH 7.4). HCT-116 cells were seeded at 5 × 10^4 cells/well in 96-well plates in McCoy's 5A medium until they reached 80% confluence. Bacteria were adjusted to 10^{4-10^5} CFU/mL in PBS buffer and added to HCT-116 cells, which were then incubated at 37˚C under 5% CO_2. After 2 hours, the medium was replaced with McCoy's 5A containing 100 µg/mL gentamicin and incubated for 1 hour to kill extracellularly adherent bacteria. The cells were then washed three times with PBS and treated with medium containing nanoparticle derivatives (100 µg/mL final concentration, OTf-based) or their OTf and SMZ equivalents, and incubated for two h at 37˚C under 5% CO_2. Cells were lysed by adding 0.8 mL of 1% Triton X-100 in PBS for 15 minutes. A portion of suspension was serially diluted with sterilised physiological saline and spotted in triplicate on TSA agar. The plates were incubated for 24 hours at 37°C, and the colony-forming units (CFU) were calculated. The results are expressed as (log CFU/10^6 cells) of two independent experiments with three wells for each derivative.

In a clear-bottom, black 96-well plate, three wells were treated similarly for each sample used in the HCT-116 cell viability assay. Cell viability was determined using the fluorescent CellTiter-Blue Assay (Promega). The results are expressed as the percentage of cell viability relative to the control. The data are presented as the mean ± SD.

Statistical analysis

Experiments were carried out in duplicate or triplicate, and mean values were used for statistical analysis. Data obtained in the study were analysed statistically using one-way ANOVA by Excel's data analysis tools. Data shows mean values, and error bars show standard deviations.

Results

Synthesis of gold nanoparticles (gNP) capped with OTf and its SMZ complex

The OTf-capped [OTf-gNP] or OTf(SMZ)-capped [OTf(SMZ)-gNP] gold nanoparticles were synthesised by the redox method at ambient temperature using sodium citrate, as outlined in **Scheme 1**. Colourimetric shift in the UV-VIS region (300-800 nm) of the formed gNPs was employed to probe incorporation of OTf onto the gNP core shell (**Figure 1**). Due to the surface plasmon resonance of colloidal gold, the absorbance peak exhibits a variation in λmax depending on the size of the nanoparticles [16]. The absorbance peak of the pure gNP was observed at a maximum of 535 nm. The incorporation of OTf [OTf-gNP] resulted in a red shift in the spectra band from λmax of 535 to 538 nm. While OTf loaded with SMZ [OTf(SMZ)-gNP] resulted in a broader peak with red shift of λmax from 537 to 554 nm. The colour change (inset photos) and red shift of the surface plasmon resonance are characteristics of an increase in the gNP size, where the red shift correlates more strongly with the size [17]. These changes in

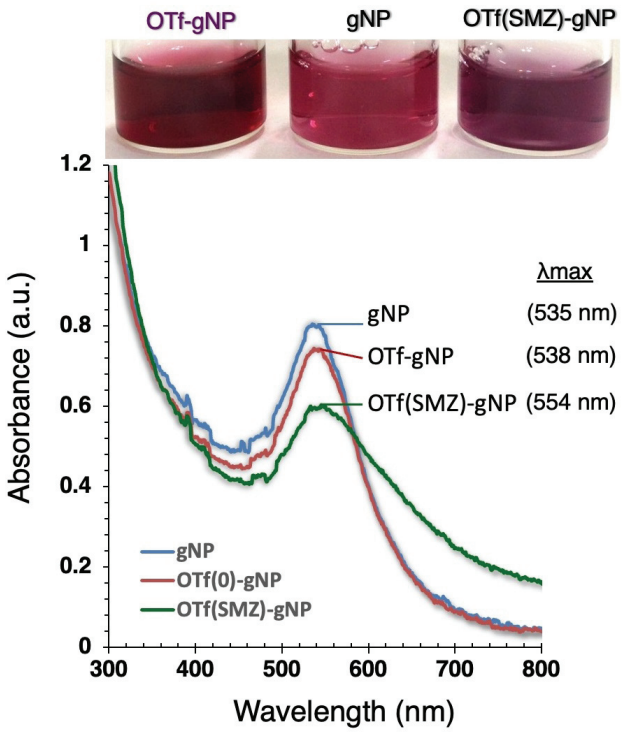


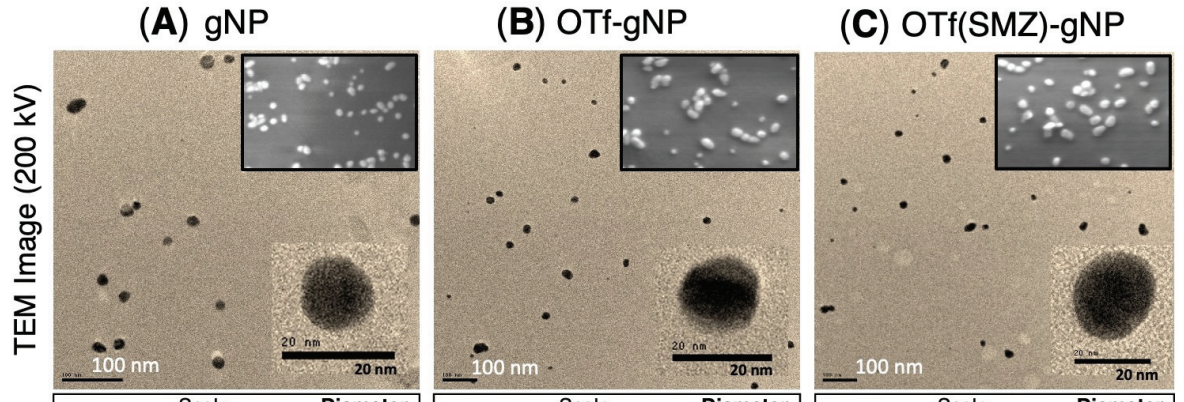
Figure 1. UV-VIS absorption spectra of gold nanoparticles synthesised in sodium citrate in the absence (gNP) and presence of OTf (OTf-gNP) or SMZ-loaded OTf (OTf(SMZ)-gNP). Each spectrum represents 24 24-hour reaction times. The inset photo shows the colour of nanoparticles formed in each preparation.

λmax suggested the formation of nanoparticles with a larger diameter. The λmax of OTf(SMZ)-gNP was shifted to a higher wavelength (red shift) and showed a broader peak, which could be attributed to the formation of larger particle sizes. The size and morphology of the nanoparticles were further characterised by transmission electron microscopy (TEM) and scanning electron microscopy (SEM) techniques (**Figure 2**). The addition of OTf or OTf(SMZ) increased the diameters of the nanoparticles. As shown in **Figure 2**, the particles were multi-dimensional, and the mean diameters of gNP (A), OTf-gNP (B), and OTf(SMZ)-gNP (C) nanoparticles were 15.72, 1.71, 19.32, 3.76, and 19.76, ±3.53 nm, respectively. The sizes of particles ranged between maximum diameters of 19.0 nm, 28.0 nm, and 24.0 nm to minimum diameters of 13.6 nm, 14.4 nm, and 13.0 nm for gNP, OTf-gNP, and OTf(SMZ)-gNP, respectively. The particles in all formulations were irregularly shaped, quasi-spherical, with most of the particles being spherical or ellipsoid in shape (**Figure 2**, inset images). Densitometric analysis of SDS-PAGE of the OTf-gNP nanoparticles (**Figure 3**) indicated that 36.29% of the total OTf is absorbed to the gNP, whereas 20.1% is monomer and 16.19% an aggregated form of bound OTf.

The determination of the loading efficiency of SMZ onto OTf(SMZ)-gNP confirmed a significant loading efficiency of ~68%. Initially, 168 µg/mL of SMZ was used in the formulation mixture, and 114 µg/mL of the SMZ remained attached to 0.84 mg/mL OTf.

Antibacterial activity of OTf(SMZ)-capped gold nanoparticles

We tested the bactericidal activities of the gold nanoparticle preparations against different bacterial strains (*Staphylococcus epidermidis*, *Staphylococcus aureus*, *Streptococcus zooepidemicus, Corynebacterium minutissimum*, *Propionibacterium acnes*, *Candida albicans, Escherichia coli* K-12, and *Salmonella enteritidis*). The activity was expressed as log CFU/mL as a function of OTf concentration in the nanoparticles. Free OTf and SMZ were tested at equivalent concentrations as in the nanoparticles. As shown in **Figure 4**, the OTf(SMZ)-gNP exhibited vigorous bacteri-



No.	Mag.	Scale bar	Scale	Diameter (nm)
1	100k	20	8.2	16.4
2	100k	20	7.5	15
3	100k	20	7.5	15
4	100k	20	7.3	14.6
5	100k	20	6.8	13.6
6	100k	20	7	14
7	100k	20	8	16
8	100k	20	9.5	19
9	100k	20	9	18
10	100k	20	7.8	15.6
Mean (nm)				**15.72**
SD				1.71
Max (nm)				19
Min (nm)				13.6

No.	Mag.	Scale bar	Scale	Diameter (nm)
1	100k	20	8.8	17.6
2	100k	20	9	18
3	100k	20	8.9	17.8
4	100k	20	7.2	14.4
5	100k	20	8.2	16.4
6	20k	100	1.8	18
7	20k	100	2	20
8	20k	100	2.2	22
9	20k	100	2.8	28
10	20k	100	2.1	21
Mean (nm)				**19.32**
SD				3.76
Max (nm)				28
Min (nm)				14.4

No.	Mag.	Scale bar	Scale	Diameter (nm)
1	100k	20	11	22
2	100k	20	10.8	21.6
3	25k	100	2	20
4	25k	100	2.2	22
5	25k	100	2.1	21
6	25k	100	2	20
7	25k	100	2.4	24
8	25k	100	1.4	14
9	25k	100	1.3	13
10	25k	100	2	20
Mean (nm)				**19.76**
SD				3.53
Max (nm)				24
Min (nm)				13

Figure 2. Nanoparticle morphology represented by TEM images of gNP (A), OTf-gNP (B) and OTf(SMZ)- gNP (C), as well as their SEM images (inset). Nanoparticle diameters of randomly selected ten particles, along with their mean diameters, are shown below the pictures. It can be seen that the gold nanoparticles are irregularly shaped, quasi-spherical, and within the nano-size range.

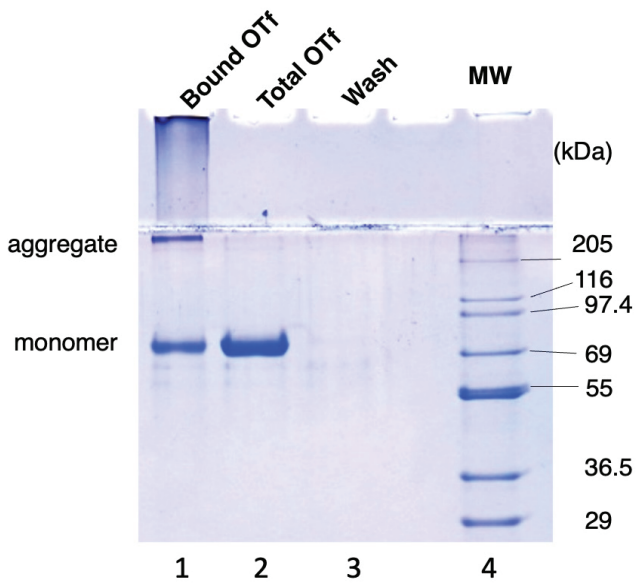


Figure 3. SDS-PAGE of OTf protein coronas formed in OTf-gNP formulation. Lane 1 is the tightly bound OTf to gNP. Lane 2 is the total OTf (free and bound OTf). Lane 3 confirms the total dissociation of OTf. Lane 4 is a protein marker.

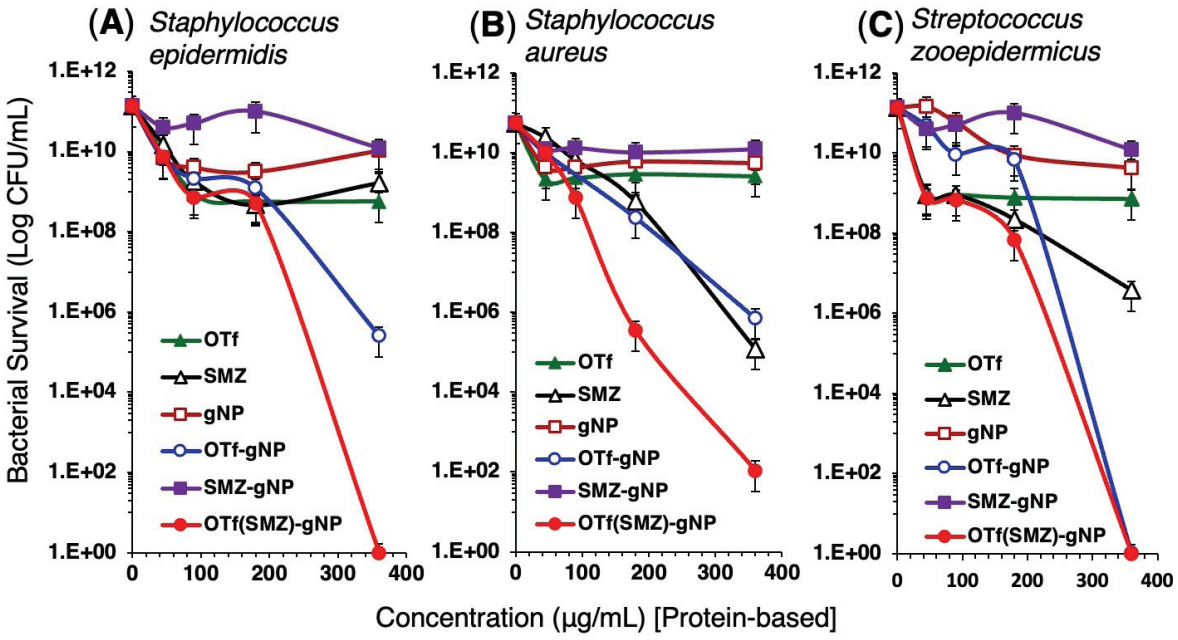


Figure 4. Bactericidal activity of various gNP formulations against three Gram-positive bacteria. The assay was performed against *Staphylococcus epidermidis* (A), *Staphylococcus aureus* (B), and *Streptococcus zooepidemicus* (C) at different doses of gNPs, as well as free SMZ and OTf at concentrations equivalent to their content in the respective formulations. The data is presented as log CFU/mL.

cidal activity, whereas severely reduced survival of the Gram-positive *Staphycoccus epidermidis* (A), *Staphycoccus aureus* (B), and *Streptococcus zooepidermicus* (C) in a dose-dependent manner. Interestingly, the OTf-gNP lacking the antibiotic exhibited remarkable dose-dependent bactericidal activity against the three strains, albeit weaker than that of OTf(SMZ)-gNP. Although free OTf and SMZ exhibited bactericidal activities, they were much weaker than OTf-gNP and OTf(SMZ)-gNP, and the effectiveness of the free OTf and SMZ varied depending on the strain.

We next tested the bactericidal activity against three strains of skin pathogens: *Corynebacterium minutissimum* (A), *Propionibacterium acnes* (B) and the fungus *Candida albicans* (C), as shown in **Figure 5**. The controls free OTf, OTf(SMZ), and gNP were almost inactive against *C. minutissimum* (**Figure 5A**) but showed weak activity against *P. acnes* and *C. albicans* (**Figures 5B** and **C**). Free SMZ showed moderate activity against the three strains tested. It should be noted that free OTf or OTf(SMZ) showed weak to moderate activity against skin pathogens, but OTf-gNP and OTf(SMZ)-gNP exhibited vigorous, dose-dependent bactericidal activity against the three strains.

While OTf(SMZ)-gNP was significantly potent bactericidal in a dose-dependent manner against *C. minutissimum* and *C. albicans* (**Figures 5A** and **C**), surprisingly, OTf-gNP showed slightly more vigorous activity than OTf(SMZ)-gNP against *P. acnes* (**Figure 5B**). Given this degree of dose-dependency, OTf-gNP appeared to serve as a potent antibacterial agent, becoming even more potent bactericidal when the capping OTf was loaded with the antibiotic SMZ (OTf(SMZ)-gNP). As shown in **Figure 6**, OTf(SMZ)-gNP also exhibited potent, dose-dependent bactericidal activity against the Gram-negative bacteria *Escherichia coli* K-12 (A) and *Salmonella enteritidis* (B). Interestingly, OTf-gNP exhibited relatively strong bactericidal activity against *Salmonella enteritidis,* although it was still less potent than OTf(SMZ)-gNP (**Figure 6B**).

OTf(SMZ) gold nanoparticles against drug-resistant *Salmonella*

We examined the bactericidal potency of OTf-gNP and OTf(SMZ)-gNP against the drug-resistant wild-type (WT) *Salmonella enteritidis* and its *tol*C-deleted mutant (Δ*tol*C). As shown in **Figure 7**, the wild-type *Salmonella* was highly resistant to OTf, OTf(SMZ), free SMZ and to the con-

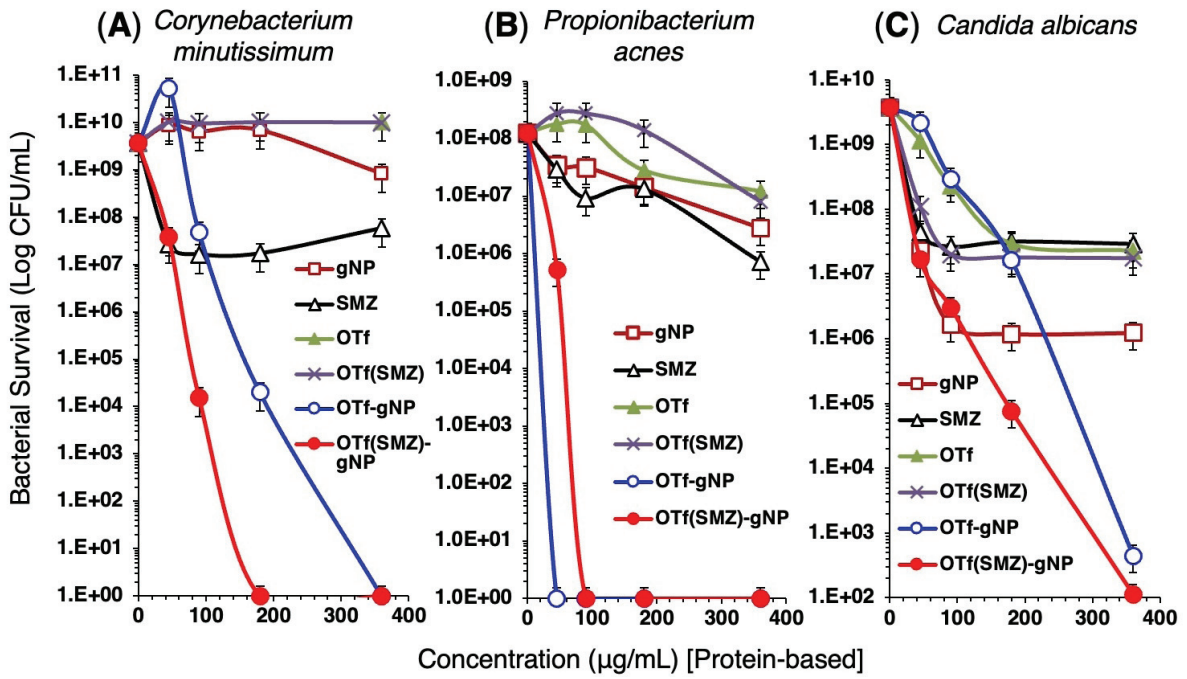


Figure 5. Bactericidal activity of various gNP formulations against skin pathogens *Corynebacterium minutissimum* (A), *Propionibacterium acnes* (B), and fungi *Candida albicans* (C). The assay was performed at different doses of gNPs, as well as free SMZ and OTf, at concentrations equivalent to their content in the respective formulations. The data is presented as log CFU/mL.

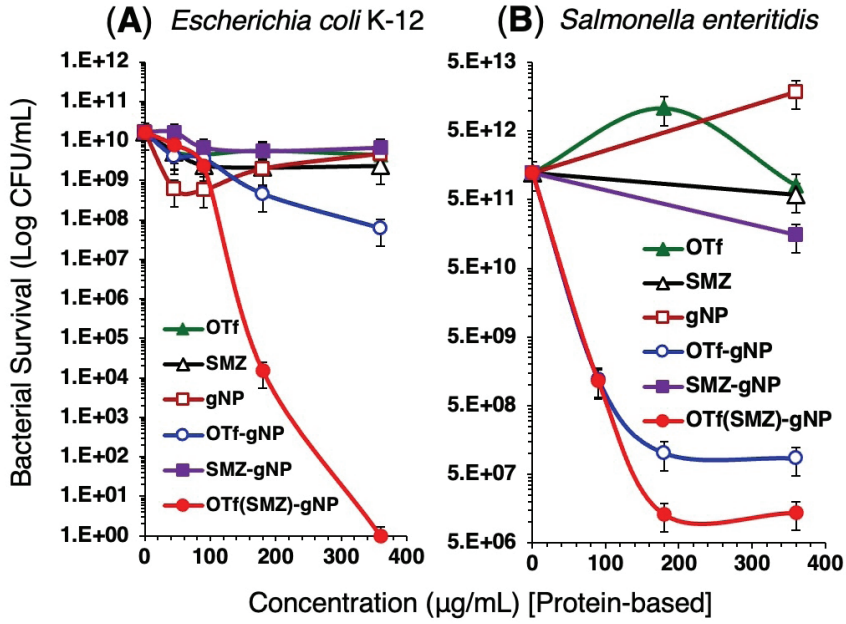


Figure 6. Bactericidal activity of various gNP formulations against Gram-negative *E. coli* (A) and *Salmonella enteritidis* (B). The assay was performed at different doses of gNPs as well as their free SMZ and OTf equivalents. The data is presented as log CFU/mL.

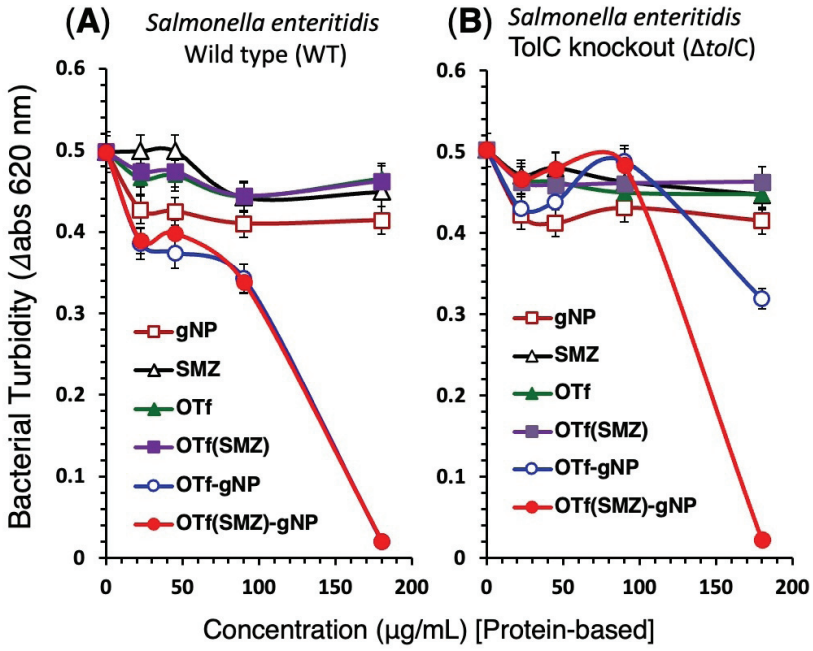


Figure 7. Antimicrobial activity of various gNP formulations against drug-resistant wild-type and *tol*C-deficient *Salmonella enteritidis*. The assay was performed against wild-type, WT (A) and *tol*C-deficient mutant, Δ*tol*C (B) strains at different doses of gNPs as well as their free SMZ and OTf equivalents. The data are presented as bacterial growth monitored at A620 nm.

trol nanoparticles (gNP), while it was sensitive to OTf-gNP and OTf(SMZ)-gNP (**Figure 7A**). Deletion of the *tol*C gene (Δ*tol*C) abolished the antibacterial effect of OTf-gNP against the wild-type (WT) *Salmonella* Enteritidis. Still, the bacteria remained significantly susceptible to OTf(SMZ)-gNP (**Figure 7B**). The results of **Figures 7A** and **B** suggest that OTf-capped gNP (OTf-gNP) is recognised and bound to TolC on the surface of wild-type *Salmonella* cells, where this physical contact effectively induces a killing effect. On the other hand, the bactericidal effect of OTf(SMZ)-gNP on *tol*C knockout *Salmonella* cells (**Figure 7B**) may be attributed to the high concentration of the antibiotic SMZ that is loaded onto the nano-formulation.

Facilitated OTf-gNP delivery of SMZ to intracellularly residing pathogens

The anti-infective effect of OTf(SMZ)-gNP on *S. aureus* (**Figure 8**) and *E. coli* (**Figure 9**) infecting HCT-116 cells intracellularly. The results were expressed as colony-forming units (CFU) per $10^{\wedge 6}$ cells. The free gNP showed a weak effect on the survival of the two intracellularly residing bacterial strains, comparable to mock-treated cells (Ctrl). The free OTf or SMZ exhibited moderate reduction of the intracellular bacterial survival of both *S. aureus* (**Figure 8A**) and *E. coli* (**Figure 9A**). While the OTf(SMZ) conjugate significantly decreased the survival of both strains, the OTf(SMZ)-gNP completely eradicated intracellular *S. aureus* (**Figure 8A**) and significantly reduced the intracellular *E. coli* (**Figure 9A**) within two hours of treatment. The OTf-gNP, lacking antibiotics, remarkably reduced intracellular bacterial survival of *S. aureus* (**Figure 8A**) but was less effective against *E. coli* (**Figure 9A**).

Subsequent post-hoc tests were conducted to determine which specific group comparisons were statistically significant. For S. aureus, the OTf(SMZ)-gNP and OTf-gNP treatments resulted in complete eradication or a remarkable reduction of intracellular bacteria. Both treatments were found to be statistically significant when compared to all other groups (Ctrl, free gNP, free OTf, and free SMZ), with a p-value < 0.05. OTf(SMZ) conjugate treatment also resulted in a statistically significant reduction of bacterial survival compared to the control group (Ctrl) and free gNP, with a p-value < 0.05. Free OTf and free SMZ treatments showed a statistically significant

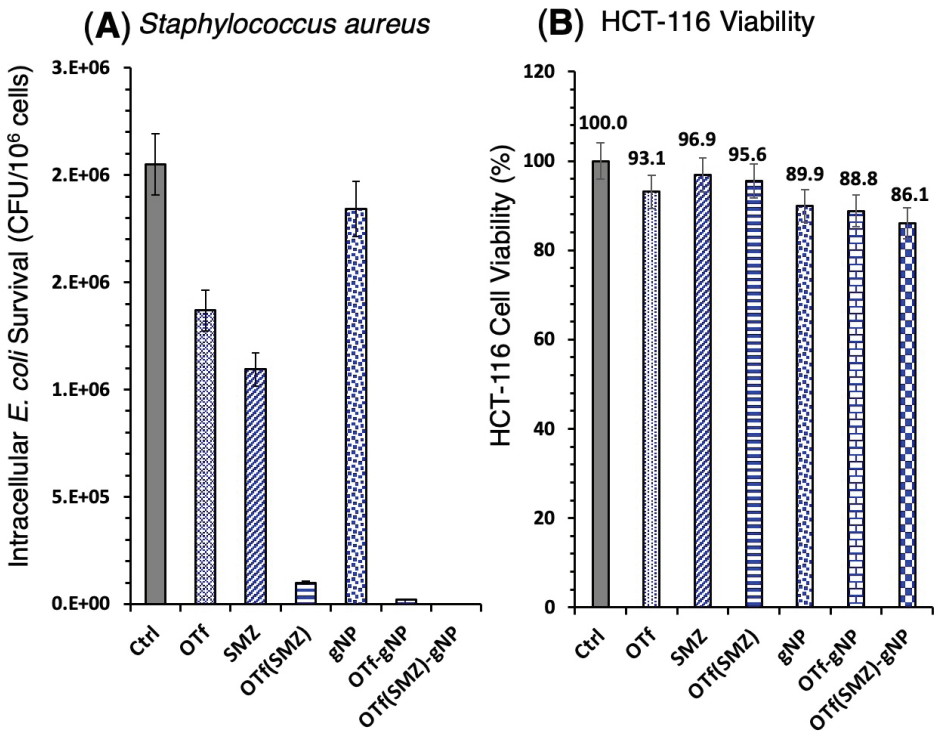


Figure 8. Recovery of *Staphylococcus aureus* (A) from the infected HCT-116 cells and viability of pre-infected HCT-116 cells (B). Cells were inoculated with 5 log cfu bacteria for 2 hours, then the extracellular bacteria were killed by gentamicin. Cells were washed and then treated with various gNP formulations or their OTf and SMZ equivalents and incubated for two hours at 37˚C under 5% CO_2. The HCT-116 cells were disrupted with Triton X-100, and the intracellular viable count was determined on TSA agar. Ctrl is cells treated with sterile distilled water. Values are expressed as the mean CFU per $10^{\wedge 6}$ cells of total protein from two experiments. Viability of pre-infected HCT-116 cells and subsequently treated with the indicated derivative for two hours was quantified by CellTiter Blue assay, and bars represent the mean with $n = 3$.

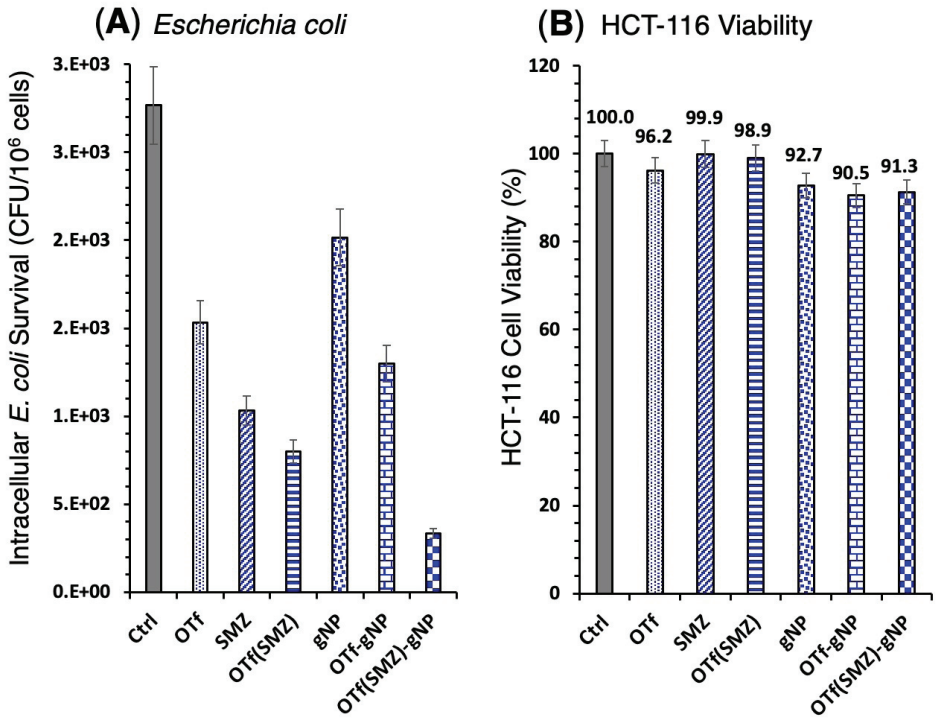


Figure 9. Recovery of *E. coli* (A) from the infected HCT-116 cells and viability of pre-infected HCT-116 cells (B). Experiments were carried out as shown in the Figure. 8.

moderate reduction in bacterial survival compared to the control group (Ctrl), with a p-value < 0.05. The difference between the OTf(SMZ)-gNP and OTf-gNP groups was not statistically significant ($p > 0.05$), suggesting a similar high level of efficacy for both treatments against *S. aureus* infections.

As for E. *coli*, OTf(SMZ)-gNP treatment resulted in a statistically significant reduction of intracellular bacteria compared to all other groups, with a p-value < 0.05. OTf-gNP treatment also caused a decrease in bacterial survival, but this effect was less significant compared to the OTf(SMZ)-gNP group, with a p-value < 0.05. OTf(SMZ) conjugate treatment led to a statistically significant decrease in bacterial survival compared to the control group (Ctrl), with a p-value < 0.05. Free OTf and free SMZ treatments exhibited a statistically significant moderate reduction in bacterial survival compared to the control group (Ctrl), with a p-value < 0.05. These results confirm that the OTf(SMZ)-gNP and OTf-gNP treatments are significantly more effective than the other treatments, with a notable difference in efficacy against *E. coli,* where the addition of SMZ in the OTf(SMZ)-gNP conjugate provides a significantly enhanced effect. The higher susceptibility of *S. aureus* to OTf-gNP compared to *E. coli* is a key finding, possibly due to the presence of transferrin receptors on the *S. aureus* surface.

To evaluate the effect of nanoparticle formulations on infected HCT-116 cells, cell viability was determined using the CellTiter Blue assay. As shown in **Figures 8B** and **9B**, OTf, SMZ, and their conjugate OTf(SMZ) exhibited negligible to weak inhibition of HCT-116 cell growth, while gNP, OTf-gNP and OTf(SMZ)-gNP showed moderate growth inhibition of HCT-116 cells. After treating for 2 hr with *S. aureus*-infected HCT-116, the growth inhibitory rates of OTf, SMZ, OTf(SMZ), gNP, OTf-gNP and OTf(SMZ)-gNP were 93.1%, 96.9%, 95.6%, 89.9%, 88.8% and 86.1%, respectively. Post-hoc analysis using Tukey's HSD test revealed several statistically significant differences in growth inhibition rates of HCT-116 cells. Specifically, the nanoparticle formulations, gNP ($p = 0.012$), OTf-gNP ($p = 0.008$), and OTf(SMZ)-gNP ($p = 0.003$), showed significantly lower growth inhibitory rates compared to the non-nanoparticle formulations OTf, SMZ, and OTf(SMZ). There were no statistically significant differences found among the nanoparticle formulations themselves or among the non-nanoparticle formulations. After treating HCT-116 cells infected with *E. coli* for 2 hours, the growth inhibitory rates of OTf, SMZ, OTf(SMZ), gNP, OTf-gNP, and OTf(SMZ)-gNP were 96.2%, 99.9%, 98.9%, 92.7%, 90.5%, and 91.3%, respectively. A Tukey's HSD post-hoc test showed a statistically significant difference between the nanoparticle formulations and the non-nanoparticle formulations. Specifically, the growth inhibition rates of gNP ($p < 0.001$), OTf-gNP ($p < 0.001$), and OTf(SMZ)-gNP ($p < 0.001$) were significantly lower than those of OTf, SMZ, and OTf(SMZ). These results suggest that while all tested formulations exhibit some level of growth inhibition, the nanoparticle-based formulations are consistently less effective at inhibiting cell growth in both *S. aureus*- and *E. coli*-infected HCT-116 cells.

Discussion

In this study, we prepared gold nanoparticles (gNPs) capped with the OTf molecule alone (OTf-gNP) or OTf preloaded with the water-insoluble antibiotic SMZ [OTf(SMZ)-gNP] as a potent antimicrobial formula. The gNPs were synthesised by the redox method using citrate as an environmentally friendly reducing agent. The OTf molecule was adsorbed onto the gNPs, increasing the diameter of gNPs from 15.72 nm to 19.32 nm with OTf and to 19.76 nm with SMZ-OTf conjugate. Both the OTf(SMZ)-gNP and OTf-gNP gold nanoparticles demonstrated potent antimicrobial activity against various bacterial isolates at low doses. At a dose of 180 μg/mL, OTf(SMZ)-gNP caused 2.43-log, 5.19-log, 3.30-log, 6.03-log and 4.98-log reduction of *S. epidermidis, S. aureus, S. zooepidermicus, E. coli* and *Salmonella enteritidis*, respectively. On the other hand, OTf-gNP at the same dose caused 2.06-log, 2.37-log, 1.31-log, 1.56-log and 4.08-log reduction of *S. epidermidis, S. aureus, S. zooepidermicus, E. coli* and *Salmonella enteritidis*, respectively. OTf(SMZ)-gNP exhibited more potent antimicrobial activity than OTf-gNP, particularly against the skin pathogens *S. epidermidis*, *S. aureus*, *C. minutissimum* and the fungus *C. albicans*. OTf(SMZ)-gNP caused 9.56-log, 8.09-log, and 4.69-log reduction at 180 μg/mL of the skin pathogens *C. minutis-*

simum, P. acnes and C. albicans, respectively. At the same time, OTf-gNP at the same dose caused 5.26-log, 8.09-log and 2.35-log reduction of *C. minutissimum, P. acnes and C. albicans,* respectively. The differences in antibacterial action between OTf(SMZ)-gNP and OTf-gNP are related to the intrinsic properties of gNPs or the nature of the cell membrane and biomolecules within the microorganism [18].

The gNPs serve as carriers for the antibacterial OTf, resulting in significantly stronger antimicrobial effects when OTf is loaded with the antibiotic SMZ. Notably, the formulated nanoparticles exhibited excellent activity against drug-resistant *Salmonella*. Both *Escherichia coli* K-12 and *Salmonella enteritidis* are known to possess outer membrane components such as TolC involved in the AcrAB-TolC multidrug resistance channel [19,20]. TolC is involved in the export of siderophores, components with a high affinity for iron [21,22], and has been reported to be activated when contacted with transferrin [23,24]. Therefore, we. Conducted tests to determine whether TolC plays a role in the bactericidal activity of OTf-capped nanoparticles. Both OTf(SMZ)-gNP and OTf-gNP caused complete inhibition (96% inhibition) of the drug-resistant wild-type *Salmonella*. Interestingly, OTf(SMZ)-gNP exhibited the same growth inhibitory efficacy (96% inhibition), but OTf-gNP showed significantly reduced inhibitory (36% inhibition) of the *tol*C-deleted *Salmonella* mutant (Δ*tol*C). These results align with a previously reported study on the mechanism of NPs affecting antibiotic resistance [25]. It has been revealed that NPs can be densely packed on the outer membrane, and some of them penetrate and kill bacterial cells, where the effects depend on the size of the NPs and the nature of the molecules at the surface of the bacterial cell [10]. Conceivably, the gold nano-formulation concentrated SMZ on the bacterial cell wall by passing the drug efflux pump (TolC) through passive delivery. The OTf on the surface of OTf-gNP interacted multivalently with the bacterial surface receptors, thereby inducing bacterial growth inhibitory action. Particularly, gNPs are known to enhance the permeability of bacterial cell membranes [8,10,11,26]. According to a study on Gram-negative bacteria, nanoparticles effectively eliminate multidrug-resistant isolates, and their action is mediated by the TolC efflux pump 23, [19,23,27]. This conclusion is supported by the observation that gNP alone exhibited significantly weaker antibacterial activity than the OTf(SMZ)-gNP and OTf-gNP nano-formulations. The superior antibacterial activities of OTf(SMZ)-gNP and OTf-gNP against *S. aureus* and skin pathogens *C. minutissimum* and *P. acne* could contribute to the treatment of skin wound healing.

Interestingly, both OTf(SMZ)-gNP and OTf-gNP exhibited excellent antimicrobial activity against the yeast/fungal *C. albicans*. Notably, the OTf(SMZ)-gNP demonstrated a more potent effect at considerably lower doses, indicating the contribution of SMZ loaded in the formulation. A recent study on the mechanisms of gNPs against *Candida* demonstrated that gNPs can interact with the H^+-ATPase-mediated proton pump, thereby disrupting the proton gradient and resulting in yeast/fungal cell death [9]. We previously demonstrated that OTf modulates the proton pump when expressed in the yeast *Pichia* [28]. Additionally, its N-terminal derived peptide can dissipate the microbial transmembrane electrochemical potential [29], thus disturbing the proton gradient. Hence, the effective activity of OTf(SMZ)-gNP against *C. albicans* is due to the disturbance of the membrane gradient by OTf and the known ability of gNPs to interact with the H^+-ATPase-mediated proton pump [9], thereby propagating the effect of SMZ. Our results, therefore, boosted the candidacy of OTf-gNP as a potent antimicrobial agent and OTf(SMZ)-gNP as a promising antibiotic-targeting carrier for treating skin infections and yeast/fungal Candidiasis.

Apart from the above antimicrobial activities, both OTf(SMZ)-gNP and OTf-gNP demonstrated significant antibacterial effects on the drug-resistant wild-type *Salmonella enterica* carrying the *tol*C resistance gene. The TolC is an outer membrane component of the AcrAB-TolC multidrug resistance channel, involved in siderophore export, thus allowing *Salmonella* to survive in iron-restricted and antibiotic environments [21,22]. An in-depth study has shown that the *tol*C promoter is activated upon contact with transferrin [19]. The sensitivity of wild-type *Salmonella* to OTf-gNP, while Δ*tol*C *Salmonella* was resistant, suggests that OTf acts as a modulator of the AcrAB-TolC efflux pump, which disrupts the growth of wild-type *Salmonella* [30].

In contrast, the intense antimicrobial action of OTf(SMZ)-gNP against both wild-type and Δ*tol*C *Salmonella* indicates that its effect is independent of the TolC channel, likely occurring through multi-mobility diffusion into the cell membrane [8,11]. However, the mechanism remains to be further confirmed.

We have explored the efficiency of OTf-gNP as a drug carrier to facilitate the delivery of the water-insoluble antibiotic SMZ into HCT-116 cells intracellularly infected with *S. aureus* or *E. coli*. The nano-formulations OTf(SMZ)-gNP and OTf-gNP potently caused the death of both intracellular bacterial strains, whereas the bacteria were poorly sensitive to pure gNP and weakly susceptible to free OTf or SMZ. In our recent work, we demonstrated the effective binding and internalisation of OTf into HCT-116 cells through transferrin receptor (TfR)-mediated uptake [5]. The higher susceptibility of intracellular *S. aureus* to OTf-gNP compared to *E. coli* may be due to the unique presence of transferrin receptor (TfR) on the peptidoglycan surface of *S. aureus* [31,32]. The significant reduction in both intracellular bacteria observed after exposure to OTf(SMZ)-gNP and OTf-gNP compared to OTf or SMZ suggests the roles of TfR-mediated uptake and direct diffusion. This mechanism facilitated the delivery and enhanced accumulation of the antibiotic (SMZ) or the gold nanoparticles in the cytosol of infected cells.

The present study thus demonstrates a novel approach for utilising an abundant egg protein, OTf, as a drug-targeting molecule through TfR-mediated endocytosis, achieving higher efficacy and minimal toxicity, thereby intervening in intracellular bacterial infections.

Conclusions

In this study, the synthesis of sulfamethoxazole (SMZ)-loaded ovotransferrin (OTf) gold nanoparticles [OTf(SMZ)-gNP] has been established via exploiting the reducing power of citrate. The OTf(SMZ)-gNP nanoparticles were irregularly shaped, with a quasi-spherical morphology within the nano-size range. Loading SMZ onto the OTf molecule and then capping the gNP significantly enhanced its antibacterial potency against both Gram-positive and Gram-negative bacteria, as well as the fungal *C. albicans* at much lower doses than the pure individual components, SMZ, OTf or gNPs. This nano-formulation demonstrated a remarkable antibacterial efficacy against drug-resistant *Salmonella* and enabled the delivery of OTf(SMZ) through TfR-mediated endocytosis into the intracellular milieu of infected cells. These results pave the way for the development of other gNP-based antibiotic formulations to combat the rising problem of drug resistance. Although gNPs are nontoxic [26], further in vivo studies are needed to examine the bioavailability, toxicity, and kinetics of OTf(SMZ)-gNP, which are necessary to assess its clinical applicability.,

Abbreviations

OTf	Ovotransferrin
SMZ	Sulfamethoxazole
gNP	Gold nanoparticles
OTf-gNP	Ovotransferrin capped gold nanoparticles
OTf(SMZ)	Ovotransferrin complex with sulfamethoxazole
OTf(SMZ)-gNP	Gold nanoparticles capped OTf(SMZ)
SMZ-gNP	Sulfamethoxazole-loaded gNP
CFU	Colony forming unit
TSA	Trypticase soy agar
PBS	Phosphate buffered saline
WT	Wild-type *Salmonella* Enteritidis
Δ*tol*C	*tol*C-deleted mutant *Salmonella* Enteritidis

Disclosures

Author Contributions

"Conceptualization, H.R. Ibrahim; methodology, H.R. Ibrahim and T. Kubo; formal analysis, H.R. Ibrahim and T. Kubo; investigation, H.R. Ibrahim; resources, H.R. Ibrahim; data curation, H.R. Ibrahim and T. Kubo; writing—original draft preparation, H.R. Ibrahim; writing—review and editing, H.R. Ibrahim; visualization, H.R. Ibrahim; supervision, H.R. Ibrahim; project administration, H.R. Ibrahim. All authors have read and agreed to the published version of the manuscript."

Funding

This research received no external funding

Acknowledgements

The Authors would like to thank the Research Support Centre of Kagoshima University for TEM analysis and the Department of Pathology, Bacteriology and Avian Diseases, Faculty of Veterinary Medicine, Ghent University, Belgium, for the wild-type and *tol*C-deficient mutant *Salmonella Enteritidis* strains.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

1. Ovung A, Bhattacharyya J. Sulfonamide drugs: structure, antibacterial property, toxicity, and biophysical interactions. Biophys Rev. 2021;13:259–272. https://doi.org/10.1007/s12551-021-00795-9.
2. Tačić A, Nikolic V, Nikolic L, Savic I. Antimicrobial sulfonamide drugs. Adv Technol. 2017;6:58–71. https://doi.org/10.5937/savteh1701058T.
3. Jia X, Li S, Wang Y, Wang T, Hou X. Adsorption Behavior and Mechanism of Sulfonamide Antibiotics in Aqueous Solution on a Novel MIL-101(Cr)@GO Composite. J Chem Eng Data. 2019;64:1265–1274. https://doi.org/10.1021/acs.jced.8b01152.
4. Dibbern DA, Montanaro A. Allergies to sulfonamide antibiotics and sulfur-containing drugs. Ann Allergy Asthma Immunol Off Publ Am Coll Allergy Asthma Immunol. 2008;100:91–100; quiz 100–103, 111. https://doi.org/10.1016/S1081-1206(10)60415-2.
5. Ibrahim HR, Tatsumoto S, Ono H, Van Immerseel F, Raspoet R, Miyata T. A novel antibiotic-delivery system by using ovotransferrin as targeting molecule. Eur J Pharm Sci Off J Eur Fed Pharm Sci. 2015;66:59–69. https://doi.org/10.1016/j.ejps.2014.10.005. Cited: in: : PMID: 25315410.
6. Ibrahim HR, Miyawaki D, Miyata T. Ovotransferrin as a novel drug-targeting molecule for cancer chemotherapy. J Stem Cell Res Med. 2020;5:1–8. https://doi.org/10.15761/JSCRM.1000141.
7. Jana S, Sen KK, Gandhi A. Alginate Based Nanocarriers for Drug Delivery Applications. Curr Pharm Des. 2016;22:3399–3410.
8. Debets VE, Janssen LMC, Šarić A. Characterising the diffusion of biological nanoparticles on fluid and cross-linked membranes. Soft Matter. 2020;16:10628–10639. https://doi.org/10.1039/D0SM00712A.
9. Tian E-K, Wang Y, Ren R, Zheng W, Liao W. Gold Nanoparticle: Recent Progress on Its Antibacterial Applications and Mechanisms. J Nanomater. 2021;2021:2501345. https://doi.org/10.1155/2021/2501345.
10. Nakamura H, Watano S. Direct Permeation of Nanoparticles across Cell Membrane: A Review. KONA Powder Part J. 2018;35:49–65. https://doi.org/10.14356/kona.2018011.
11. Hsieh C-L, Spindler S, Ehrig J, Sandoghdar V. Tracking Single Particles on Supported Lipid Membranes: Multimobility Diffusion and Nanoscopic Confinement. J Phys Chem B. 2014;118:1545–1554. https://doi.org/10.1021/jp412203t.
12. Methner U, al-Shabibi S, Meyer H. Experimental oral infection of specific pathogen-free laying hens and cocks with Salmonella enteritidis strains. Zentralblatt Vet Reihe B J Vet Med Ser B. 1995;42:459–469. https://doi.org/10.1111/j.1439-0450.1995.tb00737.x. Cited: in: : PMID: 8578920.
13. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci. 2000;97:6640–6645. https://doi.org/10.1073/pnas.120163297.
14. Mezghich S. Photochemical Degradation of the Antimicrobial Sulfamethoxazole upon Solar Light Excitation: Kinetics and Elucidation of Byproducts Using LC/ ESI+/MS2 Technique. Mass Spectrom Purif Tech [Internet]. 2016;
15. Saptarshi SR, Duschl A, Lopata AL. Interaction of nanoparticles with proteins: relation to bio-reactivity of the nanoparticle. J Nanobiotechnology. 2013;11:26. https://doi.org/10.1186/1477-3155-11-26.
16. Mekuye B. The Impact of Size on the Optical Properties of Silver Nanoparticles Based on Dielectric Function. In: Ameen S, Akhtar MS, Jiménez-Suárez A, Seisdedos G, editors. Nanotechnol Nanomater Annu Vol 2024 [Internet]. Rijeka: IntechOpen; 2023 [cited 2025 Feb 13]. Available from: https://doi.org/10.5772/intechopen.113976.
17. Sørensen LK, Khrennikov DE, Gerasimov VS, Ershov AE, Polyutov SP, Karpov SV, Ågren H. Nature of the Anomalous Size Dependence of Resonance Red Shifts in Ultrafine Plasmonic Nanoparticles. J Phys Chem C. 2022;126:16804–16814. https://doi.org/10.1021/acs.jpcc.2c03738.
18. Hoshyar N, Gray S, Han H, Bao G. The effect of nanoparticle size on in vivo pharmacokinetics and cellular interaction. Nanomed. 2016;11:673–692. https://doi.org/10.2217/nnm.16.5.
19. Horiyama T, Yamaguchi A, Nishino K. TolC dependency of multidrug efflux systems in Salmonella enterica serovar Typhimurium. J Antimicrob Chemother. 2010;65:1372–1376. https://doi.org/10.1093/jac/dkq160. Cited: in: : PMID: 20495209.
20. Yousefian N, Ornik-Cha A, Poussard S, Decossas M, Berbon M, Daury L, Taveau J-C, Dupuy J-W, Đorđević-Marquardt S, Lambert O, et al. Structural characterization of the EmrAB-TolC efflux complex from E. coli. Biochim Biophys Acta BBA - Biomembr. 2021;1863:183488. https://doi.org/10.1016/j.bbamem.2020.183488.
21. Li K, Chen W-H, Bruner SD. Microbial siderophore-based iron assimilation and therapeutic applications. Biometals Int J Role Met Ions Biol Biochem Med. 2016;29:377–388. https://doi.org/10.1007/s10534-016-9935-3. Cited: in: : PMID: 27146331.
22. Ratledge C, Dover LG. Iron metabolism in pathogenic bacteria. Annu Rev Microbiol. 2000;54:881–941. https://doi.org/10.1146/annurev.micro.54.1.881. Cited: in: : PMID: 11018148.

23. Zgurskaya HI, Krishnamoorthy G, Ntreh A, Lu S. Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. Front Microbiol. 2011;2:189. https://doi.org/10.3389/fmicb.2011.00189. Cited: in: : PMID: 21954395.
24. Wang Z, Fan G, Hryc CF, Blaza JN, Serysheva II, Schmid MF, Chiu W, Luisi BF, Du D. An allosteric transport mechanism for the AcrAB-TolC multidrug efflux pump. eLife. 2017;6:e24905. https://doi.org/10.7554/eLife.24905. Cited: in: : PMID: 28355133.
25. Xu Y, Li H, Li X, Liu W. What happens when nanoparticles encounter bacterial antibiotic resistance? Sci Total Environ. 2023;876:162856. https://doi.org/10.1016/j.scitotenv.2023.162856.
26. Enea M, Peixoto de Almeida M, Eaton P, Dias da Silva D, Pereira E, Soares ME, Bastos M de L, Carmo H. A multiparametric study of gold nanoparticles cytotoxicity, internalization and permeability using an in vitro model of blood-brain barrier. Influence of size, shape and capping agent. Nanotoxicology. 2019;13:990–1004. https://doi.org/10.1080/17435390.2019.1621398.
27. Mishra M, Kumar S, Majhi RK, Goswami L, Goswami C, Mohapatra H. Antibacterial Efficacy of Polysaccharide Capped Silver Nanoparticles Is Not Compromised by AcrAB-TolC Efflux Pump. Front Microbiol. 2018;9:823. https://doi.org/10.3389/fmicb.2018.00823. Cited: in: : PMID: 29780364.
28. Ibrahim HR, Hozono A, Fukami M, Shaban MA, Miyata T. Expression of Ovotransferrin Enhances Tolerance of Yeast Cells toward Oxidative Stress. J Agric Food Chem. 2013;61:6358–6365. https://doi.org/10.1021/jf401152e.
29. Ibrahim HR, Sugimoto Y, Aoki T. Ovotransferrin antimicrobial peptide (OTAP-92) kills bacteria through a membrane damage mechanism. Biochim Biophys Acta. 2000;1523:196–205. https://doi.org/10.1016/s0304-4165(00)00122-7. Cited: in: : PMID: 11042384.
30. Allgood Samual C., Su Chih-Chia, Crooks Amy L., Meyer Christian T., Zhou Bojun, Betterton Meredith D., Barbachyn Michael R., Yu Edward W., Detweiler Corrella S. Bacterial efflux pump modulators prevent bacterial growth in macrophages and under broth conditions that mimic the host environment. mBio. 2023;14:e02492-23. https://doi.org/10.1128/mbio.02492-23.
31. Modun B, Kendall D, Williams P. Staphylococci express a receptor for human transferrin: identification of a 42-kilodalton cell wall transferrin-binding protein. Infect Immun. 1994;62:3850–3858. https://doi.org/10.1128/iai.62.9.3850-3858.1994. Cited: in: : PMID: 8063401.
32. van Dijk MC, de Kruijff RM, Hagedoorn P-L. The Role of Iron in Staphylococcus aureus Infection and Human Disease: A Metal Tug of War at the Host—Microbe Interface. Front Cell Dev Biol [Internet]. 2022;10.

ORIGINAL PAPER

JMS Journal of Medical Science

Validation and psychometric properties of the Consumer Financial Protection Bureau Financial Well-Being Scale (CFPB-FWBS) for Polish family caregivers of individuals with rare diseases

Dariusz Walkowiak
Department of Organisation and Management in Health Care, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0001-8874-2401
Corresponding author: dariuszwalkowiak@ump.edu.pl

Piotr Jabkowski
Faculty of Sociology, Adam Mickiewicz University, Poznań, Poland
https://orcid.org/0000-0002-8650-9558

Jan Domaradzki
Department of Social Sciences and Humanities, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0002-9710-832X

https://doi.org/10.20883/medical.e1428

Keywords: CFPB-FWBS, family caregivers, financial burden, financial well-being, rare diseases

Received 2025-09-20
Accepted 2025-11-13
Published 2025-12-30

How to Cite: Walkowiak D, Jabkowski P, Domaradzki J. Validation and psychometric properties of the Consumer Financial Protection Bureau Financial Well-Being Scale (CFPB-FWBS) for Polish family caregivers of individuals with rare diseases. Journal of Medical Science. 2025 December;94(4);e1428. doi:10.20883/medical.e1428

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Introduction. Caregivers of individuals with rare diseases (RDs) face numerous challenges related to healthcare access, physical and emotional strain, social isolation, and psychological distress; however, financial burden often has the most significant impact on family well-being and the ability to provide adequate care. This study aimed to validate the Consumer Financial Protection Bureau Financial Well-Being Scale (CFPB-FWBS) among caregivers of individuals with RD in Poland.

Material and methods. Based on a sample of 942 family caregivers of individuals diagnosed with one of 159 RDs, the validation procedure involved exploratory and confirmatory factor analyses, along with evaluation of internal consistency and interpretability.

Results. Analyses supported the unidimensional structure of the Polish CFPB-FWBS. Inter-item correlations were moderate to strong (r = .48–.68), except for Q9, which showed a weaker correlation. PCA confirmed a dominant first component (eigenvalue ≈ 5.7, explaining ~57% of variance), with all items loading adequately. Cronbach's alpha was high (α = 0.92), and no item removal improved reliability. CFA indicated good model fit (χ^2/df = 5.67, CFI = 0.968, TLI = 0.959, SRMR = 0.03, RMSEA = 0.07). Latent scores (0–100 scale) approximated a normal distribution (M = 50.5, SD = 12.1, range = 14–86).

Conclusions. The Polish version of the CFPB-FWBS demonstrates strong reliability, structural validity, and meaningful score distributions among caregivers of individuals with rare diseases. These findings support its use as a standardised measure of financial well-being in Poland, enabling research, policy development, and international comparisons.

Introduction

Although no universal definition of a rare disease (RD) exists, the European Union (EU) defines RDs as conditions affecting fewer than 1 in 2,000 individuals [1,2]. However, some diseases occur even less frequently: ultra-rare diseases are those found in fewer than 1 in 100,000 people, while hyper-rare diseases affect fewer than 1 in 1,000,000 [3,4]. Current estimates indicate that more than 10,000 distinct RDs have been identified [5], impacting up to 35 million people in the EU, over 350 million worldwide, and approximately 2.5-3 million individuals in Poland [6].

Despite wide variation in aetiology, clinical manifestations, course, and prognosis, RDs share several characteristics. Approximately 80% of these conditions are of genetic origin, with about 65% resulting in severe clinical symptoms. Children account for half of all cases, and nearly one--third die before reaching the age of five [7]. Furthermore, 95% of RDs still lack approved therapy [8–10]).

Challenges related to RDs extend beyond patients, placing substantial emotional, physical, financial, and organisational demands on families and caregivers, who often require continuous support and guidance in navigating fragmented healthcare and social systems [11–14]. Families frequently face prolonged diagnostic delays, limited referral pathways, and a lack of coordinated care, often having to manage services independently [14–16]. Access to genetic testing, counselling, and innovative therapies is commonly restricted, while institutional and psychological support remains insufficient, forcing reliance on out-of-pocket resources [11,17]. A common barrier is low RD awareness among the public, policymakers, and healthcare professionals , resulting in most primary care physicians reporting a lack of expertise and feeling unprepared to care for patients with RD. This is unsurprising, as such cases represent only about 1.6% of visits [22,23].

These systemic barriers translate into significant burdens for caregivers, encompassing physical strain (e.g., fatigue, injury), emotional distress, social isolation, and psychological exhaustion, which reduce quality of life and increase the risk of depression, anxiety, and caregiver burden. Consequently, many caregivers require psychiatric support. Financial strain is also considerable, involving frequent medical visits, high out-of-pocket costs, and disparities in insurance coverage, which force many parents to reduce their working hours or leave employment entirely [24–26]. Research consistently shows that RDs generate both direct (e.g., therapies, equipment, home modifications) and indirect (notably, caregiver productivity loss) costs, further compounded by the adverse effects of caregiving on physical and mental health, which undermines both family well-being and economic stability [27,28].

Although earlier research has shown that financial well-being is an essential predictor of caregivers' quality of life and perceived burden, no dedicated instrument exists to evaluate their economic strain. This study addresses this gap by validating the Consumer Financial Protection Bureau Financial Well-Being Scale (CFPB-FWBS) [29] among Polish caregivers of individuals with RDs and by examining its psychometric properties, including reliability, validity, and factor structure in this specific population.

Methods

Study Design

A cross-sectional study was conducted in 2024, in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [30]. Data were collected through a self-administered, anonymous, computer-assisted web-based survey assessing the association between caregiving for a person with an RD and caregivers' financial well-being.

Participants and Setting

Participants were Polish-speaking adults (≥18 years) who were parents or family caregivers of individuals with a confirmed RD diagnosis, had internet access, and provided informed consent. Due to the absence of a national RD registry, caregivers were recruited via convenience sampling, with patient associations, foundations, and organisations distributing the survey link through their websites and social media.

Table 1 shows the composition of the caregiver sample according to gender (male/female) and age group (16–29, 30–39, 40–49, 50+), along

Table 1. Sample description by gender and age of caregivers.

Age (years)	Male	Female	Total
16-29	2 (1.5%)	32 (3.9%)	34 (3.6%)
30-39	44 (33.8%)	337 (41.5%)	381 (40.4%)
40-49	64 (49.2%)	350 (43.1%)	414 (43.9%)
50 and more	20 (15.4%)	93 (11.5%)	113 (12.0%)
	130 (13.8%)	**812 (86.2%)**	**942 (100%)**

with totals for the entire sample. The table provides counts and percentages, directly comparing the age structure within each gender group and against the overall distribution.

The analytical sample comprised 942 caregivers, of whom 812 were women (86.2%), and 130 were men (13.8%). Age distribution was: 16–29, 34 (3.6%; men 2 [1.5%], women 32 [3.9%]); 30–39, 381 (40.4%; men 44 [33.8%], women 337 [41.5%]); 40–49, 414 (43.9%; men 64 [49.2%], women 350 [43.1%]); 50+, 113 (12.0%; men 20 [15.4%], women 93 [11.5%]). Thus, participation was concentrated among those aged 30–49 (795/942; 84.3%), with relatively few in the 16–29 age group (3.6%) and among those aged 50 and over (12.0%). Men were more likely to be in the 40–49 age group (49.2%) than women (43.1%), whereas women were overrepresented in the 30–39 age group (41.5% vs. 33.8% for men). At the same time, although the relatively low number of male participants may suggest a potential gender bias, it should be noted that convenience sampling typically reflects the broader pattern of a marked underrepresentation of men among family caregivers [14,31,32].

Ethical Issues

The study was conducted in accordance with the principles outlined in the Declaration of Helsinki. It was approved by the Bioethics Committee of the Poznan University of Medical Sciences (KB–228/24, March 13, 2024). Informed written consent was obtained electronically via the online survey ("I agree" option).

Participation was voluntary, anonymous, and confidential, with the right to withdraw at any time. No personal identifiers were collected, and although some questions could be emotionally sensitive, respondents could skip items or discontinue participation. No financial compensation was provided.

Research Questionnaire

The survey questionnaire used in this study consisted of two sections. The first included closed-ended, single-choice questions concerning caregivers' sociodemographic characteristics. The second incorporated the CFPB-FWBS [29], a freely available, standardised instrument developed in the United States to measure financial well-being, understood as both economic security and the freedom to manage short- and long-term financial needs. The scale comprises ten Likert-type items, producing raw and standardised scores that enable comparisons across populations and allow for examining associations with other variables:

How well does this statement describe you or your situation?

Q1. I could handle a major unexpected expense
Q2. I am securing my financial future
Q3. Because of my financial situation, I feel like I will never have the things I want in life
Q4. I can enjoy life because of the way I'm managing my money
Q5. I am just getting by financially
Q6. I am concerned that the money I have or will save won't last

How often does this statement apply to you?

Q7. Giving a gift for a wedding, birthday, or other occasion would put a strain on my finances for the month
Q8. I have money left over at the end of the month
Q9. I am behind with my finances
Q10. My finances control my life

Since no Polish version of the CFPB-FWBS was available, the tool was back-translated by two independent bilingual translators and adapted to the Polish context. It was then pilot-tested in two small groups (caregivers of children with

chronic illnesses and healthy adults), confirming its clarity, cultural relevance, and internal consistency, which supported its suitability for use in the present study.

Data Collection

The data were collected between March and August 2024 among family caregivers of individuals with RDs, with the assistance of several patient organisations, foundations, and associations (see Acknowledgements). After obtaining their permission, the research coordinator distributed an invitation letter with a link to the online questionnaire via their websites and social media. In total, 73 patient groups were contacted, many of which represented multiple rare conditions or the broader RD community. Participants provided informed consent electronically before completing the survey, which took approximately 20–25 minutes. To increase response rates, three reminders were sent during the study period.

Statistical Analysis

Descriptive statistics were first computed for all items. Inter-item correlations were examined to assess initial patterns of association. Principal Component Analysis (PCA) was conducted to evaluate unidimensionality, with eigenvalues, scree plot, and component loadings used to guide interpretation. Reliability was assessed using Cronbach's alpha with 95% confidence intervals (CI) and item-deletion diagnostics. Confirmatory Factor Analysis (CFA) was then performed using polychoric correlations and a robust weighted least squares estimator, with factor variance fixed to 1 for identification. Model fit was evaluated with multiple indices, including the chi-square test, the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) with 90% CI, and the Standardised Root Mean Square Residual (SRMR). Standard thresholds for acceptable fit (CFI/TLI ≥ .95, RMSEA ≤ .06–.08, SRMR ≤ .08) were applied. Normative data were generated for the total scale scores, stratified by gender. All statistical analyses were conducted using R Statistical Software (version 4.3.1; R Foundation for Statistical Computing) [34].

Results

Figure 1 shows the matrix of correlations between the ten CFPB-FWBS items. After aligning the item

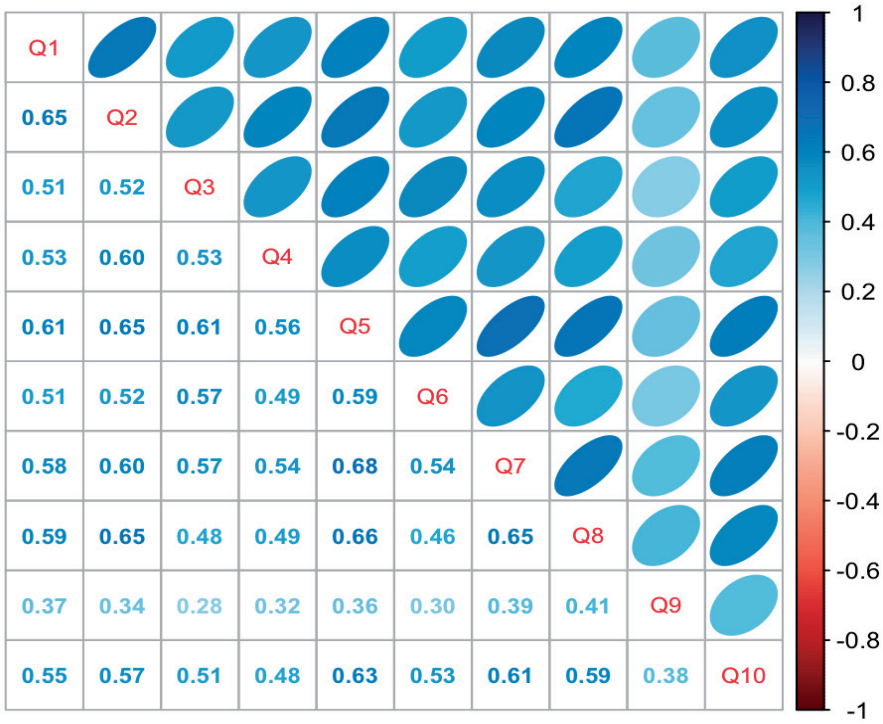


Figure 1. Correlation plot for 10 items of the CFPB-FWBS.

directions, all correlations are positive. Most of the associations between pairs of items fall within the moderate to strong range (approximately r = .48–.68), indicating that the items share a substantial amount of common variance. The strongest associations are observed among items measuring perceived security and financial management (Q1–Q8 and Q10), consistent with a largely unidimensional structure.

A notable exception is Item Q9 ('I am behind with my finances', reverse-coded), which shows markedly weaker correlations with the rest of the scale (approximately r = .28–.41). This attenuation is evident, suggesting that Q9 may capture a more specific aspect of financial strain relating to arrears rather than the broader sense of economic security and control targeted by the remaining items. Consequently, we anticipate that Q9 will contribute less to internal consistency and exhibit a lower factor loading in subsequent analyses, a point we examine formally below.

Principal Component Analysis

Next, we implemented principal component analysis (PCA) to examine the scale's unidimensionality before confirmatory modelling. PCA provides a model-free summary of how the ten items covary by reducing them to components extracted from the inter-item correlation matrix. This step addresses two central questions regarding score validity: (a) whether a single dominant component accounts for a substantial proportion of the variance, which would support a unidimensional total score, and (b) whether any items contribute weakly or inconsistently to the aggregated score.

The scree plot in **Figure 2** shows a dominant first component, followed by a sharp drop and a long, shallow tail. The first eigenvalue is approximately 5.7–5.8, accounting for around 57–58% of the total variance. The second eigenvalue is approximately 0.8, and all subsequent eigenvalues are less than or equal to 0.6. According to the Kaiser criterion (retaining eigenvalues > 1) and the apparent elbow after the first point, there is strong evidence in favour of a unidimensional structure. However, given that the second eigenvalue is <1 and the curve flattens immediately, any additional component would capture only minor, item-specific variance rather than a coherent secondary dimension.

Table 2 reports the PCA loadings and the item complexity. A clear pattern emerges: Items Q1-Q8 and Q10 load moderately to firmly on Compo-

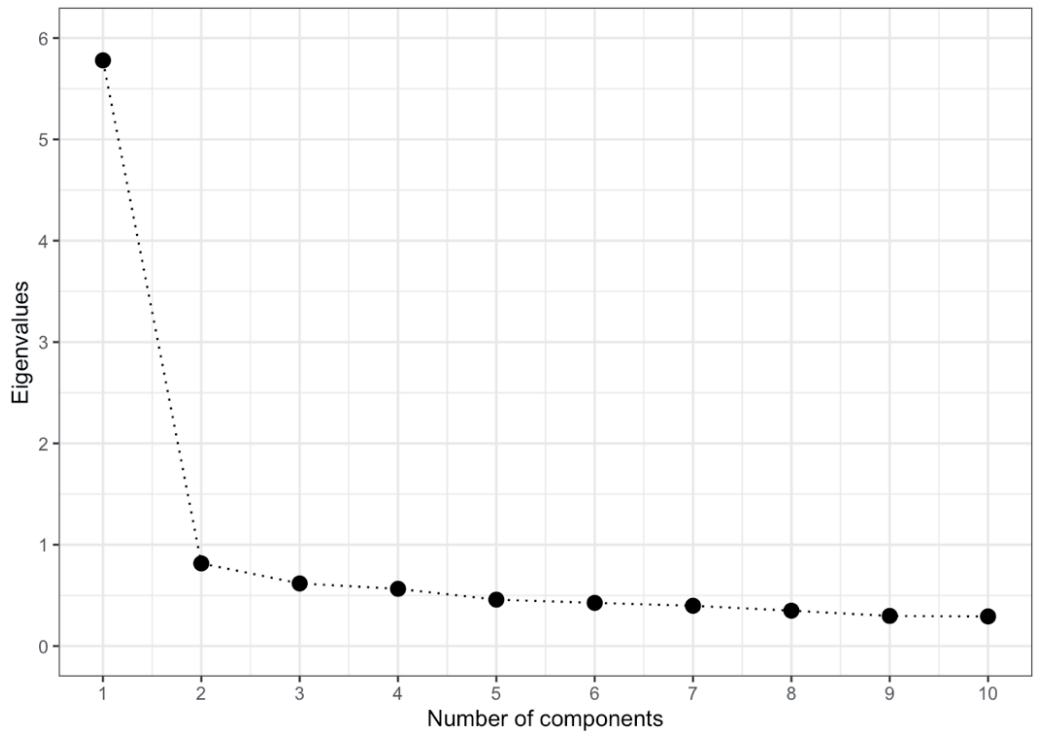


Figure 2. Scree Plot.

Table 2. PCA results for two components: number of items reduced based on factor loadings values, and complexity measure.

Item	Component 1
Q1	0.784
Q2	0.813
Q3	0.739
Q4	0.732
Q5	0.848
Q6	0.728
Q7	0.819
Q8	0.795
Q9	0.520
Q10	0.775

Table 3. Fit indices for the unidimensional solution of the CFPB-FWBS.

Chisq	df	Chisq/df	p-value	CFI	TLI	SRMR	RMSEA	RMSEA – 90%CI	RMSEA + 90%CI
198.441	35	5.67	<0.001	0.968	0.959	0.03	0.07	0.061	0.08

nent 1 (approximately 0.65–0.80; maximum Q5 loading of 0.804). Q9 shows the lowest loading on Component 1 (0.520); however, its primary loading remains adequate.

According to typical item-reduction rules (items with salient loading on the dominant component are retained, and items with high complexity are avoided), only Q9 could be recognised as flagged for exclusion from a unidimensional scale. No other item meets the removal criteria. However, in practice, retaining Q9 is helpful for comparability or for capturing strain specific to arrears; hence, we proceed further with all ten items.

Reliability analysis

The Cronbach's alpha for all ten items was 0.92 (standardised α = 0.92), with a 95% confidence interval (CI) of 0.91–0.93 and an average inter-item correlation of 0.57. This indicates high internal consistency for the retained items. Item-deletion diagnostics revealed that removing no item would meaningfully improve reliability: alphas if dropped ranged from 0.90 to 0.92 (the lowest being when Q5 was dropped, with an alpha of approximately 0.90). The reliability evidence supports the creation of a unidimensional scale based on questions 1-10; removing any single item would not yield a higher reliability score, and all retained items demonstrate adequate discrimination.

Confirmatory factor analysis

Guided by the PCA and reliability results, we tested a single-factor CFA model for the CFPB scale using all items (see **Table 3**). As the items are ordinal, we estimated the model using polychoric correlations with a robust estimator (factor variance fixed to 1 for identification). We evaluated the model's fit using the following criteria: χ^2/df, CFI, TLI, RMSEA (90% CI), and SRMR, and the standard thresholds: CFI/TLI ≥ .95, RMSEA ≤ .06–.08, and SRMR ≤ .08.

The unidimensional model demonstrated satisfactory overall fit. Although the chi-squared test was significant (i.e., χ^2 = 198.44, df = 35, p < .001; chi-squared to degrees of freedom ratio = 5.67), this is to be expected with N = 942. Practical indices were strong: the CFI and TLI exceeded conventional thresholds at 0.968 and 0.959, respectively; the SRMR indicated perfect residual fit at 0.03; and the RMSEA fell within the acceptable range at 0.07, with 90% CI [0.061, 0.08]. Together, these results support a single-factor representation of the ten items.

Latent scores of the CFPB-FWBS for the Polish population of caregivers of persons with RDs

A descriptive analysis of the normalised CFPB-FWBS scores was conducted. Responses to each item were coded on a 0–4 scale, and the total score was derived in accordance with the scoring

instructions provided by the scale's authors (30). The recommended interpretive bands are as follows: Very low (0–29), Low (30–37), Medium low (38–49), Medium high (50–57), High (58–67), and Very high (68–100). **Figure 3** displays the distribution of CFPB-FWBS latent scores, ranging from 0 to 100, calculated as a CFA-weighted sum of all ten items.

Figure 3 shows a roughly symmetric, unimodal distribution centred near 50 using the CFPB's scoring algorithm (0–100). **Table 4** indicates a mean of 50.54 (SD 12.08) and a median of 50.00, with a skewness of −0.001 and a kurtosis of 2.892. These values are consistent with a normal distribution. The Shapiro-Wilk test (W = 0.996, p = 0.010) rejects the assumption of normality, likely due to its sensitivity to large samples. Yet, the deviation is slight, and the histogram remains close to bell-shaped.

Empirical scores in a sample range from 14 to 86, with quartiles at Q1 = 42.0 and Q3 = 59.0 (IQR = 17). Thus, the central 50% of caregivers fall within the Medium Low (38–49), Medium High (50–57), and High (58–67) bands. The mode around 50–55 suggests that many respondents are in the Medium High category. Visual inspection

Table 4. Descriptive characteristics of latent scores for the CFPB-FWBS for the Polish population of caregivers of persons with RDs

Statistic	CFPB-FWBS
Mean	50.54
Standard deviation	12.08
Skewness	-0.001
Kurtosis	2.892
Shapiro-Wilk statistic	0.996
Shapiro-Wilk p-value	0.010
Minimum	14.0
Q1	42.0
Median	50.0
Q3	59.0
Maximum	86.0

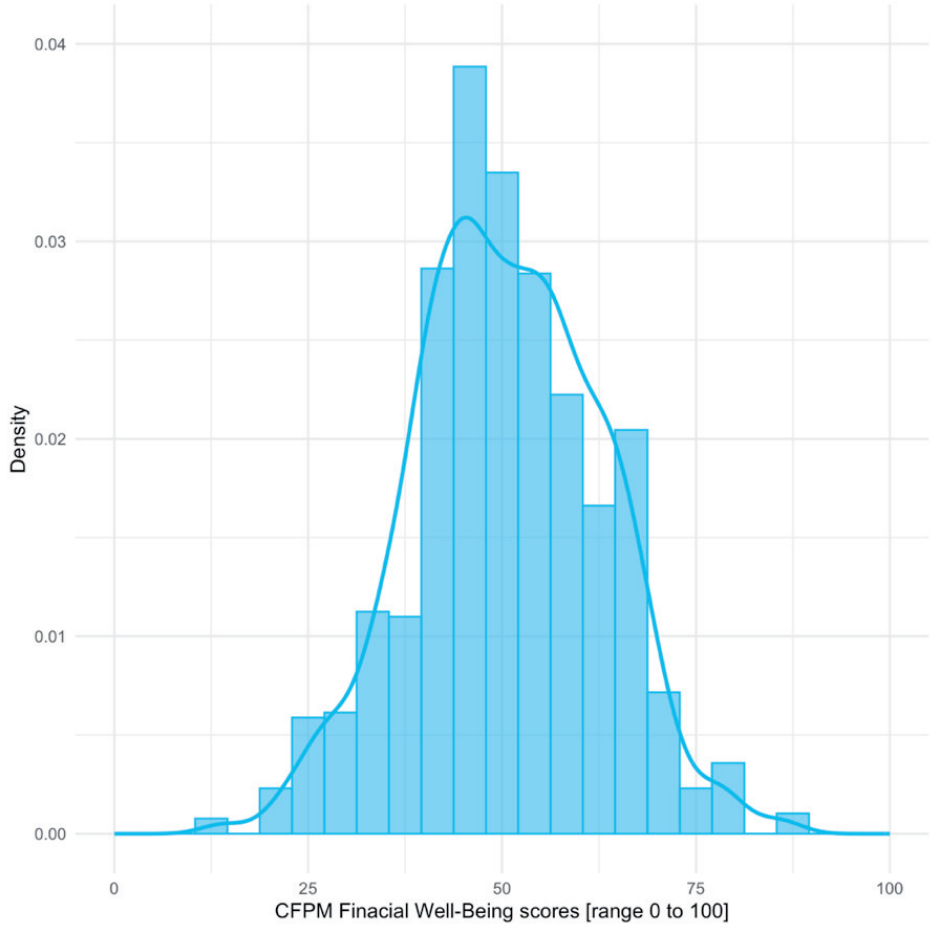


Figure 3. Histogram and density plot for the CFPB-FWBS for the Polish population of caregivers of persons with RDs

reveals thin tails, indicating relatively few cases in the Very Low (0–29) or Very High (68–100) ranges, with no evidence of floor or ceiling compression.

Overall, the score distribution is evenly spread across the interpretive bands, with most caregivers clustering in the medium to high financial well-being range—an empirical profile that supports subsequent subgroup comparisons and norm-referenced interpretation.

Normative data for the CFPB-FWBS

Normative data for the CFPB-FWBS were established using a large sample of Polish caregivers of individuals with rare diseases. The norms are stratified by gender, enabling researchers to interpret an individual's score in relation to population-specific reference values (see **Table 5**). We provide the following information to gauge each subgroup's distribution and potential score compression: means, standard errors, percentile--based summaries (Q1, median, Q3), minima, and maxima. These norms can be used for several purposes, such as benchmarking individuals or cohorts, identifying demographic segments with relatively lower financial well-being, and informing eligibility thresholds in screening or programme evaluation. As subgroup sizes vary, interpretation should consider standard errors and the width of interquartile ranges. Taken together, the norms enhance the practical utility of the scale by providing a clear demographic-specific context for CFPB-FWBS scores in this population.

Table 5 provides gender-stratified reference values. On average, men have a higher central tendency than women (mean = 53.8, standard error (SE) = 1.10 vs. mean = 50.0, SE = 0.42), which equates to a difference of ~3.8 points on the 0–100 scale (a small effect given the overall standard deviation (SD) of approximately 12). The medians mirror this pattern (55 vs. 50). However, quartiles indicate substantial overlap: men Q1–Q3 = 46–63 (IQR = 17) and women Q1–Q3 = 42–58 (IQR = 16). Thus, while men score higher on average, the distributions are similar.

Relative to the CFPB interpretive bands, both medians fall within the 'Medium High' category (50–57). For men, Q3 = 63 falls into the High category (58–67), whereas for women, Q3 = 58 sits at the High threshold. Q1 places more women in the Medium Low category (38–49) than men (42 vs. 46). The observed ranges (14–86 in both groups) suggest there is no compression of the floor or ceiling. Standard errors are smaller among women, reflecting the larger subgroup. Overall, these norms indicate that men have modestly higher financial well-being, but substantial overlap between the genders supports the use of a standard cut-point system for practical interpretation.

Table 5. CFPB-FWBS scale norms by gender for the Polish population of caregivers of persons with RDs

Statistic	Male	Female
Mean	53.8	50.0
Standard error of the mean	1.10	0.42
Minimum	14.0	14.0
Q1	46.0	42.0
Median	55.0	50.0
Q3	63.0	58.0
Maximum	86.0	86.0

Discussion

Caring for a person with RD results in a significant economic burden, encompassing both direct medical costs and indirect productivity losses [24–28]. In the United States, 379 RDs generated an annual cost of $997 billion, of which $449 billion (45%) were direct medical costs, primarily hospitalisations (32%). Prescription drugs (18%). In comparison, $437 billion (44%) stemmed from productivity losses such as absenteeism, presenteeism, and early retirement [34]. Another U.S. study of 24 RDs reported average per-patient costs of $266,000, ten times higher than for common conditions, with indirect costs making up to 45% of the total [26]. Similar findings were reported in China, where annual direct medical costs exceeded household income, with an additional 40–45% of income devoted to non-medical and indirect expenses, posing a high or extremely high burden on over half of families [35].

In Poland, studies on tuberous sclerosis complex showed that indirect costs account for 17-39% of the total burden [36]. Similarly, for adult patients with cystic fibrosis, the average annual treatment cost was €19,581, with 70% related to direct costs, mainly pharmacotherapy (€10,171) and hospitalisations (€2,878), and 30% to indire-

ct expenses, primarily lost productivity (€5,706) [37]. A recent study on parents of individuals with Angelman syndrome reported that while one--third of caregivers utilised psychological or psychiatric services, 96.2% of their costs (€31,356.79 in 2024) were privately funded [27].

European studies further confirm the high costs of RDs. Direct medical expenses for Dravet syndrome average €16,000 annually, with hospitalisations and related care accounting for the majority of expenditures [38]. In Germany, annual costs of Duchenne muscular dystrophy were €78,913 and of Becker muscular dystrophy €39,060, with informal care, productivity losses, and rehabilitation as main expenses, all rising with disease progression [39]. Numerous other studies indicate that indirect costs, lost income, reduced working hours, and job resignations remain the most significant burden [40–42]. All these findings demonstrate that the economic consequences of RDs extend far beyond healthcare expenditures, placing a profound strain on families.

This study presents the first psychometric evaluation of the Polish adaptation of the CFPB--FWBS among family caregivers of people with RDs [29]. The results indicate that the instrument functions reliably and demonstrates solid structural validity in this distinct cultural and caregiving setting.

The CFPB-FWBS was initially designed in the United States as a standardised, publicly accessible measure of financial well-being. It captures not only objective aspects of household finances but also subjective perceptions of stability and security. The Consumer Financial Protection Bureau conceptualises financial well-being as a state in which individuals can meet their present and ongoing financial commitments, feel confident about their future, and maintain the freedom to make choices that bring satisfaction in daily life. This definition was developed through extensive empirical research and qualitative input from U.S. consumers and financial experts. The instrument operationalises four interconnected domains: management of everyday financial demands, resilience to unexpected shocks, progress toward longer-term financial goals, and the freedom to make discretionary choices that improve quality of life [29].

Our analyses were consistent with the original CFPB report (CFPB, 2017), which documented a unidimensional factor structure. Principal component analysis revealed a dominant underlying factor, while confirmatory factor analysis provided strong support for a single-factor solution. Fit indices were uniformly favourable (CFI = 0.968, TLI = 0.959, SRMR = 0.03, RMSEA = 0.07), indicating that the scale performs well in this population. Although the chi-square statistic reached significance, this is expected given the large sample size and does not compromise the interpretation of the other fit indices.

Reliability was also satisfactory. Cronbach's alpha exceeded standard benchmarks, and all items loaded meaningfully onto the latent construct. These findings parallel results from other adaptations. For instance, the Brazilian Portuguese version reported by Howat-Rodrigues et al. also confirmed a unidimensional structure [43]. The consistency of results across cultural contexts underscores the instrument's robustness.

At the respondent level, scores rescaled to a 0–100 metric displayed a meaningful distribution, reflecting variation in caregivers' financial circumstances. This heterogeneity likely reflects diverse caregiving demands, household resources, and differential access to social and health system support. Caregivers of individuals with RDs often face particularly acute challenges, including high direct medical costs, uncovered out-of-pocket expenses, and lost income due to reduced labour market participation. The CFPB--FWBS's ability to capture such gradations of perceived financial well-being underscores its value for both policy and research applications.

Although initially designed for consumer-focused use in the United States, the CFPB-FWBS has since been applied across a wide array of settings, both domestically and internationally, further supporting its versatility. Research with individuals living with diabetes has shown consistently lower financial well-being compared with those without the condition, with marked disparities across racial and ethnic groups [44]. Among health professionals and trainees, the scale has been sensitive to external crises, such as Lebanon's financial collapse, during which scores declined sharply, and to structural factors, such as socioeconomic background, income, and student debt [45]. In U.S. family medicine residents, scores have typically fallen in the mid-range, with progression in training and financial education

linked to improvements [46]. Beyond these contexts, studies among trauma survivors in China revealed widespread economic insecurity, while research with older cancer patients in India found that most participants experienced poor financial well-being [47]. During the COVID-19 pandemic, the instrument was also used to assess otorhinolaryngologists in India, detecting sharp income-related declines, with higher scores observed in older and more experienced physicians [48].

Importantly, the CFPB-FWBS not only measures financial standing but also helps illuminate the broader impact of financial strain. Lower scores have been tied to psychological distress, depressive symptoms, and post-traumatic stress among trauma survivors [49]. Among patients with cancer, poorer financial well-being correlated with higher distress, greater caregiver burden, and worse mental health outcomes. More broadly, evidence shows that over-indebtedness is associated with higher rates of depression and anxiety, echoing social causation theories that link economic hardship with mental health difficulties. Lower financial well-being has also been consistently associated with reduced quality of life across multiple domains, as well as with socioeconomic vulnerabilities such as low income, debt, lack of insurance, illiteracy, and cognitive impairment.

The validation of the CFPB-FWBS in a Polish caregiving population, therefore, contributes to the international literature on financial well-being. In addition to its methodological value, the instrument provides a practical means for assessing both financial strain and resilience in health economics, social policy, and psychosocial research. Conceptually, the findings resonate with frameworks that integrate subjective perceptions and objective resources in defining financial well-being [50]. Future research should extend this work by assessing stability over time, evaluating predictive validity with health or quality-of-life outcomes, and examining measurement invariance across subgroups such as gender, employment status, and rural versus urban residence.

Several limitations should be acknowledged. The study drew on a large but non-random sample, which restricts generalizability. Its cross-sectional design prevents conclusions about sensitivity to change or causal relationships. Finally, although a unidimensional structure was supported, multidimensional aspects of financial well-being, such as short-term versus long-term security, could be relevant in specific subgroups and warrant further exploration. As only 130 fathers completed the survey, the study is constrained by an implicit gender bias. Furthermore, reliance on online recruitment may have introduced selection bias.

Despite these caveats, the present validation strengthens confidence in the CFPB-FWBS as a reliable and valid instrument for assessing financial well-being in Poland. Its application among caregivers of individuals with RDs addresses a crucial methodological gap while also shedding light on the economic challenges faced by this vulnerable group. By enabling standardised assessment, the CFPB-FWBS provides a foundation for targeted interventions, policy initiatives, and international comparisons to alleviate financial vulnerability among caregiving families.

Abbreviations

CFPB-FWBS: the Consumer Financial Protection Bureau Financial Well-Being Scale; EU: the European Union; RDs: rare diseases.

Acknowledgements

We wish to thank all the parents and caregivers who volunteered and participated in the study by sharing their time and experiences. We are also indebted to all patient associations, foundations, and organisations for their help in recruiting families: achromatopsia.pl, Alkaptonuria, BMD Dystrofia Mięśniowa Beckera, Blackfan Diamond Anemia Polska, Choroba Gauchera, Chorzy na rdzeniowy zanik mięśni (SMA), Cri du Chat Polska, Delecja 18q. Zespół de Grouchy'ego, DRAVET.PL – Grupa wspierająca rodziców i opiekunów, Dystrofia Duchenne'a w Małopolsce, FAST Poland – Foundation for Angelman Syndrome Therapeutics; Forum Sarkoidoza – Poland, FOXG1 syndrome Poland, Fundacja MATIO, Fundacja Parent Project Muscular Dystrophy, Fundacja Pomocy Chorym na Zanik Mięśni, Fundacja Pomocy Chorym na Zanik Mięśni im. Piotra Karlińskiego, Fundacja Salemander, Fundacja Saventic – choroby rzadkie, Fundacja SMA, glikogenoza, Grupa wsparcia dla chorych na stwardnienie rozsiane i ich bliskich, Grupa wsparcia dla osób z Chorobą Fabry'ego, Hemofilia Polska – Polskie

Stowarzyszenie Chorych na Hemofilię, Jesteśmy Pod Ścianą Foundation, Mowat-Wilson syndrome Polska, Mukowiscydoza, Najrzadsze i rzadkie choroby genetyczne świata – Grupa Wsparcia, Ogólnopolskie Stowarzyszenie Pomocy Osobom z Zespołem Retta, Polskie Stowarzyszenie Pomocy Osobom z Zespołem Pradera-Williego, Polskie Towarzystwo Chorób Nerwowo-Mięśniowych, Rodzice dzieci z CZD, Rodzice niepełnosprawnych dzieci – dyskusja na każdy temat, Rodzinamuko, Rzadkie choroby metaboliczne – grupa wsparcia, Specjaliści, Lekarze, Terapeuci, Rodzice – Zespół Coffin-Siris, StopDuchenne, Stowarzyszenie Chorych na Mukopolisacharydozę (MPS) i Choroby Rzadkie, Stowarzyszenie Ehlers-Danlos Polska, Stowarzyszenie Marfan Polska, Stowarzyszenie na Rzecz Dzieci z Zaburzeniami Genetycznymi GEN, Stowarzyszenie Osób z Wrodzoną Łamliwością Kości (O.I) – Polska, Stowarzyszenie Rodzin z Chorobą Fabry'ego, Stowarzyszenie Rodzin z Chorobą Gauchera, Stowarzyszenie Rodzin z Zespołem Angelmana, Stowarzyszenie Zespołu Williamsa, Stowarzyszenie 22q11 Polska, Stwardnienie guzowate TSC, Syndrom Collins'a Treachera w Polsce, Wrodzona łamliwość kości, Wyjątkowa dziewczynka – Zespół Kabuki, Wyjątkowi! Wady i choroby genetyczne, Zespół Churga-Strauss ZChS – Polska, Zespół Jouberta, Zespół KABUKI Niikawa-Kuroki POLSKA, Zespół Klinefeltera – grupa wsparcia dla dorosłych i rodziców dzieci z ZK, Zespół Noonan – wyjątkowe dzieci = nasze Noonanki, Zespół Silvera-Russella, Zespół Smith-Magenis Polska, Zespół Sotosa Polska, Zespół Turnera – Wyjątkowe Dziewczynki – Kobiety, Zespół Wolfa-Hirschhorna, Choroba Wilsona, Ziarniniak Wegenera. We also acknowledge using Grammarly (v. 1.2.135.1595) and ChatGPT to support language correction and improve text clarity. All edits were critically reviewed and revised before finalising the manuscript.

Ethics Approval and Consent to Participate

This study was performed in line with the principles of the Declaration of Helsinki. Ethics and research governance approval were obtained from the PUMS Bioethics Committee (KB–228/24). Informed written consent was obtained from all individual participants in the study.

Consent for publication

Not applicable

Clinical trial number

Not applicable

Availability of data and materials

The data supporting this study's findings are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Funding sources

There are no sources of funding to declare.

Authors' contributions

Dariusz Walkowiak and Jan Domaradzki contributed to the conception and design of the study. Jan Domaradzki collected the data and coordinated the study. Dariusz Walkowiak and Piotr Jabkowski performed material preparation and analysis. Piotr Jabkowski prepared tables and Figures. All authors contributed to the preparation of the first draft of the manuscript and read and approved the final manuscript. All authors contributed equally to this paper.

References

1. The Lancet Global Health null. The landscape for rare diseases in 2024. Lancet Glob Health. 2024 Mar;12(3):e341. https://doi.org/10.1016/S2214-109X(24)00056-1
2. Ministerstwo Zdrowia [Internet]. [cited August 14 2025]. Rada Ministrów przyjęła Plan dla Chorób Rzadkich na lata 2024-2025 - Ministerstwo Zdrowia - Portal Gov.pl. https://www.gov.pl/web/zdrowie/rada-ministrow-przyjela-plan-dla-chorob-rzadkich-na-lata-2024-2025
3. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020 Feb;28(2):165–73. https://doi.org/10.1038/s41431-019-0508-0
4. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. iScience [Internet]. 19 August 2022 [cited 28 July 2025];25(8). https://www.cell.com/iscience/abstract/S2589-0042(22)00970-1
5. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare diseases are there? Nat Rev Drug Discov. 2020 Feb;19(2):77–8. https://doi.org/10.1038/d41573-019-00180-y

6. Kancelaria Prezesa Rady Ministrów [Internet]. [cited 14 September 2025]. Projekt uchwały Rady Ministrów w sprawie przyjęcia dokumentu Plan dla Chorób Rzadkich na lata 2024-2025 - Kancelaria Prezesa Rady Ministrów - Portal Gov.pl. https://www.gov.pl/web/premier/projekt-uchwaly-rady-ministrow-w-sprawie-przyjecia-dokumentu-plan-dla-chorob-rzadkich-na-lata-2024-2025
7. What is a rare disease? [Internet]. EURORDIS-Rare Diseases Europe. [cited 28 August 2025]. https://www.eurordis.org/information-support/what-is-a-rare-disease
8. RARE Disease Facts [Internet]. Global Genes. [cited 28 August 2025]. https://globalgenes.org/rare-disease-facts
9. Domike R, Raju GK, Sullivan J, Kennedy A. Expediting treatments in the 21st century: orphan drugs and accelerated approvals. Orphanet Journal of Rare Diseases. 2024 Nov;19(1):418. https://doi.org/10.1186/s13023-024-03398-1
10. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the forty years of the Orphan Drug Act. Orphanet J Rare Dis. 2023 Jun;18(1):163. https://doi.org/10.1186/s13023-023-02790-7
11. Claessens Z, Vanneste A, Van Isterdael C, Verbeke C, Wens I, Huys I. Criteria to evaluate unmet health-related needs of persons living with rare diseases and their caregivers: rapid literature review and stakeholder consultations. Orphanet Journal of Rare Diseases. 2025 Jul;20(1):321. https://doi.org/10.1186/s13023-025-03838-6
12. Mcmullan J, Lohfeld L, McKnight AJ. Needs of informal caregivers of people with a rare disease: a rapid review of the literature. BMJ Open. 2022 Dec;12(12):e063263. https://doi.org/10.1136/bmjopen-2022-063263
13. Rihm L, Dreier M, Rezvani F, Wiegand-Grefe S, Dirmaier J. The psychosocial situation of families caring for children with rare diseases during the COVID-19 pandemic: results of a cross-sectional online survey. Orphanet J Rare Dis. 2022 Dec;17:449. https://doi.org/10.1186/s13023-022-02595-0
14. Domaradzki J, Walkowiak D. Invisible patients in rare diseases: parental experiences with the healthcare and social services for children with rare diseases. A mixed method study. Sci Rep. 2024 Jun;14(1):14016. https://doi.org/10.1038/s41598-024-63962-4
15. Currie G, Szabo J. „It is like a jungle gym, and everything is under construction": The parent's perspective of caring for a child with a rare disease. Child Care Health Dev. 2019 Jan;45(1):96–103. https://doi.org/10.1111/cch.12628
16. Anderson M, Elliott EJ, Zurynski YA. Australian families living with rare disease: experiences of diagnosis, health services use and needs for psychosocial support. Orphanet J Rare Dis. 2013 Feb;8:22. https://doi.org/10.1186/1750-1172-8-22
17. Al-Attar M, Butterworth S, McKay L. A quantitative and qualitative analysis of patient group narratives suggests common biopsychosocial red flags of undiagnosed rare disease. Orphanet J Rare Dis. 2024 Apr;19:172. https://doi.org/10.1186/s13023-024-03143-8
18. Ramalle-Gómara E, Domínguez-Garrido E, Gómez-Eguílaz M, Marzo-Sola ME, Ramón-Trapero JL, Gil-de-Gómez J. Education and information needs for physicians about rare diseases in Spain. Orphanet J Rare Dis. 2020 Jan;15(1):18. https://doi.org/10.1186/s13023-019-1285-0
19. Walkowiak D, Domaradzki J. Are rare diseases overlooked by medical education? Awareness of rare diseases among physicians in Poland: an explanatory study. Orphanet J Rare Dis. 2021 Sep;16(1):400. https://doi.org/10.1186/s13023-021-02023-9
20. Zhang H, Xiao Y, Zhao X, Tian Z, Zhang SY, Dong D. Physicians' knowledge on specific rare diseases and its associated factors: a national cross-sectional study from China. Orphanet J Rare Dis. 2022 Mar;17(1):120. https://doi.org/10.1186/s13023-022-02243-7
21. Walkowiak D, Bokayeva K, Miraleyeva A, Domaradzki J. The Awareness of Rare Diseases Among Medical Students and Practicing Physicians in the Republic of Kazakhstan. An Exploratory Study. Front Public Health. 2022 Apr;10:872648. https://doi.org/10.3389/fpubh.2022.872648
22. Rohani-Montez SC, Bomberger J, Zhang C, Cohen J, McKay L, Evans WRH. Educational needs in diagnosing rare diseases: A multinational, multispecialty clinician survey. Genet Med Open. 2023 Apr;1(1):100808. https://doi.org/10.1016/j.gimo.2023.100808
23. Jo A, Larson S, Carek P, Peabody MR, Peterson LE, Mainous AG. Prevalence and practice for rare diseases in primary care: a national cross-sectional study in the USA. BMJ Open. 2019 Apr;9(4):e027248. https://doi.org/10.1136/bmjopen-2018-027248
24. Angelis A, Kanavos P, López-Bastida J, Linertová R, Oliva-Moreno J, Serrano-Aguilar P, et al. Social/economic costs and health-related quality of life in patients with epidermolysis bullosa in Europe. Eur J Health Econ. 2016 Apr;17 Suppl 1(Suppl 1):31–42. https://doi.org/10.1007/s10198-016-0783-4
25. Chung CCY, Ng NYT, Ng YNC, Lui ACY, Fung JLF, Chan MCY, et al. Socio-economic costs of rare diseases and the risk of financial hardship: a cross-sectional study. Lancet Reg Health West Pac. 2023 Feb;34:100711. https://doi.org/10.1016/j.lanwpc.2023.100711
26. Andreu P, Karam J, Child C, Chiesi G, Cioffi G. The Burden of Rare Diseases: An Economic Evaluation. https://www.ultragenyx.com/wp-content/uploads/2024/05/Chiesi-EconomicBurdenofRareDiseasesFeb.-2022.pdf
27. Domaradzki J, Pospieszyńska-Martysiuk K, Dianow H, Węgrzyn J, Walkowiak D. Psychological and psychiatric service use among family caregivers of individuals with Angelman Syndrome: A cross-sectional study. Psychiatr Pol. 2025;1–16. https://doi.org/10.12740/PP/OnlineFirst/206031
28. Strzelczyk A, Schubert-Bast S, Bast T, Bettendorf U, Fiedler B, Hamer HM, et al. A multicenter, matched case-control analysis comparing burden-of-illness in Dravet syndrome to refractory epilepsy and seizure remission in patients and caregivers in Germa-

ny. Epilepsia. 2019 Aug;60(8):1697–710. https://doi.org/10.1111/epi.16099
29. Consumer Financial Protection Bureau [Internet]. 2024 [cited 2 September 2025]. Measuring financial well-being: A guide to using the CFPB Financial Well-Being Scale. https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-scale
30. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008 Aug;61(4):344–9. https://doi.org/10.1016/j.jclinepi.2007.11.008
31. Chiarotti F, Kodra Y, De Santis M, Bellenghi M, Taruscio D, Carè A, et al. Gender and burden differences in family caregivers of patients affected by ten rare diseases. Ann Ist Super Sanita. 2023 Apr-Jun;59(2):122–31. https://doi.org/10.4415/ANN_23_02_05
32. Chu SY, Wen CC, Weng CY. Gender Differences in Caring for Children with Genetic or Rare Diseases: A Mixed-Methods Study. Children (Basel). 2022 Apr;9(5):627. https://doi.org/10.3390/children9050627
33. R: The R Project for Statistical Computing [Internet]. [cited 14 September 2025]. https://www.r-project.org
34. Yang G, Cintina I, Pariser A, Oehrlein E, Sullivan J, Kennedy A. The national economic burden of rare disease in the United States in 2019. Orphanet J Rare Dis. 2022 Apr;17(1):163. https://doi.org/10.1186/s13023-022-02299-5
35. Yu J, Chen S, Zhang H, Zhang S, Dong D. Patterns of the Health and Economic Burden of 33 Rare Diseases in China: Nationwide Web-Based Study. JMIR Public Health Surveill. 2024 Aug;10:e57353. https://doi.org/10.2196/57353
36. Leśniowska J. Economic Burden of Rare Diseases With Common Diseases as a Comorbidity in Poland. European Management Studies. 2020;18(89):103–20. https://doi.org/10.7172/1644-9584.89.6
37. Kopciuch D, Zaprutko T, Paczkowska A, Nowakowska E. Costs of treatment of adult patients with cystic fibrosis in Poland and internationally. Public Health. 2017 Jul;148:49–55. https://doi.org/10.1016/j.puhe.2017.03.003
38. Lagae L, Irwin J, Gibson E, Battersby A. Caregiver impact and health service use in high and low severity Dravet syndrome: A multinational cohort study. Seizure. 2019 Feb;65:72–9. https://doi.org/10.1016/j.seizure.2018.12.018
39. Schreiber-Katz O, Klug C, Thiele S, Schorling E, Zowe J, Reilich P, et al. Comparative cost of illness analysis and assessment of health care burden of Duchenne and Becker muscular dystrophies in Germany. Orphanet J Rare Dis. 2014 Dec;9:210. https://doi.org/10.1186/s13023-014-0210-9
40. Jarvis J, Chertavian E, Buessing M, Renteria T, Tu L, Hoffer L, et al. The economic impact of caregiving for individuals with Angelman syndrome in the United States: results from a caregiver survey. Orphanet J Rare Dis. 2025 Feb;20(1):82. https://doi.org/10.1186/s13023-025-03551-4
41. Nabbout R, Dirani M, Teng T, Bianic F, Martin M, Holland R, et al. Impact of childhood Dravet syndrome on care givers of patients with DS, a major impact on mothers. Epilepsy Behav. 2020 Jul;108:107094. https://doi.org/10.1016/j.yebeh.2020.107094
42. Rodríguez-Santana I, Mestre T, Squitieri F, Willock R, Arnesen A, Clarke A, et al. Economic burden of Huntington disease in Europe and the USA: Results from the Huntington's Disease Burden of Illness study. Eur J Neurol. 2023 Apr;30(4):1109–17. https://doi.org/10.1111/ene.15645
43. Howat-Rodrigues ABC, Laks J, Marinho V. Translation, cross-cultural adaptation, and psychometric properties of the Brazilian Portuguese version of the Consumer Financial Protection Bureau Financial Well-Being scale. Trends Psychiatry Psychother. 2021 Apr-Jun;43(2):134–40. https://doi.org/10.47626/2237-6089-2020-0034
44. Evans E, Jacobs M. Diabetes and Financial Well-Being: Differential Hardship Among Vulnerable Populations. Sci Diabetes Self Manag Care. 2024 Aug;50(4):263–74. https://doi.org/10.1177/26350106241256324
45. Jardaly M, Antoun J, Sakr R, Doumiati H, Shaarani I. Financial Literacy and Wellbeing Among Medical Students, Residents, and Attending Physicians in Lebanon: Results From a Nationwide Multi-Centered Survey. Inquiry. 2024 Jan-Dec;61:469580241294135. https://doi.org/10.1177/00469580241294135
46. Ellwood S, Weathers J, DeMello J, Graves L, Antoun J, Soares N. Personal Financial Well-being of Family Medicine Residents and Residency Curricula: A CERA Study. PRiMER. 2023 Jan;7:415901. https://doi.org/10.22454/PRiMER.2023.415901
47. Noronha V, Tongaonkar A, Pillai A, Rao AR, Kumar A, Sehgal A, et al. Prevalence and impact of financial toxicity in older patients with cancer: a prospective observational study in India. Support Care Cancer. 2025 Apr;33(5):416. https://doi.org/10.1007/s00520-025-09252-9
48. Dosemane D, Khadilkar MN, Kanthila J, Mithra PP. Quality of life and perceived financial implications among otorhinolaryngologists during the COVID-19 pandemic across India. Acta Otorhinolaryngol Ital. 2021 Aug;41(4):289–95. https://doi.org/10.14639/0392-100X-N1229
49. Huang CHO, Leung RHL, Sit KY, Tsui TYN, Wong JYH, Fung HW. The Financial Well-Being Scale: Reliability, Validity, and Clinical Correlates Among Childhood Trauma Survivors. Research on Social Work Practice. 2025;10497315241312889. https://doi.org/10.1177/10497315241312889
50. Kempson E, Poppe C. Understanding Financial Well-Being and Capability [Internet]. Høgskolen i Oslo og Akershus; 2018. https://oda.oslomet.no/oda-xmlui/handle/20.500.12199/5357

ORIGINAL PAPER

JMS Journal of Medical Science

Nesfatin-1 and Ghrelin/GOAT as Potential Biomarkers in Adolescent Headache with Temporomandibular Disorders: A Cross-Sectional Study

Anna Sojka
Chair and Clinic of Prosthodontics and Gerodontology, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0002-0193-0491

Barbara Dorocka-Bobkowska
Chair and Clinic of Prosthodontics and Gerodontology, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0003-3659-7761

Yasmin Bartosik
Chair and Clinic of Prosthodontics and Gerodontology, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0009-0007-8932-1810

Ozgun Yetkin
Department of Developmental Neurology, Poznan University of Medical Sciences, Poznań, Poland
Doctoral School, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0003-4110-8002

Corresponding author: 90457@student.ump.edu.pl

Barbara Steinborn
Poznan University of Medical Sciences Department of Developmental Neurology, Poznań, Poland
https://orcid.org/0000-0002-9839-8162

Marcin Żarowski
Poznan University of Medical Sciences Department of Developmental Neurology, Poznań, Poland
https://orcid.org/0000-0002-9800-9446

https://doi.org/10.20883/medical.e1418

Keywords: nesfatin-1, ghrelin/GOAT, headache, temporomandibular disorders, adolescent

Received 2025-09-12
Accepted 2025-12-23
Published 2025-12-31

How to Cite: Sojka A, Dorocka-Bobkowska B, Bartosik Y, Yetkin O, Steinborn B, Żarowski M. Nesfatin-1 and Ghrelin/GOAT as Potential Biomarkers in Adolescent Headache with Temporomandibular Disorders: A Cross-Sectional Study. Journal of Medical Science. 2025 December;94(4);e1418. doi:10.20883/medical.e1418

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Introduction. Nesfatin-1 and Ghrelin O-acetyltransferase (GOAT) are peptide hormones involved in stress response and pain modulation beyond their metabolic functions. This study investigated whether these biomarkers differ between adolescents with headaches and healthy controls and examined their relationship with temporomandibular disorders (TMD).

Material and methods. This cross-sectional study measured Nesfatin-1 and Ghrelin/GOAT levels in the serum and saliva of 44 adolescents with headaches (aged 10-17 years, diagnosed per ICHD-3 beta criteria) and 20 age-matched healthy controls. TMD was assessed using the Diagnostic Criteria for TMD (DC/TMD). Statistical comparisons were performed using non-parametric tests, with significance set at $p < 0.05$.

Results. Nesfatin-1 was elevated in all headache patients compared to controls in both blood (patient median range: 39.5-45.0 ng/mL vs. control median range: 8.1-9.0 ng/mL, $p < 0.001$) and saliva (patient medi-

an range: 6.6-8.2 ng/mL vs. control median range: 1.8-2.3 ng/mL, p < 0.001) regardless of headache type (migraine vs. tension-type) or TMD status. Ghrelin/GOAT showed opposite patterns in blood versus saliva: serum levels were higher in patients (median range: 5.9-6.6 ng/mL vs. 1.4 ng/mL in controls, p < 0.001), while salivary levels were lower (median range: 0.2 ng/mL vs. 1.4 ng/mL in controls, p < 0.001). Gender differences emerged in clinical manifestations: females had more muscle pain (74% vs 38%, p = 0.016), while males reported more psychological distress (90% vs 48%, p = 0.003).
Conclusions. This study reveals altered Nesfatin-1 and Ghrelin/GOAT levels in adolescents with headaches, independent of headache type or TMD status. These findings suggest potential utility as pain-related biomarkers, though validation studies with larger sample sizes and receiver operating characteristic analyses are needed before clinical implementation. Gender specific clinical patterns underscore the importance of sex-stratified approaches in adolescent pain management.

Introduction

Headache is among the most common complaints in pediatric neurology clinics, affecting 54-58% of children and adolescents globally [1-3]. These headaches frequently lead to school absences and impaired academic performance, creating a substantial burden on young people during critical developmental years [2,3]. Both migraine and tension-type headache (TTH) can significantly impact quality of life and functioning, making early recognition and treatment essential [4,5].

Temporomandibular disorders (TMD) encompass pain and dysfunction of the masticatory muscles, the temporomandibular joint (TMJ), and surrounding structures. They are the most common cause of chronic orofacial pain and the primary source of non-dental pain in the craniofacial region [6]. A notable clinical overlap exists between TMD and headache, particularly tension-type headache (TTH), though this relationship is often overlooked in clinical practice [7,8]. The International Classification of Headache Disorders, 3rd edition (beta version) (ICHD-3 beta), acknowledges this diagnostic ambiguity, noting that when the TMD diagnosis is uncertain, the headache may be classified as 'TTH or one of its subtypes' (presumably with pericranial muscle tenderness) [9].

Current headache diagnosis in adolescents relies primarily on clinical history and examination, with a limited objective markers available to aid diagnosis or predict outcomes. The availability of reliable biomarkers would help resolve diagnostic uncertainty and potentially guide treatment decisions. While some biomarkers have been identified in adult migraine patients from peripheral blood and cerebrospinal fluid [10,11], research in paediatric populations remains limited.

Nesfatin-1, discovered in 2006, is a peptide derived from the nucleobindin-2 (NUCB2) gene that was initially characterised as a satiety signal [12]. However, subsequent research revealed broader roles in stress response, anxiety, and pain processing [13]. Nesfatin-1 is expressed in brain regions involved in both stress and pain processing, including the hypothalamus, periaqueductal grey and rostral ventromedial medulla. These areas are key components of the descending pain-modulatory system, suggesting that Nesfatin-1 may link stress and pain perception.

Similarly, ghrelin was identified in 1999 as a growth hormone-releasing neuropeptide that influences appetite and energy metabolism [14]. Ghrelin undergoes post-translational modification by ghrelin O-acyltransferase (GOAT), which attaches a fatty acid chain that is essential for biological activity [15]. Beyond its metabolic functions, ghrelin influences pain processing through central nervous system actions. The ghrelin system has been implicated in anxiety disorders and stress responses [16], conditions that frequently co-occur with chronic pain syndromes.

Adolescence represents a critical developmental period characterised by dramatic hormonal changes, brain maturation, and psychosocial transitions. These factors may influence both the pain experience and the expression of biomarkers. Additionally, gender differences in pain conditions emerge during adolescence, with females showing higher rates, including TMD [17,18].

Despite the clinical overlap between headache and TMD and the theoretical rationale for neuroendocrine biomarkers, no previous study has examined Nesfatin-1 and Ghrelin/GOAT levels in adolescents with headaches. This study aimed to address this gap with the following objectives:

- Primary Objective:
  To compare serum and salivary levels of Nesfatin-1 and Ghrelin/GOAT between adolescents with headaches and age-matched healthy controls.
- Secondary Objectives:
  1. To examine whether biomarker levels differ between migraine and TTH
  2. To assess whether biomarker levels vary across TMD diagnostic categories (myalgia, disc displacement with reduction, and no TMD)
  3. To evaluate gender-specific differences in biomarker expression and clinical manifestations

Materials and methods

Study Design and Ethical Approval

This cross-sectional study was performed at the Department of Developmental Neurology, Poznan University of Medical Sciences, from November 2017 to December 2020. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Bioethical Committee of Poznan University of Medical Sciences (approval number 550/17) [19]. Written informed consent was obtained from all participants and their legal guardians following a comprehensive explanation of the study protocol. Participant anonymity was maintained throughout the study, and participation was voluntary with no financial incentives offered. The study procedures were designed to minimise any potential burden on the participants.

Participation Selection and Recruitment

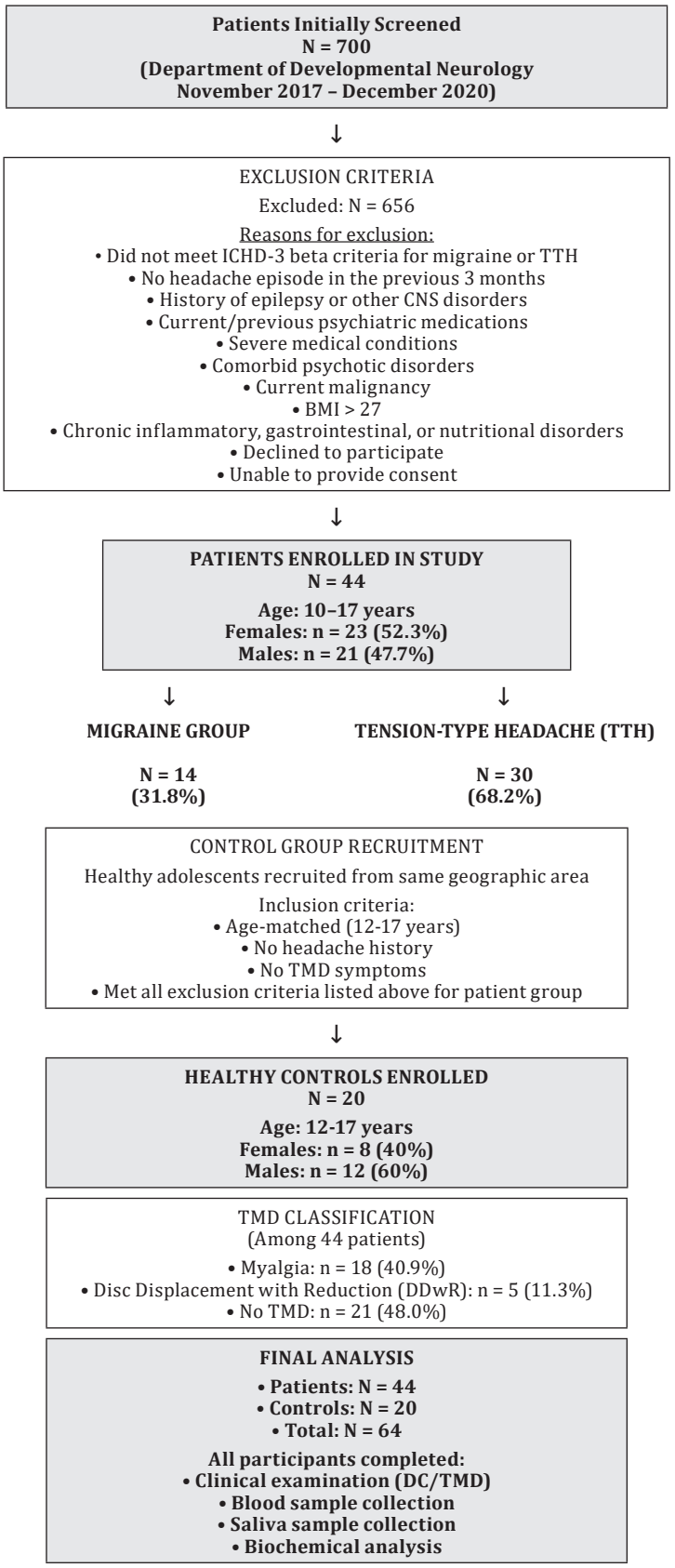
Of the 700 patients initially screened, 44 adolescents met the inclusion criteria and were enrolled in the study (see **Figure 1**). Inclusion criteria for patients: age 10-17 years; confirmed migraine or TTH according to the ICHD-3 beta criteria [9]; at least one headache episode in the previous 3 months; and ability to provide consent from themselves and their parents. Exclusion criteria for both groups: history of epilepsy or other central nervous system disorders; current or previous use of psychiatric medications; presence of severe medical conditions; comorbid psychotic disorders; current malignancy; body mass index (BMI) exceeding 27; and presence of chronic inflammatory, gastrointestinal, or nutritional disorders that could potentially affect Nesfatin-1 and Ghrelin levels. Additional exclusion criteria for the control group included any history of headaches or TMD symptoms. Subjects with a BMI greater than 27 were excluded from the study, as body mass significantly affects the secretion and circulating levels of both Nesfatin-1 and Ghrelin, which could confound the interpretation of the results [20,21].

The control group consisted of 20 age-matched healthy adolescents from the same geographic area, who did not have headache or TMD symptoms. Control patients met all exclusion criteria listed above. The patient age range (10-17 years) was defined according to ICHD-3 beta criteria. In comparison, the control group was recruited from ages 12 to 17 years to match the patient cohort's median age distribution. No significant age differences were observed between the final recruited patient group and the control group.

The smaller control group size (n = 20 vs. n = 44 patients) was determined based on power calculations. Using pilot data showing mean Nesfatin-1 levels of approximately 40 ± 16 ng/mL in patients versus 8 ± 2 ng/mL in controls, with $\alpha = 0.05$ and power = 0.80, a minimum of 15 controls were required to detect this significant effect (Cohen's $d \approx 2.5$). We recruited 20 controls to allow for potential sample quality issues. The unequal group sizes reflect pragmatic challenges in recruiting healthy adolescents for research procedures, including blood draws, which proved more difficult than enrolling patients already presenting for clinical care.

Headache diagnosis was established through a comprehensive assessment protocol that included investigator-designed parental questionnaires, a detailed medical history, and a clinical examination performed by a certified child neurologist (M.Z. and B.S.).

All diagnoses were made according to ICHD-3 beta criteria [9]. Patients were classified as having either migraine or tension-type headache.



BMI = Body Mass Index; CNS = Central Nervous System; DC/TMD = Diagnostic Criteria for Temporomandibular Disorders; ICHD-3 = International Classification of Headache Disorders, 3rd edition; TMD = Temporomandibular Disorders; TTH = Tension-Type Headache

Figure 1. STROBE flowchart showing participant recruitment, selection, and classification in the cross-sectional study of Nesfatin-1 and Ghrelin/GOAT as potential biomarkers in adolescent headache with temporomandibular disorders.

Temporomandibular Disorder Assessment

TMD examinations were performed by trained dentists (Y.B. and A.S.M.) who were blinded to participants' headache diagnoses and control-group status. The examination was conducted in accordance with the protocol outlined in the Diagnostic Criteria for TMD (DC/TMD) Axis II guidelines [22,23].

A comprehensive clinical assessment included medical and dental histories, with particular attention to parafunctional habits, such as bruxism. The intraoral examination assessed signs of bruxism, including dental wear facets, fractures, masseter hypertrophy, and soft tissue changes (linea alba and tongue scalloping). Bruxism was identified based on clinical signs during examination rather than polysomnography, and we did not differentiate between awake and sleep bruxism in this study.

Two major diagnostic categories were assessed: myalgia and disc displacement with reduction (DDwR). Myalgia was evaluated by palpation of the temporalis and masseter muscles and by pain during jaw opening, whereas DDwR was identified by joint sounds (clicking or popping). Following DC/TMD protocols, pain provocation during examination was confirmed as familiar pain matching the patient's primary complaint. Functional measurements included maximum mouth opening (measured from the maxillary to mandibular incisal edges at midline), with an opening of less than 40 mm considered restricted [24]. Mandibular deviation was defined as a displacement of 2 mm or more to either side during opening movement. The assessment also included a distress evaluation using the Patient Health Questionnaire-4 (PHQ-4), the PHQ-9, the Generalised Anxiety Disorder 7-item scale (GAD-7), and the Oral Behaviour Checklist (OBC) [23]. Parafunctional habits were systematically assessed using the OBC, which categorises oral behaviours into 'bite' parafunctions (clenching, grinding) and 'non-bite' parafunctions (nail biting, cheek chewing, lip biting). Physical symptoms were defined using the somatic symptom items from the PHQ-9 as part of the DC/TMD Axis II distress assessment [23].

Biochemical Sample Collection and Analysis

Fasting venous blood (5 mL) and unstimulated whole saliva (2 mL) samples were collected simultaneously between 08:00 and 09:00. Participants fasted overnight and refrained from eating, drinking (except water), and oral hygiene procedures for at least 1 hour before sample collection. Saliva was collected using the passive drooling method: participants were instructed to allow saliva to accumulate naturally in the floor of the mouth and drool into sterile collection tubes over approximately 5 minutes until 2 mL was obtained.

Saliva samples underwent double centrifugation (3046.4 g, 2 × 10 minutes), with supernatant collection between cycles, and were stored at -20°C in 0.5 mL aliquots. Blood samples were processed immediately and stored at -20°C until analysis, a temperature compatible with the ELISA kits used. Thyroid-stimulating hormone (TSH), fasting plasma glucose, and total cholesterol were analysed using a Cobas 8000 platform (Roche Diagnostics GmbH, Germany). Nesfatin-1 and Ghrelin/GOAT levels were duplicated using commercial *enzyme-linked immunosorbent* assay (ELISA) kits (Human NEFA/Nesfatin-1 and Human MBOAT4/GOAT).

Nesfatin-1 and Ghrelin/GOAT levels were measured using commercial enzyme-linked immunosorbent assay (ELISA) kits (Human NEFA/Nesfatin-1, Cat# E8726h; Human MBOAT4/GOAT, Cat# E3525h; Wuhan EIAab Science Inc., China). Both kits had a detection range of 0.312-20 ng/mL. The Nesfatin-1 kit demonstrated a sensitivity of <0.12 ng/mL, with an intra-assay coefficient of variation (CV) of ≤6.2%, an inter-assay CV of ≤8.5%, and an average recovery of 94%. The Ghrelin/GOAT kit demonstrated a sensitivity of <0.1 ng/mL, with intra-assay CV ≤5.9%, inter-assay CV ≤10.1%, and an average recovery of 102%.

Statistical Analysis:

Descriptive statistics included means, standard deviations, and medians for continuous variables. The Shapiro-Wilk test was used to assess the normality of the distribution.

Due to non-normal distribution, between-group comparisons were conducted using non-parametric tests: the Mann-Whitney U test for two-group comparisons and the Kruskal-Wallis test for multiple-group analyses. Categorical variables were presented as counts (N) and percentages, and differences in proportions were

analysed using chi-square tests. Statistical significance was set at $p < 0.05$. All analyses were performed using Statistica v13.0 (Dell Inc., Round Rock, TX, USA).

Although the Shapiro-Wilk test indicated non-normal distribution for some variables, data are presented as both mean±standard deviation and median for several reasons: (1) to facilitate comparison with existing literature which predominantly reports means, (2) the large effect sizes observed reduce concern about violation of normality assumptions for descriptive purposes, and (3) means provide information about the magnitude of differences. However, all inferential statistical tests (Mann-Whitney U and Kruskal-Wallis) were nonparametric and did not require the assumption of normality.

Results

Study Population and Characteristics

The study cohort comprised 44 adolescents with headaches (23 females [52.3%] and 21 males [47.7%] aged 10-17 years) (Figure 1). All participants had experienced the last episodes of headaches within the previous three months, and this was the main reason for their admission to the Department. Based on ICHD-3 beta diagnostic criteria, 14 participants (31.8%) were diagnosed with migraine, while 30 (68.1%) met the requirements for TTH [9]. According to DC/TMD Axis I criteria [22], the distribution of TMD was as follows: 18 patients (40.9%) had myalgia, five patients (11.3%) had DDwR, and 21 patients (48.0%) had no TMD symptoms. The control group comprised 20 healthy adolescents (8 females [40%] and 12 males [60%], aged 12-17 years) without headache or TMD symptoms.

Table 1 presents the demographic and biochemical characteristics of patients with headaches compared to controls, stratified by sex. No statistically significant differences were observed between female patients and female controls, or between male patients and male controls, in anthropometric measurements (BMI, weight, height) or in basic biochemical parameters (TSH, total cholesterol, glucose). The mean BMI values were comparable between groups (20.78 kg/m² in female patients vs. 19.53 kg/m² in female controls; 19.05 kg/m² in male patients vs. 19.90 kg/m² in male controls).

Table 1. Characteristics of the investigated headache patients and controls.

Characteristic	Female (N = 23) mean ± SD (median)	Female control mean ± SD (median)	p-value	Male (N = 21) mean ± SD (median)	Male control mean ± SD (median)	p-value
BMI [kg/m^2]	21.05 ± 2.41 (20.78)	20.65 ± 4.10 (19.53)	0.603	20.23 ± 2.53 (19.05)	19.74 ± 2.63 (19.90)	0.708
Weight [kg]	51.91 ± 9.25 (52.00)	45.25 ± 16.74 (47.00)	0.416	48.67 ± 13.67 (50.00)	48.63 ± 17.42 (43.75)	0.721
Height [cm]	156.43 ± 10.41 (155.00)	144.88 ± 15.99 (153.50)	0.099	153.57 ± 14.04 (155.00)	154.08 ± 19.98 (149.50)	0.895
TSH level [µU/ml]	3.47 ± 1.73 (3.41)	3.08 ± 1.01 (2.60)	0.527	3.23 ± 1.38 (2.83)	3.37 ± 1.16 (3.34)	0.369
Total cholesterol [mg/dl]	161.70 ± 25.66 (158.00)	158.50 ± 29.49 (153.50)	0.635	162.81 ± 34.33 (167.00)	161.50 ± 24.25 (161.50)	0.808
Glucose level [mg/dl]	88.74 ± 6.05 (88.00)	90.63 ± 4.96 (88.50)	0.286	89.71 ± 7.46 (93.00)	93.92 ± 5.65 (94.50)	0.106
Nesfatin-1 (blood) [ng/mL]	39.12 ± 18.28 (39.48)	9.14 ± 0.63 (9.04)	**<0.001**	43.97 ± 14.47 (45.00)	7.52 ± 2.50 (8.14)	**<0.001**
Nesfatin-1 (saliva) [ng/mL]	10.57 ± 9.62 (6.61)	2.16 ± 0.68 (2.32)	**<0.005**	11.16 ± 9.19 (8.22)	1.85 ± 0.82 (1.76)	**<0.001**
Ghrelin/GOAT (blood) [ng/mL]	6.35 ± 2.75 (5.86)	1.47 ± 0.22 (1.43)	**<0.001**	6.21 ± 2.46 (6.64)	1.46 ± 0.45 (1.39)	**<0.001**
Ghrelin/GOAT (saliva) [ng/mL]	0.29 ± 0.24 (0.21)	1.65 ± 0.58 (1.41)	**<0.001**	0.24 ± 0.25 (0.18)	1.63 ± 0.78 (1.37)	**<0.001**

Data are presented as mean ± SD (median) for continuous variables. Between-group comparisons were performed using the Mann-Whitney U test. *p < 0.05 (statistically significant). BMI, Body Mass Index; GOAT, Ghrelin O-acetyltransferase; TSH, Thyroid-Stimulating Hormone.

Primary Outcome: Neuroendocrine Biomarker Levels

Blood Nesfatin-1 levels in blood were markedly elevated in both female (patients 39.12 ± 18.28 ng/mL (median 39.48) vs. controls 9.14 ± 0.63 ng/mL (median 9.04), $p < 0.001$, Cohen's d = 2.51) and male (patients 43.97 ± 14.47 ng/mL (median 45.00) vs. controls 7.52 ± 2.50 ng/mL, (median 8.14) $p < 0.001$, Cohen's d = 3.35) compared to their controls.

Similarly, salivary Nesfatin-1 levels were significantly higher in females (patients 10.57 ± 9.62 ng/mL (median 6.61) vs. controls 2.16 ± 0.68 ng/mL (median 2.32), $p < 0.005$, Cohen's d = 1.12) and males (patients 11.16 ± 9.19 ng/mL (median 8.22) vs. controls 1.85 ± 0.82 ng/mL (median 1.76), $p < 0.001$, Cohen's d = 1.30) (see **Table 1**).

A distinctive pattern was observed with Ghrelin/GOAT: blood levels were significantly elevated in both female (patients 6.35 ± 2.75 ng/mL (median 5.86) vs. controls 1.47 ± 0.22 ng/mL (median 1.43), $p < 0.001$, Cohen's d = 2.32) and male (patients 6.21 ± 2.46 ng/mL (median 6.64) vs. controls 1.46 ± 0.45 ng/mL, (median 1.39), $p < 0.001$, Cohen's d = 2.47). In contrast, salivary levels were markedly lower in both female (patients 0.29 ± 0.24 ng/mL (median 0.21) vs. controls 1.65 ± 0.58 ng/mL, (median 1.41), $p < 0.001$, Cohen's d=-3) and male (patients 0.24 ± 0.25 ng/mL (median 0.18) vs. controls 1.63 ± 0.78 ng/mL, (median 1.37), $p < 0.001$, Cohen's d = 2.50) compared to controls. The negative Cohen's d values for salivary ghrelin/GOAT indicate lower levels in patients than in controls (see **Table 1**).

Biomarker Levels Across Headache Types

No statistically significant differences were observed between the migraine patients (n = 14) and the TTH (n = 30) groups for any measured parameter (see **Table 2**).

Biomarker Levels Across TMD Categories

Comparison between TMD subgroups revealed significant differences in physical symptoms. Patients with myalgia showed a significantly higher prevalence of physical symptoms compared to the no-TMD group (89% vs 47%, $p < 0.006$). Similarly, the DDwR group demonstrated substantially more physical symptoms than the no-TMD group (100% vs. 47%, $p < 0.035$). The Kruskal-Wallis test revealed no significant differences in Nesfatin-1 levels (blood and saliva), Ghrelin/GOAT levels (blood and saliva), or TSH levels among the Myalgia, DDwR, and no-TMD groups (see **Table 3**).

Comparison of Neuroendocrine Markers Between All Patient Subgroups and Controls

Nesfatin-1 and Ghrelin/GOAT levels revealed consistent patterns across all headache subgroups compared to controls (see **Table 4**). Nesfatin-1 levels in blood were significantly elevated in migraine patients (42.94 ± 17.04 ng/ml), TTH patients (40.73 ± 16.58 ng/ml), those with TMD symptoms (41.74 ± 16.75 ng/ml), and those without TMD symptoms (41.10 ± 16.76 ng/ml) compared to controls (all $p < 0.001$). Similarly, salivary Nesfatin-1 levels were significantly higher in all patient subgroups (ranging from 10.13 ± 9.48 to

Table 2. Neuroendocrine Markers and Clinical Characteristics in Migraine versus Tension-Type Headache Patients.

Headache	Nesfatin-1 (ng/ml) Blood mean ± SD (median)	Nesfatin-1 (ng/ml) Saliva mean ± SD (median)	Ghrelin/ GOAT (ng/ml) Blood mean ± SD (median)	Ghrelin/ GOAT (ng/ml) Saliva mean ± SD (median)	TSH (uU/ml)	Myalgia TMD N(%)	DDwR TMD N(%)	Parafunctional bite or nonbite N(%)	Distress N(%)	Physical symptoms N(%)
Migraine (N = 14)	42.94 ± 17.04 (42.94)	10.98 ± 8.23 (8.62)	6.26 ± 2.65 (5.91)	0.25 ± 0.26 (0.16)	3.57 ± 1.95 (2,89)	8 (57%)	1 (7%)	14 (100%)	8 (57%)	23 (76%)
TTH (N = 30)	40.73 ± 16.58 (41.70)	10.79 ± 9.91 (4.52)	6.30 ± 2.60 (6.09)	0.27 ± 0.23 (0.21)	3.25 ± 1.36 (3.20)	10 (33%)	4 (13%)	18 (93%)	20 (66%)	8 (57%)
p-value	0.998	0.457	0.990	0.296	0.990	0.131	0.555	0.399	0.564	0.224

Data are presented as mean ± SD (median) for continuous variables and n (%) for categorical variables. Between-group comparisons were performed using the Mann-Whitney U test for continuous variables and the chi-square test for categorical variables. *p < 0.05 (statistically significant). DDwR, Disc displacement with reduction; GOAT, Ghrelin O-acetyltransferase; TMD, Temporomandibular Disorders; TSH, Thyroid-Stimulating Hormone; TTH, Tension-Type Headache.

Table 3. Comparison of Neuroendocrine Markers and Clinical Manifestations Across TMD Diagnostic Subgroups in Headache Patients.

TMD	Nesfatin-1 (ng/ml) Blood mean ± SD (median)	Nesfatin-1 (ng/ml) Saliva mean ± SD (median)	Ghrelin/ GOAT (ng/ml) Blood mean ± SD (median)	Ghrelin/ GOAT (ng/ml) Saliva mean ± SD (median)	Physical symptoms of TMD N(%) mean ± SD (median)	Distress N(%)	Parafun-ctional bite or nonbite N(%)	TSH (uU/ml) N(%)
Myalgia(M) N = 18	40.89 ± 17.23 (40.99)	11.59 ± 9.08 (10.30)	6.45 ± 2.85 (6.34)	0.18 ± 0.06 (0.16)	16 (89[a])	12 (67)	18 (100)	3.66 ± 1.88 (3.33)
Disc displacement with reduction (DDwR) N = 5	44.79 ± 16.32 (50.00)	11.22 ± 11.29 (8.22)	7.19 ± 3.58 (6.19)	0.25 ± 0.03 (0.25)	5 (100[a])	5 (100)	5 (100)	3.74 ± 1.34 (4.18)
No TMD N = 21	41.10 ± 16.76 (45.00)	10.13 ± 9.48 (4.48)	5.92 ± 2.13 (5.61)	0.33 ± 0.33 (0.21)	10 (47[b])	12 (57)	18 (85)	2.99 ± 1.27 (2.56)
p-value	0.628	0.823	0.764	0.067	***0.006***[a] 0.486 ***0.035***[a] ***vs No TMD***	0.522 0.149 0.077	0.119 0.393 0.392	0.289

Data are presented as mean ± SD (median) for continuous variables and n (%) for categorical variables. Multiple-group comparisons were performed using the Kruskal-Wallis test for continuous variables and the chi-square test for categorical variables. Post-hoc pairwise comparisons were conducted when the overall test was significant. *p < 0.05 (statistically significant). Superscripts indicate significant pairwise differences between groups (different letters denote p < 0.05). DDwR, Disc displacement with reduction; GOAT, Ghrelin O-acetyltransferase; TMD, Temporomandibular Disorders; TSH, Thyroid-Stimulating Hormone.

Table 4. Comparison of Nesfatin-1 and Ghrelin/GOAT Levels Between Headache Subgroups, TMD Categories, and Controls.

	Migraine (N = 14)	Control (N = 20)	p-value	TTH (N = 30)	Control (N = 20)	p-value	TMD (N = 23)	Control (N = 20)	p-value	No TMD (N = 21)	Control (N = 20)	p-value
Nesfatin-1 (ng/ml) **Blood** mean ± SD (median)	42.94 ± 17.04 (42.94)	8.16 ± 2.10 (8.64)	<0.001	40.73 ± 16.58 (41.70)	8.16 ± 2.10 (8.64)	***<0.001***	41.74 ± 16.75 (42.33)	8.16 ± 2.10 (8.64)	<0.001	41.10 ± 16.76 (45.00)	8.16 ± 2.10 (8.64)	***<0.001***
Nesfatin-1 (ng/ml) **Saliva** mean ± SD (median)	10.98 ± 8.23 (8.62)	1.97 ± 0.77 (2.01)	<0.001	10.79 ± 9.91 (4.52)	1.97 ± 0.77 (2.01)	***<0.001***	11.51 ± 9.32 (8.92)	1.97 ± 0.77 (2.01)	<0.001	10.13 ± 9.48 (4.48)	1.97 ± 0.77 (2.01)	***<0.001***
Ghrelin/ GOAT (ng/ml) **Blood** mean ± SD (median)	6.26 ± 2.65 (5.91)	1.46 ± 0.37 (1.41)	<0.001	6.30 ± 2.60 (6.09)	1.46 ± 0.37 (1.41)	***<0.001***	6.61 ± 2.95 (6.19)	1.46 ± 0.37 (1.41)	<0.001	5.92 ± 2.13 (5.61)	1.46 ± 0.37 (1.41)	***<0.001***
Ghrelin/ GOAT (ng/ml) **Saliva** mean ± SD (median)	0.25 ± 0.26 (0.16)	1.64 ± 0.69 (1.38)	<0.001	0.27 ± 0.23 (0.21)	1.64 ± 0.69 (1.38)	***<0,001***	0.19 ± 0.06 (0.19)	1.64 ± 0.69 (1.38)	<0.001	0.33 ± 0.33 (0.21)	1.64 ± 0.69 (1.38)	***<0.001***

Data are presented as mean±SD (median). Between-group comparisons were performed using the Mann-Whitney U test. *p < 0.05 (statistically significant). GOAT, Ghrelin O-acetyltransferase; TMD, Temporomandibular Disorders; TTH, Tension-Type Headache.

11.51 ± 9.32 ng/ml) compared to controls (1.97 ± 0.77 ng/ml) (all p < 0.001). Ghrelin/GOAT demonstrated a tissue-specific pattern: blood levels were significantly higher in all headache groups (ranging from 5.92 ± 2.13 to 6.61 ± 2.95 ng/ml) compared to controls (1.46 ± 0.37 ng/ml) (all p < 0.001), while salivary levels showed the opposite trend, being significantly lower in all patient groups (ranging from 0.19 ± 0.06 to 0.33 ± 0.33 ng/ml) compared to controls (1.64 ± 0.69 ng/ml) (all p < 0.001).

Gender Specific Clinical Manifestations

Gender-specific analysis revealed significant differences in TMD manifestations. Muscle pain upon palpation was significantly more prevalent in females than males (74% n = 17/23 vs 38% n = 8/21, p < 0.016). Conversely, males exhibited a higher prevalence of TMD-free cases than females (66%, n = 14/21 vs. 30%, n = 7/23, p < 0.016). Psychological distress was significantly more prevalent in males compared to females (90% n = 19/21 vs 48% n = 11/23, p < 0.003) (see **Table 5**).

Myalgia diagnosis, according to DC/TMD criteria, is also more frequent in females (52%/23 vs. 28%, n = 6/23), but there is no significant difference (p < 0.105). While both groups showed comparable rates of parafunctional habits, males demonstrated a higher frequency of both bite (52% vs. 21%, 43% vs. 23%, p < 0.550) and non-bite (90% vs. 21%, 82% vs. 23%, p < 0.447) parafunctions. Physical symptoms were similarly distributed between females and males (70% vs. 23% and 71% vs. 21%, respectively). No significant differences were observed in any of the other assessed parameters (see **Table 5**).

Discussion

The current study provides new perspectives into the relationship between neuroendocrine markers and headaches with TMD in adolescents. The most notable finding was the consistent and significant increase in Nesfatin-1 levels in both blood and saliva among all headache groups, regardless of headache type or TMD status. This pattern, together with the tissue-specific changes in Ghrelin/GOAT levels, suggests that these peptides may serve as biomarkers for headache disorders in adolescents.

Neuroendocrine Dysregulation in Adolescent Headache

The elevated Nesfatin-1 levels reported in our headache patients are consistent with recent evidence regarding this peptide's functions beyond energy regulation. Nesfatin-1 has been associated with multiple stress-related disorders, as evidenced by Hofmann et al., who identified positive associations between Nesfatin-1 levels and scores of anxiety, depression, and perceived stress in individuals of normal weight [25]. Xiao et al. showed a favourable correlation between plasma Nesfatin-1 levels and the degree of depression [26]. Considering the recognised correlation between psychological discomfort and pain

Table 5. Gender Distribution of TMD, Parafunctional Habits, and Associated Symptoms in Adolescents with Headache.

Signs and symptoms of TMD	Female (N = 23)	Male (N = 21)	p-value
Pain symptoms in the temple area	16 (70%)	18 (86%)	0.203
Muscle pain with palpation	17 (74%)	8 (38%)	***0.016***
Bite parafunctions	10 (43%)	11 (52%)	0.550
Non-bite parafunctions	19 (82%)	19 (90%)	0.447
Signs of bruxism	10 (43%)	10 (48%)	0.739
Physical symptoms N(%)	16 (70%)	15 (71%)	0.942
Distress (Depression&Anxiety) N(%)	11 (48%)	19 (90%)	***0.003***
Diagnosis of TMD – Myalgia	12 (52%)	6 (28%)	0.105
Diagnosis of TMD – DDwR	3 (13%)	2 (9%)	0.673
No TMD	7 (30%)	14 (66%)	***0.016***

Data are presented as n (%). Between-group comparisons were performed using the chi-square test. *p < 0.05 (statistically significant). DDwR, Disc displacement with reduction; TMD, Temporomandibular Disorders.

problems, our findings further substantiate that Nesfatin-1 may function as a neurobiological mediator connecting stress and pain perception in adolescents experiencing headaches.

The tissue-specific distribution of Ghrelin/GOAT, augmented in blood yet diminished in saliva, is particularly intriguing. The mechanisms underlying this differential expression across body fluids remain unclear and warrant further investigation. Potential explanations include differences in local production, degradation rates, or transport mechanisms between systemic circulation and salivary glands, though these hypotheses require direct experimental validation. Previous research has shown elevated plasma Ghrelin levels in children with anxiety disorders [16], but our study is the first to demonstrate this distinct pattern in adolescents with headaches. The mechanism underlying the inverse association between blood and salivary Ghrelin/GOAT levels warrants further examination and may yield insights into the pathophysiology of headache disorders.

Comparison with Established Headache Biomarkers

Our findings should be contextualised within the broader landscape of headache biomarker research. Calcitonin gene-related peptide (CGRP) is the most extensively studied biomarker for migraine in adults, and therapeutic agents targeting CGRP are now approved for migraine prevention [10, 11]. Pituitary adenylate cyclase-activating polypeptide (PACAP) also shows promise as a migraine marker [27]. However, these established biomarkers have limitations in pediatric populations, and their specificity for particular headache types may limit utility in mixed or uncertain diagnoses [3,10,11,27].

Compared to these established markers, Nesfatin-1 and Ghrelin/GOAT offer potential advantages: they can be measured in both blood and saliva (which is essential for pediatric populations, where minimally invasive sampling is preferred) [13,16,20], and they may reflect broader pain and stress mechanisms relevant to multiple headache types, rather than being specific to migraine alone [1-5]. However, without receiver operating characteristic (ROC) curve analysis, we cannot determine diagnostic accuracy parameters (sensitivity, specificity, optimal cutoffs) needed for clinical implementation.

Gender-Specific Clinical Manifestations

Our data revealed prominent gender-specific differences, with palpation-induced muscular discomfort occurring significantly more frequently in females (74%) than in males (38%; $p < 0.016$). In contrast, males demonstrated elevated levels of psychological distress (90% vs 48%, $p < 0.003$) and were more frequently asymptomatic for TMD (66% vs 30%, $p < 0.016$). These gender disparities align with previous research demonstrating higher TMD prevalence in females [28, 29], possibly due to hormonal influences, different pain thresholds, and gender-specific stress responses. The paradoxical observation of increased psychological distress in boys, despite lower incidences of TMD symptoms, indicates intricate gender-specific correlations between psychological factors and pain expression in adolescents.

Headache-TMD Comorbidity and Parafunctional Habits

A correlation between TMD and headache was apparent in our population, with 52% of patients with headache exhibiting TMD symptoms. This association corroborates the findings of Nilsen et al. regarding the strong correlation between TMD pain and headache in adolescents [30]. Recent Polish studies have also documented high prevalence and significant overlap between headaches and TMD, with headache being more frequent among those with painful TMD [7, 31]. Liljestrom et al. also reported connections between TMD and headaches in teenagers, urging the diagnosis of TMD when headaches are accompanied by ear pain, difficulty in mouth opening, and tiredness [32]. Our data support these suggestions, indicating that a thorough evaluation of TMD should be included in the standard assessment for adolescents presenting with headaches.

The lack of significant differences in Nesfatin-1 and Ghrelin/GOAT levels between patients with migraine and those with TTH is particularly noteworthy. This finding suggests that changes in these neuroendocrine markers may reflect overarching pain mechanisms rather than headache-specific pathways. Similarly, the lack of significant differences in these markers across TMD subgroups (myalgia, DDwR, no TMD) supports their potential as general biomarkers of pain and/or stress rather than as indicators of specific diseases.

Parafunctional habits were prevalent in our study group, with both bite and non-bite parafunctions observed in more than 80% of cases. This prevalence is more than the reported ranges in literature (5.9-49.6%) [33, 34] and highlights the possible contribution of these activities to the onset and persistence of TMD and headache symptoms. The elevated incidence of physical symptoms in individuals with myalgia (89%) and disc displacement with a decrease (100%) relative to those without TMD (47%) underscores the clinical importance of these conditions and their effect on adolescents' quality of life [34]. Recent evidence from Polish TMD populations further emphasises the complex relationships between sleep bruxism, reported pain, headache, and various health factors in patients with temporomandibular disorders [31].

Study Limitations

Several limitations of our study should be acknowledged. The limited sample size, particularly within specific subgroups, restricts the generalizability of our results. Subgroup analyses are statistically underpowered (e.g., DDwR, n = 5) and should be considered exploratory rather than confirmatory. The cross-sectional approach inhibits our ability to establish causality or ascertain whether the identified neuroendocrine abnormalities are primary factors or indirect effects of headache and TMD.

There is a slight difference in age ranges between the patient (10-17 years) and control (12-17 years) groups. However, it did not result in significant age differences in the final recruited cohorts, which warrants consideration of uniform age criteria in future studies.

Although TMD examinations were conducted according to standardised DC/TMD protocols, we did not formally assess inter-rater reliability between examiners, which would strengthen diagnostic accuracy in future studies.

We did not perform formal correlation analysis between blood and salivary biomarker levels. The opposing directional changes observed (Ghrelin/GOAT elevated in blood but decreased in saliva) indicate differential regulatory mechanisms in distinct biological compartments rather than simple linear relationships. Future mechanistic studies examining production, degradation, and transport in different tissues would better elucidate these tissue-specific patterns than correlation analysis.

We did not apply multiple-comparison corrections (e.g., Bonferroni or Benjamini-Hochberg) to our statistical analyses. Our primary comparisons (patients vs. controls for four biomarkers) were pre-specified, and the observed effect sizes were extensive (Cohen's d = 2.3-3.4), making false-positive findings highly unlikely. Subgroup analyses (headache type, TMD categories, gender) were clearly identified as exploratory. Applying strict multiple comparison corrections in this context would be overly conservative and risk false-negative conclusions in this hypothesis--generating study.

Moreover, although we controlled for several confounding factors (BMI via exclusion criteria; age, sex, and basic biochemical parameters via group balancing and stratification), our sample size (n = 44) precluded multivariable regression analyses that would have allowed simultaneous adjustment for multiple confounders. Future studies with larger cohorts are needed to perform multivariable modelling incorporating stress, parafunctional habits, TMD status, and headache type. Additionally, we did not assess pubertal stage (Tanner staging), which may influence neuroendocrine marker levels and represents a significant biological variable in adolescent research.

Clinical Implications and Future Directions

Future research should focus on longitudinal studies to investigate the temporal relationship between neuroendocrine indicators and headache/TMD symptoms, as well as the potential predictive value of these markers for treatment efficacy. Investigating the processes governing the tissue-specific patterns of Ghrelin/GOAT may yield significant insights into the pathophysiology of headache diseases. Ultimately, interventional studies focusing on these neuroendocrine pathways may provide novel therapeutic strategies for adolescents suffering from headaches and TMD.

Conclusion

This cross-sectional study demonstrates consistent elevations in Nesfatin-1 and contrasting patterns of Ghrelin/GOAT expression (elevated

in blood, reduced in saliva) in adolescents with headaches compared to healthy controls, independent of specific headache type or TMD status. These findings suggest that neuroendocrine dysregulation may represent a common pathway in adolescent headache rather than being specific to particular diagnostic subtypes.

In conclusion, our findings reveal substantial changes in Nesfatin-1 and Ghrelin/GOAT levels in adolescents experiencing headaches, irrespective of headache classification or TMD status. The persistent increase in these markers across various clinical manifestations indicates their potential utility as biomarkers in adolescent headache disorders. The gender-specific disparities in TMD symptoms and psychological suffering underscore the necessity of treating sex as a biological variable in pain studies. These findings enhance our comprehension of the intricate interactions among neuroendocrine variables, psychological distress, and pain perception in adolescents experiencing headaches and TMD.

Declarations

Authors' contributions

Conceptualisation: M.Z., B.S., B.D.B., A.S.: data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software; A.S., Y.B., M.Z.; supervision: M.Z., B.S., B.D.B.; validation, visualisation, writing: A.S., Y.B., O.Y., M.Z. All authors have read and approved the final version of the manuscript.

Conflicts of interest

Authors disclose any financial and personal relationships with others or organisations that could inappropriately influence the work.

Human ethics and consent to participate

The study was conducted in accordance with the Declaration of Helsinki and received ethical approval from the Bioethical Committee of Poznan University of Medical Sciences (approval number 550/17). For child participants, written informed consent was obtained from their parents or legal guardians, and assent was obtained from the children themselves after a full explanation of the study procedures. Participant anonymity and confidentiality were maintained throughout the study. Participation was voluntary, with no financial or other obligations imposed on the participants. All procedures were designed to minimise any potential burden or risk to the participants.

Ethics approval declaration

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Bioethical Committee of Poznan University of Medical Sciences (approval number 550/17).

Funding declaration

This research received no external funding.

Clinical trial number

Not applicable.

References

1. Dao JM, Qubty W. Headache Diagnosis in Children and Adolescents. Curr Pain Headache Rep. 2018;22(3):17. https://doi.org/10.1007/s11916-018-0675-7.
2. Wöber-Bingöl C. Epidemiology of migraine and headache in children and adolescents. Curr Pain Headache Rep. 2013;17(6):341. https://doi.org/10.1007/s11916-013-0341-z.
3. Abu-Arafeh I, Razak S, Sivaraman B, Graham C. Prevalence of headache and migraine in children and adolescents: a systematic review of population-based studies. Dev Med Child Neurol. 2010;52(12):1088-97. https://doi.org/10.1111/j.1469-8749.2010.03793.x.
4. Hershey AD, Powers SW, Vockell AL, LeCates SL, Segers A, Kabbouche MA. Development of a patient-based grading scale for PedMIDAS. Cephalalgia. 2004;24(10):844-9. https://doi.org/10.1111/j.1468-2982.2004.00757.x.
5. Hershey AD. Current approaches to the diagnosis and management of paediatric migraine. Lancet Neurol. 2010;9(2):190-204. https://doi.org/10.1016/s1474-4422(09)70303-5.
6. Gauer RL, Semidey MJ. Diagnosis and treatment of temporomandibular disorders. Am Fam Physician. 2015;91(6):378-86. PMID: 25822556.
7. Wieckiewicz M, Grychowska N, Nahajowski M, Hnitecka S, Kempiak K, Charemska K, et al. Prevalence and Overlaps of Headaches and Pain-Related Temporomandibular Disorders Among the Polish Urban Population. Journal of Oral & Facial Pain and Headache. 2020;34(1):31-9. https://doi.org/10.11607/ofph.2386.
8. Błaszczyk B, Waliszewska-Prosół M, Smardz J, Więckiewicz M, Wojakowska A, Martynowicz H. Exploring the associations of sleep bruxism and obstructive sleep apnea with migraine among patients with temporomandibular disorder: A polysomnographic study. Headache: The Journal of Head and Face Pain. 2025;65(2):242-57. https://doi.org/10.1111/head.14892.
9. The International Classification of Headache Disorders, 3rd edition (beta version). Cephalalgia. 2013;33(9):629-808. https://doi.org/10.1177/0333102413485658.

10. van Dongen RM, Zielman R, Noga M, Dekkers OM, Hankemeier T, van den Maagdenberg AM, et al. Migraine biomarkers in cerebrospinal fluid: A systematic review and meta-analysis. Cephalalgia. 2017;37(1):49-63. https://doi.org/10.1177/0333102415625614.
11. Loder E, Rizzoli P. Biomarkers in migraine: their promise, problems, and practical applications. Headache. 2006;46(7):1046-58. https://doi.org/10.1111/j.1526-4610.2006.00498.x.
12. Oh IS, Shimizu H, Satoh T, Okada S, Adachi S, Inoue K, et al. Identification of nesfatin-1 as a satiety molecule in the hypothalamus. Nature. 2006;443(7112):709-12. https://doi.org/10.1038/nature05162.
13. Aydin S, Dag E, Ozkan Y, Erman F, Dagli AF, Kilic N, et al. Nesfatin-1 and ghrelin levels in serum and saliva of epileptic patients: hormonal changes can have a major effect on seizure disorders. Mol Cell Biochem. 2009;328(1-2):49-56. https://doi.org/10.1007/s11010-009-0073-x.
14. Kojima M, Hosoda H, Date Y, Nakazato M, Matsuo H, Kangawa K. Ghrelin is a growth-hormone-releasing acylated peptide from stomach. Nature. 1999;402(6762):656-60. https://doi.org/10.1038/45230.
15. Kojima M, Hamamoto A, Sato T. Ghrelin O -acyltransferase (GOAT), a specific enzyme that modifies ghrelin with a medium-chain fatty acid. The Journal of Biochemistry. 2016;160(4):189-94. https://doi.org/10.1093/jb/mvw046.
16. Ozmen S, Şeker A, Demirci E. Ghrelin and leptin levels in children with anxiety disorders. J Pediatr Endocrinol Metab. 2019;32(10):1043-7. https://doi.org/10.1515/jpem-2019-0229.
17. Minervini G, Franco R, Marrapodi MM, Fiorillo L, Cervino G, Cicciù M. Prevalence of temporomandibular disorders in children and adolescents evaluated with Diagnostic Criteria for Temporomandibular Disorders: A systematic review with meta-analysis. J Oral Rehabil. 2023;50(6):522-30. https://doi.org/10.1111/joor.13446.
18. Nilsson IM, Drangsholt M, List T. Impact of temporomandibular disorder pain in adolescents: differences by age and gender. J Orofac Pain. 2009;23(2):115-22. PMID: 19492536.
19. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013;310(20):2191-4. https://doi.org/10.1001/jama.2013.281053.
20. Gutierrez-Castro K, Kornhauser-Araujo C, Gomez A, Luevano-Contreras C, Fafutis-Morris M, Gugliucci A, et al. Serum nesfatin-1 levels are higher in obese vs lean adolescents and associate with cardiometabolic dyslipidemia. Journal of Clinical and Molecular Medicine. 2020;4. https://doi.org/10.15761/JCMM.1000140.
21. Shiiya T, Nakazato M, Mizuta M, Date Y, Mondal MS, Tanaka M, et al. Plasma ghrelin levels in lean and obese humans and the effect of glucose on ghrelin secretion. J Clin Endocrinol Metab. 2002;87(1):240-4. https://doi.org/10.1210/jcem.87.1.8129.
22. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet JP, et al. Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: recommendations of the International RDC/TMD Consortium Network* and Orofacial Pain Special Interest Group†. J Oral Facial Pain Headache. 2014;28(1):6-27. https://doi.org/10.11607/jop.1151.
23. Ohrbach R, Knibbe W. Diagnostic criteria for temporomandibular disorders (DC/TMD) scoring manual for self-report instruments. NY, US and ACTA, Amsterdam, The Netherlands: University at Buffalo. 2018. Corpus ID: 45047162.
24. Zawawi KH, Al-Badawi EA, Lobo SL, Melis M, Mehta NR. An index for the measurement of normal maximum mouth opening. Journal (Canadian Dental Association). 2003;69(11):737-41.PMID: 14653940.
25. Hofmann T, Ahnis A, Elbelt U, Rose M, Klapp BF, Stengel A. NUCB2/nesfatin-1 Is Associated with Elevated Levels of Anxiety in Anorexia Nervosa. PLoS One. 2015;10(7):e0132058. https://doi.org/10.1371/journal.pone.0132058.
26. Xiao MM, Li JB, Jiang LL, Shao H, Wang BL. Plasma nesfatin-1 level is associated with severity of depression in Chinese depressive patients. BMC Psychiatry. 2018;18(1):88. https://doi.org/10.1186/s12888-018-1672-4.
27. Christensen CE, Ashina M, Amin FM. Calcitonin Gene -Related Peptide (CGRP) and Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) in Migraine Pathogenesis. Pharmaceuticals. 2022;15(10):1189. https://doi.org/10.3390/ph15101189
28. Pallegama RW, Ranasinghe AW, Weerasinghe VS, Sitheeque MA. Anxiety and personality traits in patients with muscle related temporomandibular disorders. J Oral Rehabil. 2005;32(10):701-7. https://doi.org/10.1111/j.1365-2842.2005.01503.x.
29. AKHTER R, MORITA M, ESAKI M, NAKAMURA K, KANEHIRA T. Development of temporomandibular disorder symptoms: a 3-year cohort study of university students. Journal of Oral Rehabilitation. 2011;38(6):395-403. https://doi.org/10.1111/j.1365-2842.2010.02195.x.
30. Nilsson IM, List T, Drangsholt M. Headache and co-morbid pains associated with TMD pain in adolescents. J Dent Res. 2013;92(9):802-7. https://doi.org/10.1177/0022034513496255.
31. Orzeszek S, Martynowicz H, Smardz J, Wojakowska A, Bombała W, Mazur G, et al. Assessment of sleep quality in patients with orofacial pain and headache complaints: A polysomnographic study. Dent Med Probl. 2024;61(4):549-62. https://doi.org/10.17219/dmp/177008.
32. Liljeström M-R, Le Bell Y, Anttila P, Aromaa M, Jämsä T, Metsähonkala L, et al. Headache Children with Temporomandibular Disorders have Several Types of Pain and other Symptoms. Cephalalgia. 2005;25(11):1054-60. https://doi.org/10.1111/j.1468-2982.2005.00957.x.
33. Machado E, Dal-Fabbro C, Cunali PA, Kaizer OB. Prevalence of sleep bruxism in children: a systematic review. Dental Press J Orthod. 2014;19(6):54-61. https://doi.org/10.1590/2176-9451.19.6.054-061.oar.
34. Manfredini D, Restrepo C, Diaz-Serrano K, Winocur E, Lobbezoo F. Prevalence of sleep bruxism in children: a systematic review of the literature. J Oral Rehabil. 2013;40(8):631-42. https://doi.org/10.1111/joor.12069.

ORIGINAL PAPER

JMS Journal of Medical Science

Passiflora incarnata extract-induced VEGF and TGF-β1 mRNA expression in the cardiovascular system using the rat model of type 2 diabetes

Anna Wesołowska
Poznan University of Medical Sciences, Department of Pharmacology, Poznań, Poland
https://orcid.org/0000-0001-5029-2265
Corresponding author: anna.wesolowska@ump.edu.pl

Marcin Ożarowski
Institute of Natural Fibres and Medicinal Plants – National Research Institute, Department of Biotechnology, Poznań, Poland
https://orcid.org/0000-0001-5881-2966

Przemysław Ł. Mikołajczak
Poznan University of Medical Sciences, Department of Pharmacology, Poznań, Poland
https://orcid.org/0000-0002-1252-6523

Agnieszka Stelmaszyk
Poznan University of Medical Sciences, Department of Pharmacology, Poznań, Poland
https://orcid.org/0000-0001-6091-8751

Anna Maliszewska-Dworacka
Collegium Da Vinci, Poznań, Poland
https://orcid.org/0009-0004-2044-399X

Saule Iskakova
Kazakh National Medical University, Department of Pharmacology, Almaty, Kazakhstan
https://orcid.org/0000-0001-7624-1570

Bartosz A. Frycz
Poznan University of Medical Sciences, Department of Biochemistry and Molecular Biology, Poznań, Poland
https://orcid.org/0000-0001-6729-1832

Paweł P. Jagodziński
Poznan University of Medical Sciences, Department of Biochemistry and Molecular Biology, Poznań, Poland
https://orcid.org/0000-0002-9046-6802

Marzena Dworacka
Poznan University of Medical Sciences, Department of Pharmacology, Poznań, Poland
https://orcid.org/0000-0002-4277-0615

https://doi.org/10.20883/medical.e1443

Keywords: *Passiflora incarnata*, vascular endothelial growth factor, vascular endothelial growth factor receptors, transforming growth factor β1, type 2 diabetes mellitus

Received 2025-10-01
Accepted 2025-12-01
Published 2025-12-31

How to Cite: *Passiflora incarnata* extract-induced VEGF and TGF-β1 mRNA expression in the cardiovascular system using the rat model of type 2 diabetes. Wesołowska A, Ożarowski M, Mikołajczak PŁ, Stelmaszyk A, Maliszewska-Dworacka A, Iskakova S, Frycz BA, Jagodziński PP, Dworacka M. Journal of Medical Science. 2025 December;94(4);e1443. doi:10.20883/medical.e1443

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Aim. To investigate the possible effects of *Passiflora* extract on VEGF and TGF-β1 mRNA expression – growth factors closely related to the development of diabetes in the cardiovascular system.
Material and methods. The animal model of type 2 diabetes, rich-fat/STZ Wistar rats, was used. Animals were randomised into four groups: *Passiflora*-treated type 2 diabetes mellitus model group; metformin-treated type 2 diabetes mellitus model group; placebo-treated type 2 diabetes mellitus model group; and placebo-treated non-diabetic control group. *Passiflora incarnata* leaf extract was administered orally once daily

for eight consecutive weeks. mRNA VEGF, VEGF-R1, VEGF-R2 and TGF-β1 expression were measured in the myocardium and the aorta.
Results. *Passiflora incarnata* extract increases VEGF and VEGFR2 mRNA expression in the myocardium of rats with type 2 diabetes and decreases expression in the aortal wall. The expression of TGF-β1 mRNA in both the myocardium and the aorta is reduced in *Passiflora incarnata*-treated rats with diabetes. Most observed effects are independent of the animals' current metabolic status.
Conclusions. The current data provide novel findings on the beneficial effects of *Passiflora* extract on the myocardium in the context of type 2 diabetes, potentially through mechanisms involving VEGF and TGF-β1. However, the significance of the impact of *Passiflora* extract on the aorta wall via VEGF and TGF-β1 is uncertain.

Introduction

It is undeniable that diabetes mellitus and its cardiovascular complications constitute a serious health problem due to the high frequency of occurrence and mortality rates [1]. Disturbed angiogenesis leads to the progression of atherosclerosis. It is widely believed that neovascularisation of the atherosclerotic plaque increases its susceptibility to damage, potentially leading to rupture [2]. On the other hand, collateral or compensatory arteriogenesis in response to occlusive ischaemia appears to be beneficial for patients with diabetes [3]. Overexpressed fibrosis is another process that is closely linked to cardiovascular dysfunction in diabetes [4]. In these two processes, growth factors and their receptors play a crucial role. This particularly concerns vascular endothelial growth factor (VEGF) and its receptors, as well as transforming growth factor-β1 (TGF-β1).

VEGF is a proangiogenic factor acting directly on vascular endothelial cells. VEGF exerts its effects primarily through binding to appropriate receptors: VEGF receptor 1 (VEGFR-1; fms-related tyrosine kinase 1 – Flt1) and VEGF receptor 2 (VEGFR-2; kinase insert domain receptor – KDR; foetal liver kinase 1 – Flk1). VEGFR-1 has approximately ten times higher affinity for VEGF than VEGFR-2, but its tyrosine kinase activity is relatively weaker, and thus the primary signal transducer in the endothelium is VEGFR-2 [5]. TGF-β1 activates a matrix-preserving phenotype in cardiac fibroblasts and is the central effector of myocardial fibrosis [6]. The role of TGF-β1 as a regulator of the vascular endothelium is complex; among other effects, it strongly induces VEGF production, thereby promoting angiogenesis and vascular remodelling [7]. Moreover, TGF-β1 affects atherosclerotic plaque formation by stimulating multiple pro-inflammatory factors that are important in the early phase of atherosclerotic lesion formation [6].

Passiflora incarnata possesses numerous pharmacological properties and contains flavonoids, maltol, cyanogenic glycosides and indole alkaloids [8]. In several experiments, *Passiflora incarnata* extracts exhibited potential effects for the treatment of anxiety, insomnia, hypertension, and epilepsy [9]. The positive influence of *Passiflora* species on the glycaemic profile has also been demonstrated in numerous experimental studies, including *Passiflora incarnata* leaves [10]. Some studies have focused on the anti-inflammatory and antioxidant activities of *Passiflora incarnata* [8], which may have cardioprotective effects. It's necessary to underline that glucose control is not the only therapeutic goal for patients with diabetes. Very essential, but highly complex to achieve, is to delay the onset of chronic diabetic complications, especially those affecting the cardiovascular system.

The current study aimed to investigate the possible effects of *Passiflora* extract on the mRNA expression of VEGF and TGF-β1– key growth factors that are closely related to the development of diabetic complications within the cardiovascular system.

Materials and methods

Animals

The animal experiment was performed in accordance with Directive 2010/63/EU of the European

Parliament and of the European Council and with Polish governmental regulations (2005.01.21). It was approved by the Local Ethics Committee on the Use of Laboratory Animals in Poznan, Poland (No. 54-55/2013).

The experiment was carried out on 51 8-week-old male Wistar rats. The animals were housed in individual cages in an environmentally controlled room at 22°C under a reversed 12:12 h light: dark cycle to acclimatise for 1 week. The animals were randomly assigned to one of four groups: *Passiflora*-treated type 2 diabetes mellitus model group (P-DM), n = 13; a metformin-treated type 2 diabetes mellitus model group (M-DM), n = 14; a placebo-treated type 2 diabetes mellitus model group (DM), n = 12; and a placebo-treated control group without diabetes mellitus (nonDM), n = 12.

Developing a model of type 2 diabetes mellitus

The type 2 diabetes model (low-dose-STZ-treated/fat-fed rats) was induced by an intraperitoneal injection of 20 mg/kg streptozotocin (STZ) (streptozotocin – Sigma-Aldrich, Saint Louis, USA) dissolved in a citrate buffer, followed by 4 weeks of feeding with a rich-fat diet (Labofeed B, 61% of fat) (groups P-DM, M-DM, DM). Animals in the non-diabetic control group (nonDM) received citrate buffer as a vehicle and commercial standard chow (Labofeed B standard, 2.8% fat) for the same duration. All animals received water *ad libitum*. Such a model was previously described to replicate the natural history and metabolic characteristics of type 2 diabetes mellitus in humans, including obesity, insulin resistance, and mildly impaired insulin secretion [11].

To confirm or exclude the development of diabetes, an intragastric glucose tolerance test (GTT) was performed. After 12 h of overnight fasting, the rats were given 1 g/kg b.w. of glucose dissolved in sterile water via intragastric gavage. The glucose concentration in the blood was measured using a glucometer at 0 and 1 h in samples obtained from superficial capillary vessels at the tip of the rats' tails. Rats showing a fasting glucose of 126 mg/dL (7.0 mmol/L) and/or a glucose concentration measured one hour after glucose solution administration of 140 mg/dL (7.8 mmol/L) were designated for the study. Rats with glucose levels > 200 mg/dL (11.1 mmol/L) and/or with clinical symptoms of hyperglycaemia were excluded from the study.

Study design

The main experiment lasted for 8 weeks. During this period, rats in the P-DM group received *Passiflora* extract (200 mg/kg) [12] via intragastric gavage, and rats in the M-DM group received 300 mg/kg of metformin (Siofor®, Berlin Chemie) suspended in 1 % methylcellulose solution. Participants in the DM and non-DM groups were administered an equivalent volume of 1% methylcellulose as a placebo. All rats were treated similarly with respect to daily handling. At the end of the experiment, the animals were euthanised by decapitation. Samples of blood and selected tissues, including heart muscle and aortic wall, were collected and stored for further analysis.

Plant material

Leaves of *Passiflora incarnata L*. were obtained from plants grown in the greenhouse of the Botanical Garden, Poznan University of Medical Sciences (Poland). The material was identified by the Department of Medicinal and Cosmetic Natural Products, Faculty of Pharmacy, Poznan University of Medical Sciences. The following conditions for cultivated plants were: temperature (25–40 °C), humidity (60–70%), and a peat substrate in pots. The voucher specimen has been deposited in the Herbarium of the Institute of Natural Fibres and Medicinal Plants in Poznan, Poland.

Preparation of the extracts

Healthy leaves after drying, comminuting (0.5kg), and portioning (5 portions, 100g each) were extracted with pure methanol (Sigma-Aldrich, Poznan, Poland) (1:10, m/V) three times (each part) for one hour by reflux using a rotavapor. All collected extracts were combined and concentrated under vacuum to remove methanol. The yield of the dry extract was 22% for *P. incarnata*. The extraction conditions and efficiency were the same as in our previous studies [13]. The analysis performed at that time identified 52 chemical compounds (C-glycosides, O-glycosides, phenolic acids, and others). The most representative metabolites were glycosides of apigenin, luteolin, vitexin, isovitexin, isoorientin, chrysin and quercetin [13].

Analysis

Metabolic parameters and angiogenic factors

The glucose concentration in the blood samples was measured using a glucometer (Diagnostic GOLD Strip/System; Diagnosis S. A.; Poland). The concentration of 1,5-anhydro-D-glucitol (1,5-AG), a retrospective marker of acute hyperglycaemia, was analysed using the enzymatic colourimetric method [14]. C-peptide in the serum was measured with a rat-specific ELISA (C-peptide rat ELISA, R&D Systems, Minneapolis, USA).

mRNA expression

Relative mRNA expression of VEGF, VEGF-R1, VEGF-R2, and TGF-β1 in the myocardium and aortic wall was determined by reverse transcription and real-time quantitative polymerase chain reaction (RQ-PCR). The tissue samples derived from the rats' heart muscles and aortas were stored at -80 °C until homogenisation in liquid nitrogen. Total RNA was isolated with Trizol reagent (Thermo Fisher Scientific, Waltham, USA) [15] and treated with DNase I (Roche, Mannheim, Germany). RNA quantity was assessed using a Nano-100 spectrophotometer (Hangzhou Allsheng, Zhejiang, China), and RNA integrity was evaluated by gel electrophoresis. Reverse transcription of 1 µg of RNA isolated from each rat tissue sample into cDNA was performed according to the manufacturer's protocol (Life Technologies, Carlsbad, USA). RQ-PCR was carried out in a Light Cycler480 II Real-time PCR detection system (Roche Diagnostics GmbH, Mannheim, Germany). SYBR® Green I was used as the detection dye. Transcript levels in tissues were expressed as the multiplicity of the cDNA concentration relative to the calibrator. The calibrator was prepared as a cDNA mix from all the rat samples. Successive calibrator dilutions were used to construct a standard curve and to estimate reaction efficiency, according to the manufacturer's instructions. For amplification, one µL of the cDNA solution was added to 9 µL of qPCR Master Mix (EURx, Gdansk, Poland) containing the following primer sequences: VEGF, 5′-CTCCACCATGCCAAGTGGTC-3′ and 5′-AATAGCTGCGCTGGTAGACG-3′; VEGFR-1, 5′-GCAGAGCCAGGAACATATAC-3′ and 5′-GAGGTTTTGAAGCAGGAGTG-3′; VEGFR-2, 5′-CGCGTTTTCAGAGTTGGTGG-3′ and 5′-TGAGGTAGGCAGGGAGAGTC-3′; TGF-β1 5′-CCATCTGTTGTTGTGCCTC-3′ and 5′-CAGTATGTGGGTTCAATTCC-3′. Primers for VEGF amplification were previously published [16,17], whereas the remaining primers were designed using OLIGO Primer Analysis Software (version 5.0; Molecular Biology Insights, Inc., Colorado Springs, USA). The number of transcripts for the analysed genes in each sample was standardised to the β-actin reference gene level.

Statistical analysis

Statistical calculations were performed with STATISTICA v. 13 on the licence owned by Poznan University of Medical Sciences. All data is shown as mean ± standard deviation and/or median. ANOVA or Kruskal-Wallis tests were used to compare the four groups with regular or non-normal distribution, respectively. Analysis of covariance (ANCOVA) was performed to adjust for the influence of metabolic factors on group differences. The correlation between variables was analysed using Spearman's rank correlation test. A p-value ≤ 0.05 was considered to be significant.

Results

Characteristics of the studied groups

There were no differences in initial and final body weight or serum C-peptide levels across groups. Fasting glycemia, glucose levels at 1 hour in the GTT and HbA1c blood concentrations were higher in all groups with diabetes than in the control group. In contrast, the plasma 1,5-AG level was lower in all diabetic groups than in the control group (see **Table 1**).

The effect of *Passiflora incarnata* extract on VEGF expression

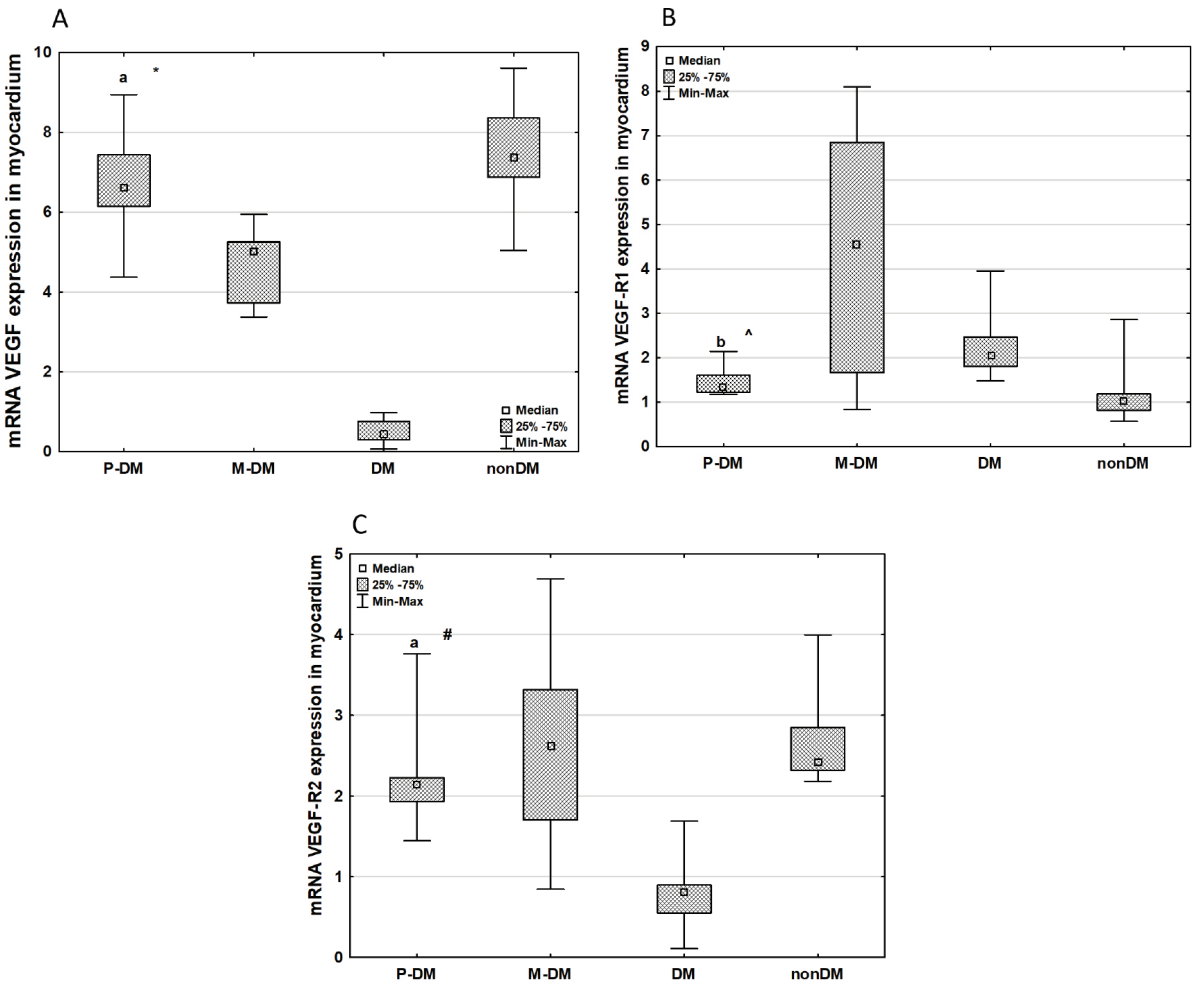
VEGF mRNA and VEGFR2 mRNA expression in the myocardium were higher in the P-DM group than in the DM group, while there were no differences in these two factors between the P-DM group and all the other groups (see **Figure 1A** and **1C**; tags for ANOVA or ANOVA Kruskal-Wallis test). VEGFR1 mRNA expression in the myocardium in the P-DM group was lower than in the M-DM group and similar to the other groups (see **Figure 1B**; tags for ANOVA or ANOVA Kruskal-Wallis test).

VEGF mRNA expression in the aorta wall in the P-DM group was higher than in the non-DM group,

Table 1. The general characteristics of the examined groups.

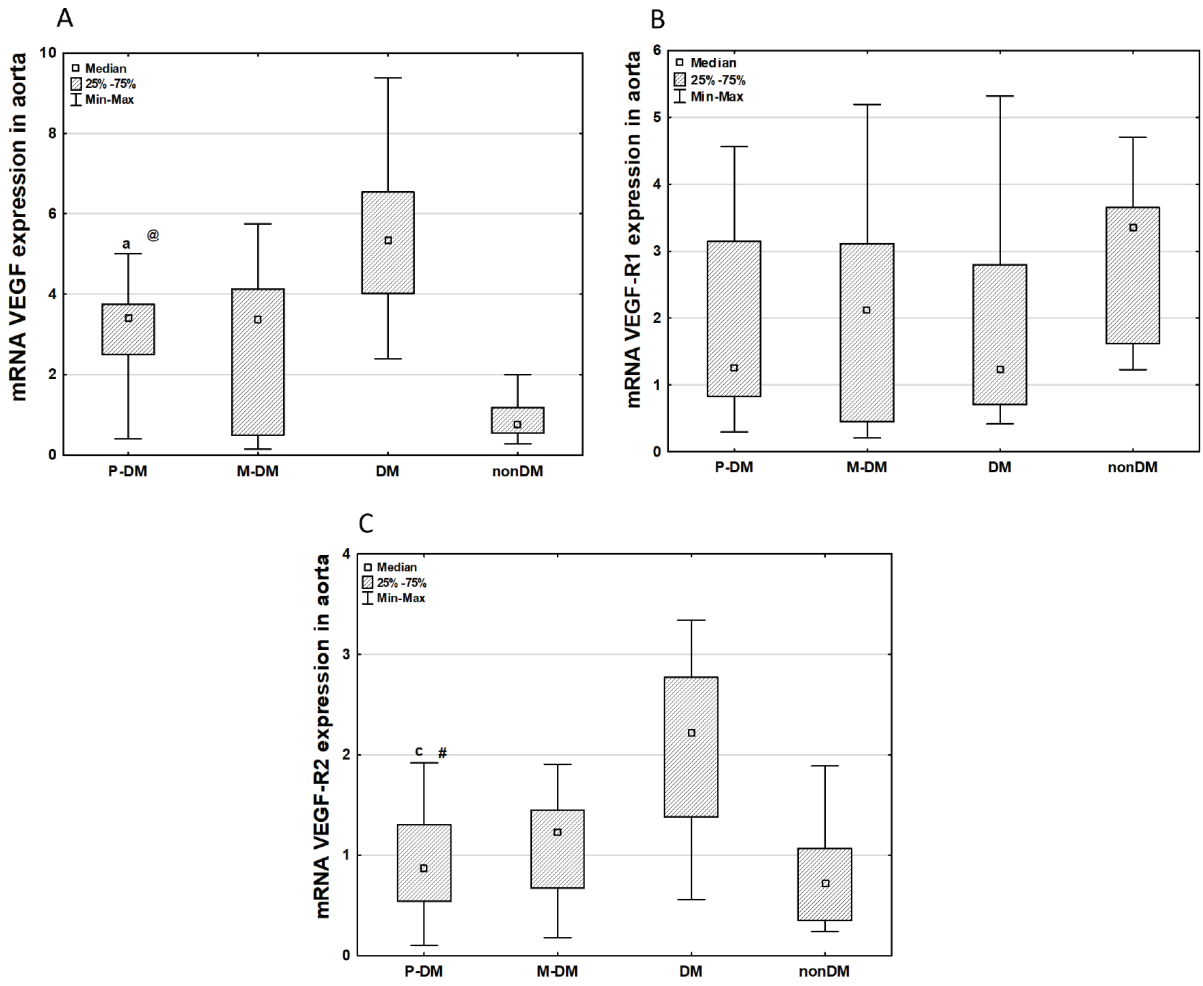
	P-DM n = 13	M-DM n = 14	DM n = 12	nonDM n = 12
Initial body weight [g]	230.00 ± 14.14 (230.00)	232.14 ± 17.62 (230.00)	238.33 ± 13.37 (235.00)	245.83 ± 11.65 (250.00)
Final body weight [g]	553.85 ± 50.75 (530.00)	541.43 ± 70.59 (569.00)	568.33 ± 44.48 (555.00)	569.17 ± 32.88 (570.00)
GTT 0h [mg/dL]	131.23 ± 19.28[a] (129.00)	136.21 ± 9.70[a] (135.00)	160.83 ± 34.34[a] (159.50)	94.00 ± 8.46 (93.50)
GTT 1h [mg/dL]	220.54 ± 94.18[a] (192.00)	230.29 ± 74.72[a] (206.50)	286.08 ± 98.13[a] (279.50)	116.67 ± 30.31 (110.50)
HbA_1c [%]	4.82 ± 0.84 (4.50)	4.51 ± 0.77 (4.60)	6.58 ± 0.56[b] (6,70)	3.47 ± 0.38[b] (3.65)
1,5-AG [µg/mL]	8.54 ± 0.55[a] (8.63)	9.02 ± 2.07[a] (8.42)	9.43 ± 2.02[a] (9.15)	15.68 ± 0.56 (15.79)
C-peptide [pmol/L]	511.65 ± 219.88 (507.79)	554.74 ± 208.22 (603.29)	452.04 ± 174.41 (425.90)	693.97 ± 479.55 (562.85)

Values are expressed as Mean ± SD (Median); a – statistically significant against the nonDM group; $p \leq 0.05$ (ANOVA Kruskal-Wallis test); b – statistically significant against the other groups; $p \leq 0.05$ (ANOVA Kruskal-Wallis test).



a – statistically significant against the DM group; b – statistically significant against the M-DM group; $p \leq 0.05$; ANOVA or ANOVA Kruskal-Wallis test for data with normal or non-normal distribution, respectively; * – statistically significant against the M-DM and DM groups; # – statistically significant against the DM group; ^ – statistically significant against the M-DM group; @ – statistically significant against the nonDM and DM groups; $p \leq 0.05$; ANCOVA, Scheffe test.

Figure 1. VEGF, VEGFR1, and VEGFR2 mRNA expression in the myocardium of the examined groups.



a – statistically significant against the DM group; c – statistically significant against the nonDM group; $p \leq 0.05$; ANOVA or ANOVA Kruskal-Wallis test for data with normal or non-normal distribution, respectively; * – statistically significant against the M-DM and DM groups; # – statistically significant against the DM group; ^ – statistically significant against the M-DM group; @ – statistically significant against the nonDM and DM groups; $p \leq 0.05$; ANCOVA, Scheffe test.

Figure 2. VEGF, VEGFR1, and VEGFR2 mRNA expression in the aorta of the examined groups.

and there were no differences between the P-DM, M-DM and DM groups (see **Figure 2A**; tags ANOVA or ANOVA Kruskal-Wallis test). No differences were found among the studied groups considering VEGFR1 mRNA expression in the aorta wall, while VEGFR2 expression in the P-DM group was lower compared to the DM group and similar to the M-DM and non-DM groups (see **Figure 2B** and **2C**; tags ANOVA or ANOVA Kruskal-Wallis test).

Taking into consideration the possible influence of some metabolic factors, such as HbA_1c, 1,5-AG or C-peptide, on VEGF, VEGFR1, and VEGFR2 expression, the analysis of covariance (ANCOVA) was performed. The ANCOVA results revealed that the therapeutic model (group) determined mRNA expression for all examined parameters, except for VEGFR1 mRNA in the aorta and VEGFR2 mRNA in the myocardium (see **Table 2**). However, the ANCOVA post hoc test results indicate that VEGF mRNA expression in the myocardium of the P-DM group is not only higher than in the DM group but also higher than in the M-DM group (see **Figure 1A**, **1B**, **1C**; ANCOVA tags). Simultaneously, VEGF mRNA expression in the aorta wall was not only lower in the P-DM group than in the DM group, but also higher in comparison to the non-DM group (see **Figure 2A**, **2B**, **2C**; tags for ANCOVA).

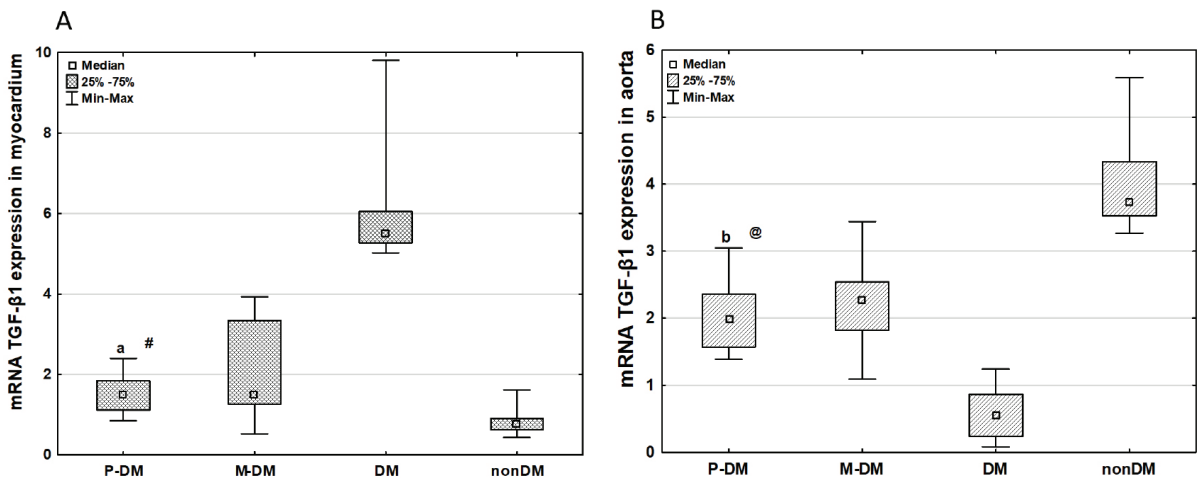
Table 2. Analysis of covariance (ANCOVA) results – the model with C-peptide,1,5-AG and HbA_1c as independent variables and VEGEF, VEGFR1, VEGFR2 and TGF-β1 mRNA expression In the myocardium or in the aorta wall as dependent variables.

F value for the model	R^2 value for the model	p for the model	independent variable which determines intergroup differences in the dependent variable
dependent variable: VEGF mRNA in myocardium			
46.38	0.84	0.001*	Group
dependent variable: VEGF mRNA in aorta			
8.26	0.47	0.000005*	Group
dependent variable: VEGFR1 mRNA in myocardium			
8.85	0.49	0.000002*	Group
dependent variable: VEGFR1 mRNA in aorta			
0.88	-0.01	0.51	None
dependent variable: VEGFR2 mRNA in myocardium			
9.74	0.52	0.000001*	HbA_1c
dependent variable: VEGFR2 mRNA in aorta			
5.07	0.33	0.0005*	Group
dependent variable: TGF-β1 mRNA in myocardium			
34.78	0.80	0.000001*	Group
dependent variable: TGF-β1 mRNA in aorta			
35.50	0.80	0.000001*	Group

*statistically significant

The effect of *Passiflora incarnata* extract on mRNA expression of TGF-β1 in the aorta wall and in the myocardium

TGF-β1 mRNA expression in the myocardium of the P-DM group was lower than in the DM group and similar to the other groups, while expression of TGF-β1 mRNA in the aorta wall of the P-DM group was also higher compared to the DM group and lower than observed in the non-DM group (see **Figure 3A**, **3B**; tags for ANOVA or ANOVA Kruskal-Wallis test). For both TGF-β1 mRNA expression in the myocardium and in the aorta wall, ANCOVA confirmed that the differences among the groups are determined only by the examined group, and the post-hoc tests proved the same differences among the groups as the ANOVA tests (see **Figure 3A**, **3B**; tags for ANCOVA).



a – statistically significant against the DM group; b – statistically significant against the DM and non-DM groups; $p \leq 0.05$; ANOVA or ANOVA Kruskal-Wallis test for data with normal or non-normal distribution, respectively; * – statistically significant against the M-DM and DM groups; # – statistically significant against the DM group; ^ – statistically significant against the M-DM group; @ – statistically significant against the non-DM and DM groups; $p \leq 0.05$; ANCOVA, Scheffe test.

Figure 3. TGF-β1 mRNA expression in the myocardium and the aorta wall of the examined groups.

The relationship between metabolic control in the course of type 2 diabetes and angiogenic factors

HbA_1c was correlated with TGF-β1 mRNA expression in the myocardium [r = 0.7] and in the aorta [(-0.76)], with VEGF mRNA expression in the myocardium [(-0.64)] and with VEGFR2 mRNA expression in the myocardium [(-0.67)]. There were no correlations between plasma 1,5-AG levels or serum C-peptide levels and the analysed growth factors.

Discussion

In recent years, there has been growing interest in plant-derived substances for the treatment of type 2 diabetes mellitus, as these compounds may offer additional benefits beyond standard therapeutic regimens with a relatively low risk of adverse effects.

It is also well established that effective management of diabetes should go beyond metabolic control, aiming not only to regulate glycaemia but also to provide additional benefits that help reduce the risk of long-term cardiovascular complications [18].

Both angiogenesis and fibrosis are processes closely related to atherosclerosis and can modify cardiovascular function [19,20]. However, the relationship between pharmacological treatment for patients with diabetes and these two processes remains poorly established, particularly for plant-derived substances. A promising source of new medicinal drugs may be species from the *Passifloraceae* family [21].

It is well established that diabetes interferes with angiogenesis in the heart muscle, increasing its susceptibility to hypoxia, creating conditions favourable for the development of ischaemic heart disease and raising the risk of myocardial infarction [22]. In the *Passiflora* extract-treated animals with type 2 diabetes, VEGF mRNA expression in the myocardium was higher than in the placebo-treated rats with diabetes; moreover, VEGF mRNA expression in the myocardium in the *Passiflora* extract-treated animals did not differ from that of the non-diabetic control group. These observations suggest that *Passiflora* extract may stimulate angiogenesis, which plays a protective role for the heart muscle affected by metabolic derangement. Interestingly, in *Passiflora* extract-treated rats, VEGF mRNA expression in the myocardium was even higher than in metformin-treated animals with type 2 diabetes. The metformin-treated group in our study served as an additional control to compare the pharmacological effects of *Passiflora* extract with those induced by the standard antidiabetic agent. The same conclusion supports the observation of VEGFR2 mRNA expression in the myocardium of *Passiflora* extract-treated animals because the main effect of VEGFR2 stimulation by VEGF is the induction of angiogenesis [23].

Angiogenesis in the course of diabetes involves two opposing angiogenic tissue-specific conditions. In the myocardium under diabetic conditions, angiogenesis plays a crucial role in collateral vessel growth [24,25]. It is worth emphasising that a very similar observation was made by Sasso et al. [26], who also revealed an increase in VEGF mRNA expression in the myocardium of animals with induced type 2 diabetes. Our results are noteworthy, given that TGF-β1 mRNA expression in the myocardium of animals treated with *Passiflora* extract was lower than in placebo-treated diabetic rats. Liu et al. [27] and Jesmin et al. [28] also observed an increase in myocardial TGF-β1 mRNA expression in the course of diabetes, just as it was found in our experiment. Increased expression of TGF-β1 in the myocardium is considered to initiate the process of fibrosis and, consequently, the loss of contractile function of the heart muscle. It increases extracellular matrix synthesis and reduces its degradation by inhibiting proteinases. Consequently, excessive TGF-β1 activity in the heart muscle appears unfavourable [7]. It cannot be overlooked, however, that TGF-β1 in the cardiovascular system may promote both pro- and anti-angiogenic effects, and these effects are highly context-dependent [29].

VEGF activity has been confirmed in aortic endothelial cells, foam cells, and smooth muscle cells that cover the plaque, demonstrating a crucial role for this molecule in the progression of atherosclerotic lesions [30]. VEGF functions as a pro-angiogenic and pro-inflammatory factor within the vascular wall, and newly formed vessels constitute the primary route for inflammatory cells to reach the atherosclerotic plaque [31]. The ingrowth of vessels within the atherosclerot-

ic plaque may rupture it and, consequently, accelerate the clinical progression of coronary artery disease [32]. VEGF mRNA expression in the aorta wall of the rats with type 2 diabetes was significantly lower than in the placebo-treated animals with diabetes. Simultaneously, it was higher than in the placebo-treated non-diabetic rats. At the same time, the expression of VEGFR2 mRNA in the *Passiflora* extract-treated group was significantly lower in comparison to the placebo-treated diabetic animals.

The role of TGF-β1 in the development of atherosclerotic lesions in the aortic wall appears complex. According to previous research, TGF-β1 expression in the aorta was reduced in the early stages of atherosclerotic lesions (stages I–II according to the American Heart Association). In contrast, in advanced stages (III and VI), TGF-β1 expression was increased to levels higher than in healthy vessels [30]. TGF-β signalling may contribute to the development of atherosclerosis. Still, it promotes a stable lesion phenotype by stimulating smooth muscle cell differentiation and preventing the switch from contractile to proliferative smooth muscle cells [29].

On the other hand, TGF-β1 modulates the inflammatory components of the early atherosclerotic lesion [29,33]. In our study, groups of animals with early-phase type 2 diabetes do not develop atherosclerosis; therefore, the lower TGF-β1 mRNA expression in the aortic wall than in placebo-treated rats with diabetes is not associated with limited fibrosis, which does not play a significant role at this stage of diabetes-induced vasculature impairment. Instead, it appears to be linked to anti-inflammatory effects in the aortic wall, potentially playing an important role in protecting against the progression of atherosclerosis in diabetes.

It should be emphasised that most of the effects of *Passiflora* extract on the analysed growth factors were independent of the current metabolic status (ANCOVA test) expressed by HbA_1c, 1,5-AG and C-peptide concentrations in the serum. The only exception was VEGR2 mRNA expression in the myocardium, as HbA1c values accounted for the observed differences among groups. This observation appears particularly important because it is well established that the expression of VEGF and TGF-β1 is modified by hyperglycaemia [5,6].

In summary, the current data provide novel findings on the beneficial effects of *Passiflora* extract on the myocardium in the context of type 2 diabetes, potentially mediated by VEGF and TGF-β1. However, the precise significance of *Passiflora* extract's influence on the aortic wall mediated by these factors remains uncertain. To our knowledge, *Passiflora incarnata* extract has been first documented to be beneficial in inhibiting or delaying cardiovascular fibrosis and promoting angiogenesis in the heart muscle, which, in turn, may positively affect chronic cardiovascular complications associated with diabetes.

Declarations

Statement on the welfare of animals

All procedures and protocols adhered to the Guide for the Care and Use of Laboratory Animals from the Institute for Laboratory Animal Research, National Research Council, Washington, D.C., National Academy Press, 2011. Procedures and protocols were in accordance with Polish governmental regulations (2005.01.21) and were approved by the Local Ethics Committee on the Use of Laboratory Animals in Poznan, Poland (permissions 54-55/2013).

Acknowledgements

Poznan University of Medical Sciences supported this work [grant number 502-01-03310421-02598]. The sponsor was not involved in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Conflicts of interest

None.

Statement of individual author contributions

Anna Wesołowska: study design, laboratory experiments, data analysis and interpretation, preparation of the first draft of the manuscript; **Marcin Ożarowski**: acquisition of research material, laboratory experiments, manuscript revision; **Przemysław Ł. Mikołajczak**: scientific supervision, expert consultation, manuscript revision; **Agnieszka Stelmaszyk**: laboratory experiments, molecular analysis, participation in data analysis and interpretation; **Anna Maliszewska-Dworacka**: data analysis and interpretation; **Saule Iskakova**:

study design, participation in results interpretation; **Bartosz A. Frycz**: laboratory experiments, molecular analysis; **Paweł P. Jagodziński**: scientific and expert consultation; **Marzena Dworacka**: study conception and design, data analysis and interpretation, scientific supervision, expert consultation, revision and final approval of the manuscript.

Final approval

All authors confirm that they have read and approved the final version of the manuscript before its submission to the journal.

References

1. Raghavan S, Vassy JL, Ho YL, Song RJ, Gagnon DR, Cho K, Wilson PWF, Phillips LS. Diabetes Mellitus-Related All-Cause and Cardiovascular Mortality in a National Cohort of Adults. J Am Heart Assoc. 2019 Feb 19;8(4):e011295. https://doi.org/10.1161/JAHA.118.011295. PMID: 30776949; PMCID: PMC6405678.
2. Michel JB, Virmani R, Arbustini E, Pasterkamp G. Intraplaque haemorrhages as the trigger of plaque vulnerability. Eur Heart J. 2011 Aug;32(16):1977-85, 1985a, 1985b, 1985c. https://doi.org/10.1093/eurheartj/ehr054. Epub 2011 Mar 12. PMID: 21398643; PMCID: PMC3155759.
3. Attanasio S, Schaer G. Therapeutic angiogenesis for the management of refractory angina: current concepts. Cardiovasc Ther. 2011 Dec;29(6):e1-e11. https://doi.org/10.1111/j.1755-5922.2010.00153.x. Epub 2010 Apr 9. PMID: 20406245.
4. Tuleta I, Frangogiannis NG. Fibrosis of the diabetic heart: Clinical significance, molecular mechanisms, and therapeutic opportunities. Adv Drug Deliv Rev. 2021 Sep;176:113904. https://doi.org/10.1016/j.addr.2021.113904. Epub 2021 Jul 29. PMID: 34331987; PMCID: PMC8444077.
5. Pérez-Gutiérrez L, Ferrara N. Biology and therapeutic targeting of vascular endothelial growth factor A. Nat Rev Mol Cell Biol. 2023 Nov;24(11):816-834. https://doi.org/10.1038/s41580-023-00631-w. Epub 2023 Jul 25. PMID: 37491579.
6. Frangogiannis N. Transforming growth factor-β in tissue fibrosis. J Exp Med. 2020 Feb 13;217(3):e20190103. https://doi.org/10.1084/jem.20190103. PMID: 32997468; PMCID: PMC7062524.
7. Dobaczewski M, Chen W, Frangogiannis NG. Transforming growth factor (TGF)-β signaling in cardiac remodeling. J Mol Cell Cardiol. 2011 Oct;51(4):600-6. https://doi.org/10.1016/j.yjmcc.2010.10.033. Epub 2010 Nov 6. PMID: 21059352; PMCID: PMC3072437.
8. Kim M, Lim HS, Lee HH, Kim TH. Role Identification of Passiflora Incarnata Linnaeus: A Mini Review. J Menopausal Med. 2017 Dec;23(3):156-159. https://doi.org/10.6118/jmm.2017.23.3.156. Epub 2017 Dec 29. PMID: 29354614; PMCID: PMC5770524.
9. Miroddi M, Calapai G, Navarra M, Minciullo PL, Gangemi S. Passiflora incarnata L.: ethnopharmacology, clinical application, safety and evaluation of clinical trials. J Ethnopharmacol. 2013 Dec 12;150(3):791-804. https://doi.org/10.1016/j.jep.2013.09.047. Epub 2013 Oct 17. PMID: 24140586.
10. Gupta RK, Kumar D, Chaudhary AK, Maithani M, Singh R. Antidiabetic activity of Passiflora incarnata Linn. in streptozotocin-induced diabetes in mice. J Ethnopharmacol. 2012 Feb 15;139(3):801-6. https://doi.org/10.1016/j.jep.2011.12.021. Epub 2011 Dec 28. PMID: 22212504.
11. Srinivasan K, Viswanad B, Asrat L, Kaul CL, Ramarao P. Combination of high-fat diet-fed and low-dose streptozotocin-treated rat: a model for type 2 diabetes and pharmacological screening. Pharmacol Res. 2005 Oct;52(4):313-20. https://doi.org/10.1016/j.phrs.2005.05.004. PMID: 15979893.
12. Suciu F, Pușcașu C, Mihai DP, Ungurianu A, Andrei C, Ancuceanu RV, Gîrd CE, Ciobanu AM, Blebea NM, Popovici V, Ghi ă CIV, Negres S. Evaluation of the Antihyperalgesic Potential of Morus alba, Angelica archangelica, Valeriana officinalis, and Passiflora incarnata in Alloxan-Induced Diabetic Neuropathy in Rats. Curr Issues Mol Biol. 2025 Sep 4;47(9):719. https://doi.org/10.3390/cimb47090719. PMID: 41020842; PMCID: PMC12468037.
13. Ozarowski M, Piasecka A, Paszel-Jaworska A, de A. Chaves D. S, Romaniuk A, Rybczynska M, Gryszczynska A, Sawikowska A, Kachlicki P, Mikolajczak P.Ł, Seremak-Mrozikiewicz A, Klejewski A, Thiem B. Comparison of bioactive compounds content in leaf extracts of Passiflora incarnata, P. caerulea and P. alata and in vitro cytotoxic potential on leukemia cell lines. Rev. Bras. Farmacogn. 28, 179–191 (2018). https://doi.org/10.1016/j.bjp.2018.01.006.
14. Dworacka M, Winiarska H. The application of plasma 1,5-anhydro-D-glucitol for monitoring type 2 diabetic patients. Dis Markers. 2005;21(3):127-32. https://doi.org/10.1155/2005/251068. PMID: 16276006; PMCID: PMC3850583.
15. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 1987 Apr;162(1):156-9. https://doi.org/10.1006/abio.1987.9999. PMID: 2440339.
16. Tham E, Gielen AW, Khademi M, Martin C, Piehl F. Decreased expression of VEGF-A in rat experimental autoimmune encephalomyelitis and in cerebrospinal fluid mononuclear cells from patients with multiple sclerosis. Scand J Immunol. 2006 Dec;64(6):609-22. https://doi.org/10.1111/j.1365-3083.2006.01851.x. Erratum in: Scand J Immunol. 2007 Mar;65(3):310. PMID: 17083617.
17. van Albada ME, du Marchie Sarvaas GJ, Koster J, Houwertjes MC, Berger RM, Schoemaker RG. Effects of erythropoietin on advanced pulmonary vascular remodelling. Eur Respir J. 2008 Jan;31(1):126-34. https://doi.org/10.1183/09031936.00035607. Epub 2007 Sep 26. PMID: 17898019.

18. American Diabetes Association Professional Practice Committee. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes-2022. Diabetes Care. 2022 Jan 1;45(Suppl 1):S125-S143. https://doi.org/10.2337/dc22-S009. PMID: 34964831.
19. Perrotta P, Emini Veseli B, Van der Veken B, Roth L, Martinet W, De Meyer GRY. Pharmacological strategies to inhibit intra-plaque angiogenesis in atherosclerosis. Vascul Pharmacol. 2019 Jan;112:72-78. https://doi.org/10.1016/j.vph.2018.06.014. Epub 2018 Jun 19. PMID: 29933080.
20. Lan TH, Huang XQ, Tan HM. Vascular fibrosis in atherosclerosis. Cardiovasc Pathol. 2013 Sep-Oct;22(5):401-7. https://doi.org/10.1016/j.carpath.2013.01.003. Epub 2013 Jan 30. PMID: 23375582.
21. Ożarowski M, Karpiński TM. Extracts and Flavonoids of Passiflora Species as Promising Anti-inflammatory and Antioxidant Substances. Curr Pharm Des. 2021;27(22):2582-2604. https://doi.org/10.2174/1381612826666200526150113. PMID: 32452323.
22. Chou E, Suzuma I, Way KJ, Opland D, Clermont AC, Naruse K, Suzuma K, Bowling NL, Vlahos CJ, Aiello LP, King GL. Decreased cardiac expression of vascular endothelial growth factor and its receptors in insulin-resistant and diabetic States: a possible explanation for impaired collateral formation in cardiac tissue. Circulation. 2002 Jan 22;105(3):373-9. https://doi.org/10.1161/hc0302.102143. PMID: 11804995.
23. Shibuya M. VEGF-VEGFR Signals in Health and Disease. Biomol Ther (Seoul). 2014 Jan;22(1):1-9. https://doi.org/10.4062/biomolther.2013.113. PMID: 24596615; PMCID: PMC3936422.
24. Waltenberger J. New Horizons in Diabetes Therapy: The Angiogenesis Paradox in Diabetes: Description of the Problem and Presentation of a Unifying Hypothesis. Immun Endoc Metab Agents Med Chem. 2007;7:87-93. https://doi.org/10.2174/187152207779802536
25. Waltenberger J. VEGF resistance as a molecular basis to explain the angiogenesis paradox in diabetes mellitus. Biochem Soc Trans. 2009 Dec;37(Pt 6):1167-70. https://doi.org/10.1042/BST0371167. PMID: 19909240.
26. Sasso FC, Carbonara O, Persico E, D'Ambrosio R, Coppola L, Nasti R, Campana B, Moschella S, Torella R, Cozzolino D. Increased vascular endothelial growth factor mRNA expression in the heart of streptozotocin-induced diabetic rats. Metabolism. 2003 Jun;52(6):675-8. https://doi.org/10.1016/s0026-0495(03)00064-7. PMID: 12800090.
27. Liu C, Liu R, Fu H, Li J, Wang X, Cheng L, Korantzopoulos P, Tse G, Li G, Liu T. Pioglitazone attenuates atrial remodeling and vulnerability to atrial fibrillation in alloxan-induced diabetic rabbits. Cardiovasc Ther. 2017;35(5):3218–21.
28. Jesmin S, Sakuma I, Hattori Y, Fujii S, Kitabatake A. Long-acting calcium channel blocker benidipine suppresses expression of angiogenic growth factors and prevents cardiac remodelling in a Type II diabetic rat model. Diabetologia. 2002 Mar;45(3):402-15. https://doi.org/10.1007/s00125-001-0765-6. PMID: 11914746.
29. Goumans MJ, Ten Dijke P. TGF-β1 Signaling in Control of Cardiovascular Function. Cold Spring Harb Perspect Biol. 2018 Feb 1;10(2):a022210. https://doi.org/10.1101/cshperspect.a022210. PMID: 28348036; PMCID: PMC5793760.
30. Panutsopulos D, Papalambros E, Sigala F, Zafiropoulos A, Arvanitis DL, Spandidos DA. Protein and mRNA expression levels of VEGF-A and TGF-beta1 in different types of human coronary atherosclerotic lesions. Int J Mol Med. 2005;15:603–10.
31. Azimi-Nezhad M. Vascular endothelial growth factor from embryonic status to cardiovascular pathology. Rep Biochem Mol Biol. 2014 Apr;2(2):59-69. PMID: 26989723; PMCID: PMC4757048.
32. Jin X, Ge X, Zhu DL, Yan C, Chu YF, Chen WD, Liu J, Gao PJ. Expression and function of vascular endothelial growth factor receptors (Flt-1 and Flk-1) in vascular adventitial fibroblasts. J Mol Cell Cardiol. 2007 Sep;43(3):292-300. https://doi.org/10.1016/j.yjmcc.2007.06.002. Epub 2007 Jun 18. PMID: 17651752.
33. Kimura C, Konishi S, Hasegawa M, Oike M. Development of vascular smooth muscle contractility by endothelium-derived transforming growth factor β proteins. Pflugers Arch. 2014 Feb;466(2):369-80. https://doi.org/10.1007/s00424-013-1329-6. Epub 2013 Jul 26. PMID: 23887380.

REVIEW PAPER

JMS Journal of Medical Science

The Toolbox for Rating Diagnostic Tests: A Guide to Classification Metrics

Wiktoria Zasada
Department of Computer Science and Statistics, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0001-9329-3495
Corresponding author: 81576@student.ump.edu.pl

Przemysław Guzik
Department of Cardiology—Intensive Therapy, Poznan University of Medical Sciences, Poznań, Poland
University Centre for Sports and Medical Studies, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0001-9052-5027

Katarzyna B. Kubiak
Department of Computer Science and Statistics, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0002-1467-4853

Barbara Więckowska
Department of Computer Science and Statistics, Poznan University of Medical Sciences, Poznań, Poland
https://orcid.org/0000-0002-1811-2583

https://doi.org/10.20883/medical.e1474

Keywords: statistical analysis, binary classification, prediction model, diagnostic test, ROC curve, model evaluation

Received 2025-12-09
Accepted 2025-12-30
Published 2025-12-31

How to Cite: Zasada W, Guzik P, Kubiak KB, Więckowska B. The Toolbox for Rating Diagnostic Tests: A Guide to Classification Metrics. Journal of Medical Science. 2025 December;94(4);e1474. https://doi.org/10.20883/medical.e1474

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Evaluating a classifier's performance is critical for its successful application. This paper explores various metrics used for binary classification tasks, highlighting their strengths and limitations.
Simple threshold metrics, such as Accuracy and Sensitivity, are efficient for binary data and a single cutoff point. However, their reliance on a single threshold and sensitivity to imbalanced data can be drawbacks.
For more robust evaluation, ranking metrics such as Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves provide a threshold-agnostic approach, enabling comparison across different cutoff points. Additionally, probabilistic metrics like Brier Score and Log Loss assess the model's ability to predict class probabilities.
The choice of metric depends on the specific classification problem and the characteristics of the data. When dealing with imbalanced data or complex decision-making processes, using multiple metrics is recommended to gain a comprehensive understanding of the model's performance.
This paper emphasises the importance of understanding metric limitations and of selecting appropriate metrics for a specific classification task. By doing so, researchers and practitioners can ensure a more accurate and informative evaluation of their models, ultimately leading to the development of reliable tools for various applications.

Introduction

In today's data-driven world, accurately distinguishing between healthy and sick individuals is crucial across various sectors. Diagnostic tests play a vital role in medicine, public health, and research by enabling objective evaluation of patients and the diagnosis of conditions [1].

With challenges like the COVID-19 pandemic, there's a growing awareness of the need for accurate diagnostic tests and continual improvement [2,3].

There is a vast array of diagnostic tests (classifiers), each with its own set of quality metrics. Choosing the appropriate metric depends on several factors.

First, the disease prevalence must be considered. Is the disease common or rare in the population being tested? Screening tests for diseases with low prevalence, like mammograms for breast cancer, use different metrics than tests for suspected cases. Secondly, the type of variable measured by the test is crucial. Can the test result be a number (e.g., body temperature), an ordered category (e.g., mild pain), or something else (e.g., gender or smoking status) [4]? Thirdly, the purpose of the test must be determined. Are we simply classifying someone as healthy or sick, or are we trying to predict future health outcomes, like dead or alive?

Classification vs. Prediction: Two Sides of the Same Coin

Classification and prediction are closely linked. While we often think of classifying things in the present and predicting future events, the technical difference is not always clear-cut. Although it is possible to study attachments in many categories, we will focus on just two. We can classify both present ("sick" vs. "healthy") and future ("will get sick" vs. "will stay healthy") states. We can also estimate the likelihood of an event occurring now (sick vs. healthy) or in the future (becoming sick vs. remaining healthy). Studies use a "training set" of data to determine how to classify or predict future situations and then test this method on new data sets.

A perfect classifier would flawlessly assign people to the sick or healthy groups. A useless classifier would not distinguish between groups and would always guess randomly.

This review explores various tools scientists use to assess test quality, like Sensitivity, Specificity, Matthews Correlation Coefficient (MCC), Area Under the Receiver Operating Characteristic – AUC (ROC) curve, Brier Score, and more. We will discuss their strengths, weaknesses, and limitations for classifying and predicting health outcomes. We will also show how to interpret these matrices and compute them using widely used software such as R and Python.

The Power of Metrics: Assessing Classification Quality in Diverse Fields

Accurately evaluating diagnostic tests and predictive models enables a wide range of applications in healthcare. These tools play a crucial role in forecasting disease outbreaks [5,6], patient admissions, and treatment outcomes across various fields like epidemiology, general healthcare, and therapeutic interventions [7,8]. They also contribute significantly to clinical trials by aiding in participant selection, identifying patients at higher risk of complications, and assessing individual risk for chronic diseases [9,10]. Beyond trials, they assist in clinical practice by supporting disease diagnosis through patient data and biomarkers [11,12], personalising treatment plans and predicting their effectiveness [13,14], evaluating genetic disease risk and susceptibility to adverse drug reactions [15,16], and guiding preventive interventions [17,18].

While numerous metrics exist to evaluate the predictive capabilities of variables, algorithms, and models, choosing the most appropriate set can be challenging. Mathematicians and statisticians continually develop new metrics to better characterise the nuances of various tasks and their outcomes [19-21]. This review introduces and explains the most commonly used metrics in the health sciences, with particular emphasis on those designed explicitly for recent machine learning techniques.

Metrics to assess the quality of classification

A classifier categorises objects based on their characteristics. For example, it can classify patients into specific classes, categories, or groups. Classifiers can be simple, relying on a single variable for categorisation. Conversely, they can be complex, derived from models that consider multiple variables.

Evaluating classifiers and prediction models often involves multiple metrics. With dozens of

indicators available, selecting the appropriate one can be challenging. Ferri et al. categorise classification metrics into three groups: Threshold, Rank, and Probability Metrics (see **Figure 1**) [22].

Threshold metrics evaluate classification performance at a fixed threshold, such as the commonly used 6.5% cutoff for diagnosing diabetes using glycated haemoglobin A1c (HbA1c). If a patient's HbA1c is above this, they are classified as having diabetes.

Rank metrics assess how well a classifier ranks predictions. They consider the ordering of all scores, rather than a single cutoff, for example, in diagnosing diabetes using HbA1c, where sorted values of this variable can serve as consecutive cutoff points. Among the possible values, those above the cutoff point (6.5%) are generally considered indicative of diabetes [23]. Similarly, on the BIRADS (Breast Imaging Reporting and Data System) scale, sorted cut-off points range from 1 (negative) to 5 (high cancer probability) [24]. Conversely, probability metrics are used in models that calculate the exact likelihood of a disease or event, focusing on the specific value rather than a single cut-off or order.

In a nutshell, Threshold metrics are like on/off switches, ignoring prediction order. Rank metrics are like assigning grades (A, B, C), focusing on the order of predictions. Probability metrics are like percentages, considering both order and closeness to reality.

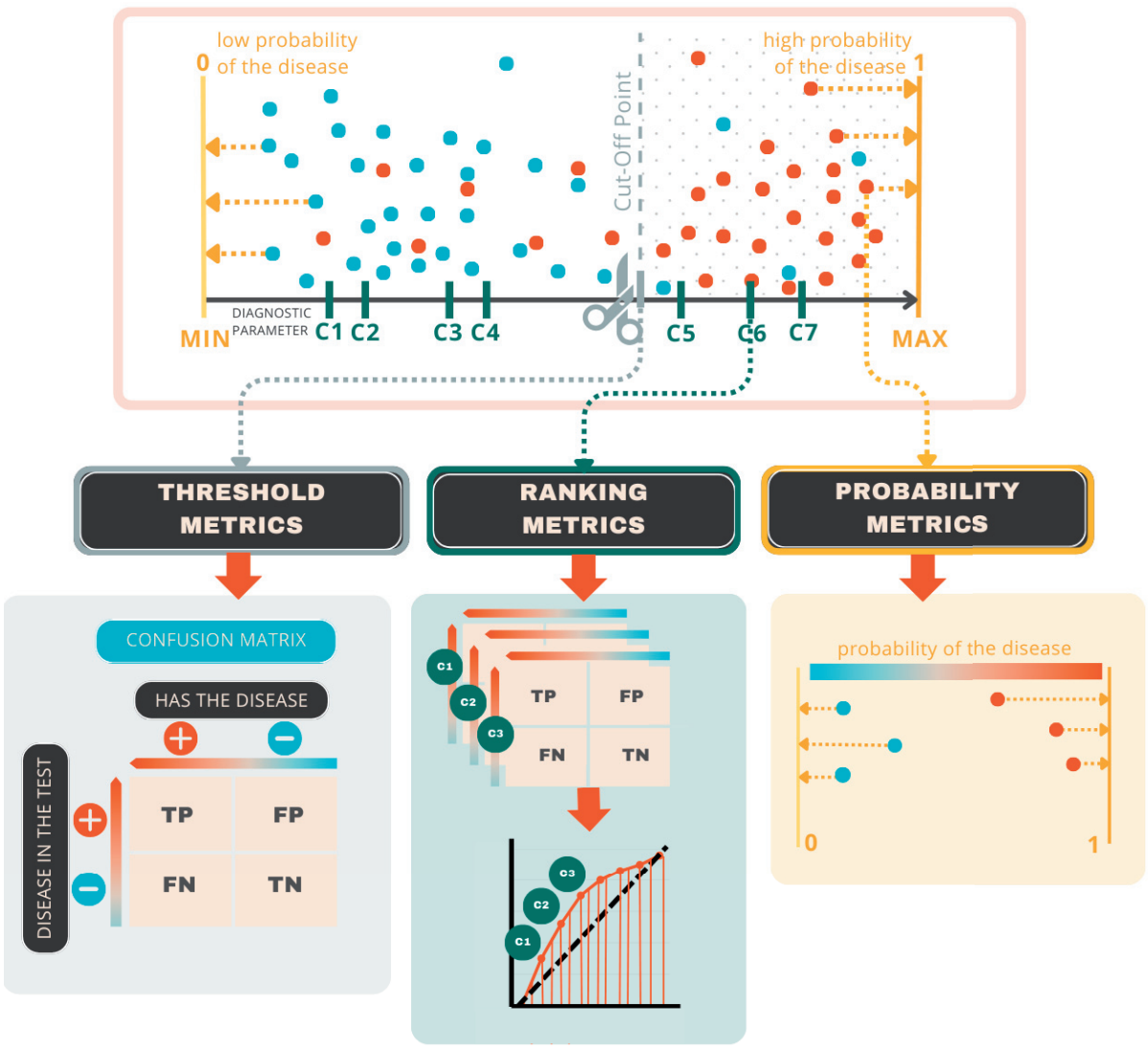


Figure 1. Illustration of the fundamental concepts underlying threshold, ranking, and probabilistic metrics in binary classification. The diagram uses colour coding, with blue indicating patients (positive class) and red indicating healthy individuals (negative class), to represent the core components of classification evaluation visually.

Confusion matrix

A confusion matrix is a table used to define the performance of a classification algorithm. The matrix has two dimensions: predicted classification and actual classification. The predicted classification is the classification the investigator assigns to each patient (or other test subject), whereas the actual classification is the correct classification of each patient. The confusion matrix is shown in **Figure 2**, using the BIRADS scale for breast cancer detection as an example.

Such confusion matrices are used to compute threshold metrics. For Rank Metrics, multiple confusion matrices are created by varying cutoff points; in the example, separate tables are produced for BIRADS = 1, BIRADS = 2, BIRADS = 3, BIRADS = 4, and BIRADS = 5. Probability Metrics do not use such a matrix; instead, they are derived directly from a disease-specific probability value calculated for each patient and from the patient's distance to the actual value.

Metrics for assessing the quality of classification

Table 1 summarises commonly used measures to assess the quality of a classifier, divided into Threshold Metrics, i.e. based on a single confusion matrix, Rank Metrics based on multiple confusion matrices determined for sequentially ordered cut-off points, and Probabilistic Metrics based on classifiers that determine the probability of an event occurring, i.e. a value between 0 and 1.

Threshold Metrics

A simple classifier is binary. For instance, whether a patient exhibits symptoms or has a disease. There are also more complex binary classifiers, such as Naive Bayes, decision trees, and neural networks. Many of these models output continuous scores or class probabilities rather than direct class labels, and a confusion matrix can be obtained only after selecting a decision thresh-

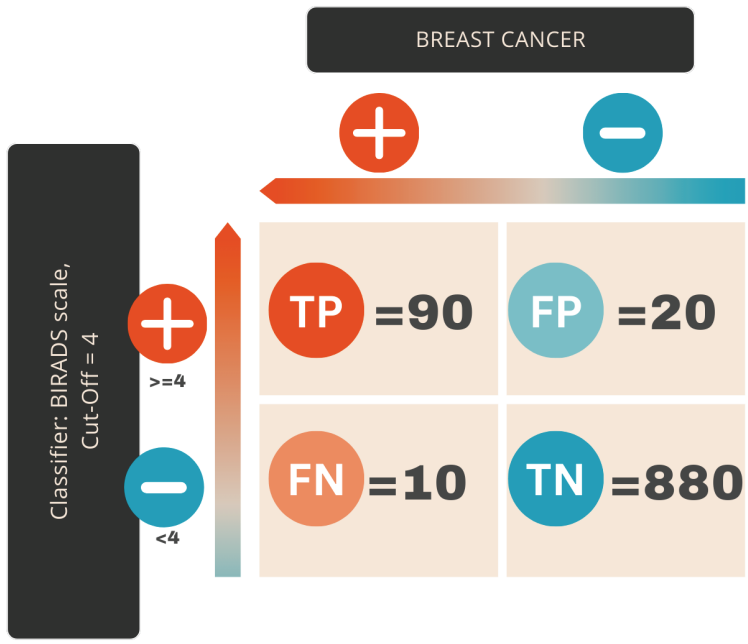


Figure 2. Example of a confusion matrix. In the columns of this matrix, a positive value indicates a positive class (occurrence of an event, in this case, breast cancer), and a negative value indicates a negative class (no event). In the rows of the table, a positive value indicates the detection of an event by the classifier (BIRADS scale >=4 indicates breast cancer), and a negative value indicates the non-detection of this event.

- The four cells of the confusion matrix represent the following: True Positive (TP): 90 patients (who have cancer) and are classified correctly as positive patients (cancer detected according to BIRADS).
- False Positive (FP): 20 patients (healthy), but are classified incorrectly as positive (cancer detected according to BIRADS).
- True Negative (TN): 880 patients (healthy) are classified correctly as negative (no cancer detected according to BIRADS).
- False Negative (FN): 10 patients (who have cancer), but are classified incorrectly as negative (no cancer detected according to BIRADS).

Table 1. Main metrics to evaluate the quality of classification and their definitions.

Term	Definition and ranges of metrics along with interpretation
Measures based on the Threshold Metrics	
Accuracy TP FP FN TN	The proportion of correct predictions among all predictions. Takes a value between 0 (no accuracy) and 1 (complete accuracy). $\frac{TP + TN}{TP + TN + FP + FN}$
Error TP FP FN TN	The proportion of incorrect predictions among all predictions. Takes a value between 0 (no error) and 1 (maximum error). Complementary metrics to Accuracy. $\frac{FP + FN}{TP + TN + FP + FN}$
Sensitivity (Recall or True Positive Rate) TP FN	The proportion of true positive predictions (correctly identified sick individuals) out of all actual sick individuals. Takes a value between 0 (no sensitivity) and 1 (maximum sensitivity). $\frac{TP}{TP + FN}$
Specificity FP TN	The proportion of true negative predictions (correctly identified healthy individuals) out of all actual healthy individuals. Takes a value between 0 (no specificity) and 1 (maximum specificity) $\frac{TN}{TN + FP}$
G-mean TP FP FN TN	The geometric mean is sensitivity and specificity combined into a single result that balances both concerns. A G-mean value of 1 indicates a perfect balance between sensitivity and specificity, while a value close to 0 indicates an imbalance or poor performance of the classifier $\sqrt{sensitivity \times specificity} = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}}$
PPV (Precision) TP FP	Precision, otherwise known as Positive Predictive Value (PPV) is the proportion of true positives among all positive predictions. Takes a value between 0 (lack of prediction skills for the positive class) and 1 (perfect prediction skills for the positive class) $\frac{TP}{TP + FP}$
NPV FN TN	Negative Predictive Value (NPV) is the proportion of true negatives among all negative predictions. Takes a value between 0 (lack of prediction skills for the negative class) and 1 (perfect prediction skills for the negative class) $\frac{TN}{TN + FN}$
F1-score TP FP FN	It focuses on the model's ability to identify positive instances. Precision and Recall combined into one result that tries to balance both concerns. It is calculated as the harmonic average of Precision and Recall. F-score values range from 0 to 1. The higher the F-score value, the better the classifier performs in balancing precision and recall. $2 \times \frac{precision \times recall}{precision + recall} = (\frac{2 \times TP}{2 \times TP + FP + FN})$
F(β) -score TP FP FN	It focuses on the model's ability to identify positive instances. A weighted harmonic mean of Precision and Recall. In this formula, β determines the weight assigned to Recall compared to Precision. A higher value of beta gives more weight to Recall, while a lower value of β favors Precision. When beta is equal to 1, the F-beta score is equivalent to the F1-score, which balances Precision and Recall equally $\frac{(1 + \beta^2) \times (precision \times recall)}{\beta^2 \times precision + recall} = \frac{(1 + \beta^2) \times (\frac{TP}{TP + FP} \times \frac{TP}{TP + FN})}{\beta^2 \times \frac{TP}{TP + FP} + \frac{TP}{TP + FN}}$

Table 1. Continuued.

Term	Definition and ranges of metrics along with interpretation
DOR 	Diagnostic Odds Ratio (DOR) is the ratio of two chances: the chance of a positive classifier result from a diseased person to the chance of a positive classifier result from a healthy person. A higher DOR indicates a better discriminatory power of the classifier, with values greater than 1 suggesting higher odds of a positive classifier result in individuals with the condition compared to those without.
	$$\frac{TP \div FN}{FP \div TN}$$
MCC [26] 	Mathews Correlation Coefficient (MCC) metrics the correlation of the true classes with the predicted labels. MCC ranges in the interval [−1,+1], with −1 meaning perfect misclassification and +1 perfect classification. MCC generates a high score in its interval only if the classifier scores a high value in all of the following: sensitivity, specificity, precision, and negative predictive value.
	$$\frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$
Kappa (Cohen's Kappa Coefficient) 	This metric evaluates the agreement between predicted and actual classes. Also, it considers that some correspondence between predicted and actual classes could occur by chance and eliminates random correspondence. Cohen's Kappa value ranges from -1 to 1, where a value of 1 indicates full agreement and a value of 0 indicates complete randomness. Negative values denote agreement weaker than random is rarely achieved in practice.
	$$\frac{observed\ agreement - expected\ agreement}{1 - expacted\ agreement} \text{ or } 2 \times \frac{TP \times TN - FN \times FP}{(TP+FP) \times (FP+TN) + (TP+FN) \times FN+TN}$$
Measures based on the Ranking Metrics (Methods)	
ROC Curve 	The Receiver Operating Characteristic Curve (ROC) is a graphical representation of the performance of a binary classifier system as its discrimination cut-off is varied. The higher the ROC curve rises above the diagonal (random line), the better the performance of the classifier. A classifier without skill formulates a line that winds along the diagonal.
	The Y-axis presents sensitivity (true positive rate, TPR), X-axis presents 1- specificity (false positive rate, FPR). Both values are obtained from the successive confusion matrices determined for each of the possible classifier cut-off points.
AUC(ROC) 	The AUC(ROC) (Area Under the Curve (Receiver Operating Characteristic Curve) is interpreted as the probability that the model will return a higher probability of illness to an arbitrarily selected sick person than an arbitrarily selected healthy person. Where a value of 1 indicates ideal classification (perfect skill classifier), and a value of 0.5 indicates random classification (no skill classifier).
	AUC(ROC) is calculated as the sum of the areas of the trapezoids where the area under the ROC curve is divided.
PR Curve 	Precision-Recall Curve (PR Curve) is a graphical representation of the performance of a binary classifier system that focuses on TP cases at different decision cut-offs. The closer the Precision-Recall curve is to the point (1,1), the better the performance of the classifier is. The no-skill line changes according to the balance of the classes. This is a horizontal line representing the proportion of positive cases in the data set. In the case of a balanced dataset, it is 0.5; if, for example, the sickness rate is 20%, the line is at 0.2.
	Build a graph with Precision as the y-axis and Recall as the x-axis. Both values are obtained from the successive confusion matrices determined for each of the possible classifier cut-off points.
AUC(PR) 	The area under the Precision-Recall curve (AUC(PR) is used as a measure of the overall performance of the classifier, where a value of 1 indicates perfect classification (perfect skill classifier) and a value of 0 indicates the worst possible result (no skill classifier).
	AUC(PR) is calculated as the sum of the areas of the trapezoids where the area under the Precision-Recall curve is divided.

Table 1. Continuued.

Term	Definition and ranges of metrics along with interpretation
	Probabilistic Metrics
LogLoss (cross-entropy)	Logarithmic Loss (LogLoss), also known as the logarithmic loss function or cross-entropy loss, is a measure of the magnitude of error used in classification problems. It measures the degree of deviation between the actual values (0 for the no-event class and 1 for the event class) and the probabilities predicted by the classifier. LogLoss values are always non-negative, where a value of 0 indicates a perfect match between actual class values and predicted probabilities, a random model would have a log loss of around 0.693. The higher the LogLoss value, the greater the deviation between the predicted and actual values
	For binary classification is calculated as: $Log\ Loss = = \frac{-1}{N} \times \sum \quad [y \times log(\underline{y}) + (1-y) \times (1-y) \times log(1-\underline{y})]$ where y is the actual value of the class (0 or 1), and $\underline{y}$ is the predicted probability of an event (disease)
Brier Score [Brier GW, Mean squared error)	Brier's score measures the mean square error between actual values (0 for the no-event class and 1 for the event class) and the probabilities predicted by the classifier. Brier's index takes values from 0 to 1. Lower values indicate better classification. An ideal classification would achieve a Brier index of 0, while a completely wrong, unreliable model would achieve a Brier index value of 1 and a random model would have a Brier score of around 0.25.
	$Brier\ Score = \frac{1}{N} \times \sum_{i=1}^{n} (\hat{y_i} - y_i)^2$

old that converts these scores into binary predictions. Once such a threshold is specified, a confusion matrix can be constructed and threshold-based metrics derived from it.

Variables with multiple possible cut-off points are standard. During model development and comparative evaluation, relying on a single confusion matrix at a single chosen threshold can be problematic because threshold-based measures are sensitive to the selected cutoff, and the "optimal" threshold may differ across datasets, even for the same biomarker. In contrast, for a clinically implemented biomarker or diagnostic test, establishing and validating a single, pre-specified threshold that can be applied consistently across laboratories and settings is a strength rather than a limitation. Accuracy and Error may also be misleading in imbalanced datasets, which are frequent in medical studies where patient groups are much smaller than control groups (healthy individuals). Their practical application in current medical research often leads to the "Accuracy Paradox," particularly in imbalanced datasets. For instance, in rare disease screening where pathology prevalence is low (e.g., 1%), a naive classifier predicting all patients as healthy achieves 99% accuracy but fails clinically due to 0% sensitivity. This phenomenon is frequently observed in large-scale health record analyses, where neglecting metrics like Balanced Accuracy or G-mean can obscure a model's inability to detect the minority class of interest [25]. Positive predictive value (PPV) and negative predictive value (NPV) are often more clinically informative in these settings because they quantify the probability of disease given a test result; however, they are strongly dependent on disease prevalence, which limits their transportability between populations.

Threshold metrics, by contrast, offer greater versatility. These metrics can be determined regardless of the type of data on which the marker was measured. It is possible to decide on both quantitative and qualitative data, structured and binary data. The simplicity of calculating and interpreting these metrics is also essential. For instance, utilising a confusion matrix facilitates straightforward computation of various performance measures (as shown in **Table 1**) for BIRADS with the cut-off shown in **Table 2**.

As outlined in **Figure 6**, these metrics are the primary choice only when the core task is binary

Table 2. The calculation results obtained for the confusion matrix presented in Figure 1, together with their interpretation

Threshold Metric	Calculations and Conclusions
Accuracy	(880 + 90) / 1000 = 0.97
	97% of individuals were classified correctly
Error	(10 + 20) / 1000 = 0.03
	3% of individuals were classified incorrectly
Sensitivity (Recall or True Positive Rate)	90 / (10 + 90) = 0.9
	90% of sick individuals were correctly classified
Specificity	880 / (20 + 880) = 0.98
	98% of healthy individuals were correctly classified
G-mean	sqrt(0.9 * 0.98) = 0.94
	Mean sensitivity and specificity is 94%
PPV (Precision)	90 / (20 + 90) = 0.82
	82% of individuals with a positive result (BIRADS>=4) were actually sick.
NPV	880 / (10 + 880) = 0.99
	99% of individuals with a negative result (BIRADS < 4) were actually healthy.
F1-score	2 * ((0.82 * 0.9) / (0.82 + 0.9)) = 0.86
	0.86 indicates a very good balance between precision and recall.
F2-score	F2 score = (1 + 2^2) * ((0.82 * 0.9) / ((2^2 * 0.82) + 0.9)) = 0.88
	With the assumption that recall is twice as important as precision, we still point out that BIRADS classifies fairly accurately.
DOR	(90/10)/(20/880) = 396
	An individual with a positive test result is 396 times more likely to have the disease compared to someone with a negative test result.
MCC	((90*880)-(10*20))/sqrt((90+20)*(90+10)*(880+20)*(880+10)) = 0.842
	0.842 indicates a strong positive correlation and a substantial agreement between BIRADS and reality.
Kappa (Cohen's Kappa Coefficient)	2*(90*880-10*20)/((90+20)*(20+880)+(90+10)*(10+880)) = 0.84
	Agreement for classification between BIRADS and reality after taking into account the agreement.

classification, and the chosen threshold is fixed and clinically validated.

Threshold Optimisation: Selecting the Optimal Cut-off Point

The selection of an optimal cut-off point is crucial for threshold-based classification metrics such as Sensitivity, Specificity, and Accuracy. The choice of threshold directly impacts the classifier's performance and its clinical utility. While a fixed threshold may be appropriate in some cases, optimal threshold selection is often necessary to balance false positives and false negatives effectively. Several methods exist for determining the best threshold, including Youden's Index, cost-benefit analysis, and clinically relevant decision-making frameworks.

Youden's Index: Maximising Sensitivity and Specificity

Youden's Index is one of the most commonly used methods to select an optimal threshold. It is defined as:

$$J = Sensitivity + Specificity - 1$$

The optimal threshold is the point at which Youden's Index is maximised, meaning it provides the best trade-off between true positive rate and actual negative rate. This method is widely used in diagnostic tests where equal importance is placed on detecting disease (high Sensitivity) and avoiding misclassification of healthy individuals (high Specificity).

Cost-Benefit Analysis: Accounting for Clinical Consequences

In many real-world applications, the costs associated with FP and FN classifications are not equal. A cost-benefit analysis helps determine a threshold that minimises the overall impact of classification errors. This approach assigns a weight or cost to each type of error based on its clinical consequences. The optimal threshold is the one that minimises the expected total cost, which is calculated as:

$$Total\ Cost = (CFP \times FP\ Rate) + (CFN \times FN\ Rate)$$

Where *C*FP and *C*FN represent the relative costs of false positives and false negatives, respectively, for example, in cancer screening, a false negative (missed diagnosis) may have a significantly higher price (delayed treatment) than a false positive (leading to additional but unnecessary testing).

Clinically Relevant Decision-Making Frameworks

Beyond mathematical optimisation, clinical decision-making frameworks integrate real-world impact into threshold selection. For instance, in sepsis prediction models, a lower threshold may be preferred to increase early detection rates, even at the cost of a higher false-positive rate. Conversely, in cardiac risk stratification, a stricter threshold may be needed to prevent unnecessary interventions.

One example is the Net Benefit Approach, which considers both the relative utility of true positives and the harm of false positives in medical decision-making. This approach is frequently used in risk-based screening guidelines.

Ranking Metrics

Receiver operating characteristic (ROC) and precision-recall curves are standard methods for evaluating classifier performance. They enable comparison of different diagnostic parameters (or classification models), regardless of the decision threshold. We adjust the threshold by classifying test results as positive (sick) or negative (healthy) to derive these curves (see **Figure 4**). We then construct a confusion matrix to plot the ROC and PR curves for each threshold.

The ROC curve is based on Sensitivity (True Positive Rate, TPR) and a value of 1 1-Specificity (False Positive Rate, FPR = 1-Specificity). For different thresholds, points are plotted on the graph, creating an ROC curve that shows how the test balances between Sensitivity (Y-axis) and 1-Specificity (X-axis) as the classifier's thresholds are successively applied. The PR curve is based on Precision (PPV) and Recall (Sensitivity). We plot points for different thresholds, creating a PR curve that shows how the test balances Precision (Y-axis) and Recall (X-axis) as a function of the classifier's successive tapping thresholds.

Although each of these metrics is based on successive cut-off points (thresholds) and successive confusion matrices, they take into account slightly different aspects of the assessment of classification quality and have distinct advantages and disadvantages (see **Table 3**)

The widespread use of ROC curves is supported by the development of statistical tests to assess the significance of the AUC and to compare ROC curves, such as the DeLong method and the Hanley-McNeil test [26]. Additionally, techniques such as Youden's method facilitate the determination of optimal cutoff points based on ROC curves [26]. It is critical to distinguish two properties: while AUC (ROC) is robust in the sense that its value is mathematically independent of changes in class prevalence (resampling), it can be highly misleading or overly optimistic in imbalanced settings. This "optimism" stems from the fact that AUC (ROC) treats all false alarms (FP) and missed diagnoses (FN) equally, regardless of the actual costs and the clinical dominance of the minority class.

Consider a dataset with a 99:1 ratio of negative to positive examples. A conservative, non-trivial classifier might achieve very high Specificity (low FPR) but low Sensitivity TPR), leading to a misleadingly good AUC(ROC). This optimism arises because the ROC curve does not reflect the low overall prevalence of the positive class. To avoid this misleading optimism, a different metric, such as PR curves, is often preferred.

There are several ways to address this problem. One is to use a different metric, such as Precision-Recall curves. This metric is more sensitive to class distribution and can give a more accurate picture of a model's performance when the minority class is the sick class. See the examples of L-selectin and P-selectin in the detection of psoriasis, and the example of heart rate in the detection of coronary disease (see **Figure 5**).

Figure 3. Advantages and disadvantages of individual Threshold Metrics.

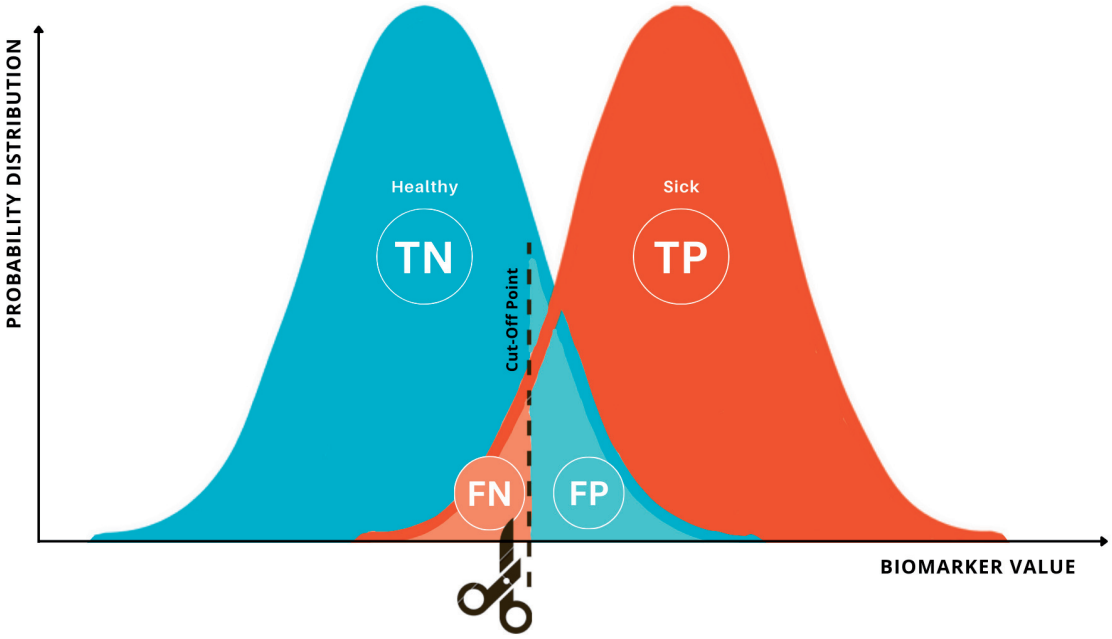
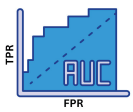
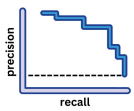


Figure 4. Outcomes of cut-off point (threshold) selection for a continuous biomarker as a potential classifier. The field sizes obtained at the indicated cutoff point yield four quantities: TP, TN, FP, and FN, which are entered into the confusion matrix.

Table 3. Advantages and Disadvantages of Ranking Metrics.

Ranking Metrics	Advantages	Disadvantages
ROC 	**Visual Representation:** ROC curves provide a visual representation of the performance of a classification model across all possible thresholds. **Comprehensive Performance Summary:** AUC (ROC) is a single metric that summarizes the overall performance of a diagnostic test. Represents the probability that a randomly chosen positive case will receive a higher test score than a randomly chosen negative case. **Threshold-Independent Comparison:** Two or more ROC curves can be compared directly even if they are derived from different variables with different units. **Independence from Prevalence:** The ROC coordinates (TPR, FPR) are mathematically independent of the class distribution (prevalence), making the AUC (ROC) value determinable even if the prevalence changes	**Limited Clinical Interpretation:** AUC (ROC), despite its popularity, may not directly translate into meaningful information for clinicians, patients, or healthcare providers. A test with an AUC of 0.9 might be considered "better" than one with an AUC of 0.8, but this difference may not have a significant impact on patient outcomes or treatment decisions. **Focus on All Thresholds:** AUC (ROC) considers the performance of a test across all possible thresholds, including those that may not be clinically relevant or useful in practice. This can lead to an overemphasis on thresholds that are not practical or important for decision-making. **Can be misleading in imbalanced settings:** This is because AUC (ROC) gives equal weight to TPR and FPR, regardless of the actual prevalence. In highly imbalanced datasets, the resulting high AUC may mask poor performance on the minority class, as the FPR remains small due to the dominant TN count. This often leads to a misleadingly high AUC (ROC) when compared to AUC (PR)
PR 	**Visual Representation:** PR curves provide a visual representation of the performance of a classification model at different Recall levels. This allows for a quick and intuitive understanding of the model's ability to identify positive cases (e.g., disease presence) while considering the trade-off between Precision and Recall **Focus on Positive Class:** PR curves specifically focus on evaluating the performance of a model in identifying positive cases (e.g., disease presence), making them particularly useful in scenarios where the detection of these cases is of primary importance. They provide precise information about the model's ability to correctly classify positive instances, even when they are rare or difficult to identify.	**Focus on Positive Class:** PR curves primarily focus on evaluating the performance of a model in identifying positive cases (e.g., disease presence), disregarding the number of true negative results (e.g., disease absence). This can make them less suitable for tasks where both positive and negative classifications are equally important. **Neglect of Healthy Individuals:** PR curves do not directly assess the model's ability to correctly classify healthy individuals (true negatives). **Sensitivity to Data Imbalance:** This sensitivity makes it challenging to compare PR curves from different studies or datasets with varying class imbalances.

Methodological note for illustrative examples

To ensure consistency, calculations for the provided examples (L-selectin, P-selectin, Heart Rate) were performed using individual probabilities obtained from a univariate Logistic Regression model trained on the respective variable (Table 6, Figure 5). Ranking metrics (AUC-ROC and AUC-PR) and probability metrics (Brier Score, Log Loss) were calculated directly from these continuous outputs without applying any arbitrary classification thresholds. The R code used to perform these calculations is available on GitHub (https://github.com/marpatra/Metrics.Selectins-HR)

The ROC curve for L-selectin lies along the line of identity. At the 0.5 level with a balanced dataset (50:50 ratio), the prevalence line is at 0.5 on the PR curve, indicating that L-selectin is a minimal-skill classifier (see **Figure 5**). While the AUC(PR) of 0.523 is technically above the random expectation baseline of 0.50, the value is negligibly close to the no-skill classifier line, justifying its practical classification as non-discriminatory. The ROC and PR curves for P-selectin indicate that low levels of this selectin have discriminatory potential; at high levels, the curves lie along a line indicating no skill.

For the balanced cardiac dataset, we obtained ROC and PR curves, yielding high fields and their plots, and also demonstrating strong classification performance for the heart rate variable. However, for the imbalanced data, the AUC(ROC) value of 0.776 is higher than that for the balanced data (0.745). This finding is due to the substantial increase in the TN, which pushes the FPR close to zero across many thresholds. This results in the ROC curve appearing overly optimistic. In sharp contrast, the AUC(PR) for the same imbalanced data drastically drops to 0.494 (from 0.704 in the balanced setting). This low AUC (PR) indicates the classifier's weakness in the minority class (patients). Specifically, the optimistic AUC(ROC) masks a clinically unacceptable drop in PPV, where many optimistic predictions are actually FP relative to the few TP available. In clinical practice, this would mean a high rate of false alarms, wasting resources and causing unnecessary patient anxiety, a critical factor missed by the ROC curve alone.

To mitigate imbalance during model training, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or undersampling can be employed to balance class distribution. Alternatively, cost-sensitive learning can be used to assign higher penalties to misclassifying the minority class.

Figure 6 directs the user to these ranking metrics when the primary goal is classification, but when the data are imbalanced or a threshold-agnostic comparison between models is required.

Real-World Applications: Robust Metrics and Resampling Genomic Studies (Metric Selection): As highlighted in recent bioinformatics literature, standard metrics like F1-score can be misleading when the 'negative' class is biologically significant. Chicco and Jurman [21] demonstrated that, in genomic binary classification, the F1-score remained optimistically high even when the model failed to detect negative samples correctly. In contrast, the Matthews Correlation Coefficient (MCC) declined markedly in these scenarios, indicating the model's poor performance. Thus, for omics data, MCC is recommended over F1 as a more truthful performance indicator.

Pandemic Surveillance (Data Resampling):
In scenarios such as a COVID-19 diagnosis, datasets are often skewed toward negative cases. Research on radiomics applied to COVID-19 [2] has shown that models trained on imbalanced internal cohorts typically exhibit high Sensitivity but suffer a sharp decline in Specificity when validated externally due to overfitting. To mitigate this, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) have been successfully employed to balance training sets, preventing the model from becoming biased toward the majority class (healthy individuals) and ensuring reliable detection of infected patients.

Robust Metrics and Resampling

Probabilistic Metrics

As shown in **Figure 6**, these metrics become the priority when the core modelling objective is the prediction of a trustworthy probability value (P), essential for accurate clinical risk assessment

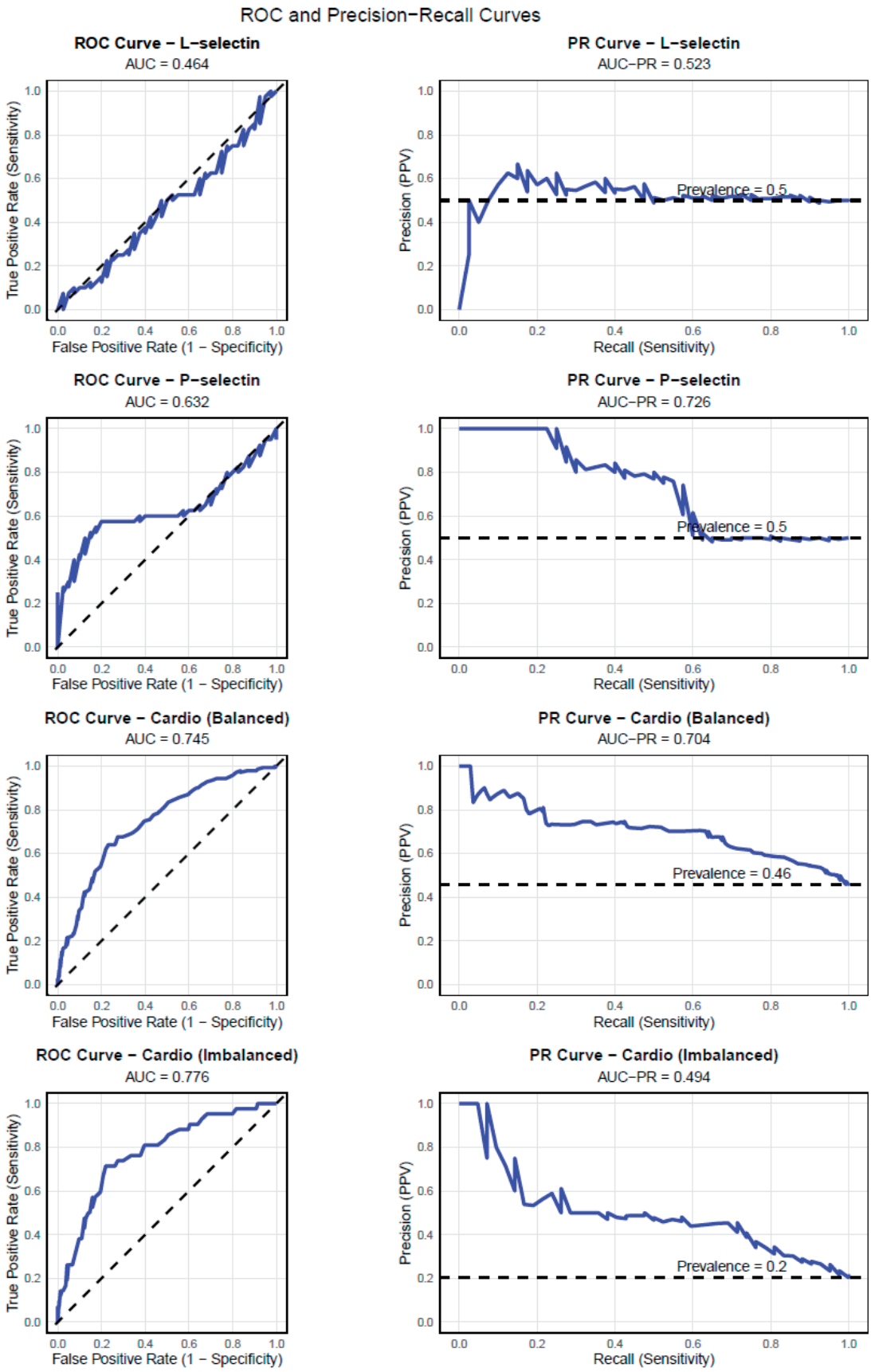


Figure 5. ROC and PR curves for balanced data: no-skill classifier using the example of L-selectin in psoriasis detection, partially-skill classifier using the example of P-selectin in psoriasis detection, skilful classifier using the example of heart rate in coronary disease detection for balanced data; partially skilful classifier using the example of heart rate in coronary disease detection for imbalanced data. Calculations were performed using individual probabilities obtained from a logistic regression model trained on the input variable. The results obtained without using the logistic regression model are identical.

When to Prefer Brier Score and Log Loss

Brier Score and Log Loss are essential metrics for evaluating probabilistic models, particularly when the focus is on the accuracy of predicted probabilities rather than binary classifications. These metrics are preferred over threshold-based metrics (e.g., accuracy, F1-score, sensitivity, specificity, ROC curves, and PR curves) in the following scenarios:

1. Probabilistic Models: When using models such as logistic regression, neural networks, or Bayesian classifiers, which output probabilities rather than binary predictions. These metrics are particularly suited for assessing the quality of probability estimates, which are often more informative than binary decisions in medical applications.
2. Calibration Assessment: When evaluating how well the predicted probabilities align with actual outcomes. For example, a well-calibrated model predicting a 30% risk of an event should observe the event occurring approximately 30% of the time. Calibration is critical in clinical decision-making, where accurate probability estimates are necessary for risk stratification and treatment planning.
3. Sensitivity to Small Errors: When the model's performance depends on accurately predicting probabilities, especially for rare events or imbalanced datasets. Log Loss, in particular, penalises overconfidence in incorrect predictions, making it a valuable tool for training and fine-tuning probabilistic models.

Evaluation of Continuous Probabilities: The Two-Stage Process

The evaluation of a diagnostic test often involves a critical two-stage process that requires different metrics: Stage 1: probability prediction (calibration), and Stage 2: final classification (thresholding).

Stage 1 metrics, such as the Brier Score and Log Loss, assess the quality of the model's raw probability output (P) before any decision threshold is applied. They quantify the model's calibration, ensuring that if the model predicts a probability P, the outcome occurs approximately P per cent of the time. This is critical because, for clinical decision-making, the expected probability P must be trustworthy.

The importance of calibration: an illustrative example

Metrics such as AUC (ROC) are primarily ranking metrics; they assess only the model's ability to correctly order positive cases above negative cases, regardless of the actual probability values. Two models can have identical AUC values but vastly different calibration quality. Consider a small dataset of 10 cases (4 positive (1), 6 negative (0)). Two hypothetical models, Model M1 (well-calibrated) and Model M2 (poorly-calibrated), produce the following probabilities (see **Table 4**):

Despite providing different probability scores, the rank order of cases is identical for both models. Consequently, both Model M1 and Model M2 achieve a perfect AUC(ROC) of 1.0. Based solely on the AUC, we would conclude that both models

Table 4. Comparison of Ranking (AUC) and Probabilistic (Brier Score) Metrics for Two Models with Identical Ranking Ability but Different Calibration (Numerical Example). The R code used to perform these calculations is available on GitHub (https://github.com/marpatra/Metrics.Selectins-HR)

Case	True label	M1 (calibrated prob.)	M2 (uncalibrated prob.)
1	1	0.60	0.90
2	1	0.55	0.85
3	1	0.40	0.70
4	1	0.30	0.60
5	0	0.25	0.55
6	0	0.15	0.40
7	0	0.10	0.30
8	0	0.05	0.20
9	0	0.02	0.10
10	0	0.01	0.05
Summary	AUC-ROC	1.00	1.00
	Brier Score	0.09	0.13

are perfect classifiers. However, when assessing calibration using the Brier Score (BS), the difference becomes clear: BS(M1) = 0.089, BS(M2) = 0.131. Since a lower Brier Score indicates better performance, Model M1 is significantly better calibrated than Model M2. Model M2 systematically underpredicts the risk for positive cases (e.g., predicting 0.60 instead of 0.90 for Case 1). Using Model M2 in a clinical setting would lead practitioners to be consistently overconfident that patients are not sick when a moderate probability is predicted, resulting in poorer clinical decisions despite the model's perfect AUC ranking.

For a more comprehensive visualisation and assessment of calibration, it is standard practice to use Calibration Plots or Reliability Diagrams alongside Brier Score and Log Loss. These graphical tools compare predicted probabilities with observed frequencies across multiple bins, providing an intuitive way to identify systemic biases in the probability output (e.g., over- or underestimation).

Stage 2 involves applying a decision threshold (cutoff point) to the calibrated probability output to produce the final binary classifications (e.g., 'sick' or 'healthy'). At this stage, threshold-dependent metrics such as Sensitivity, Specificity, PPV, NPV, F1-Score and Accuracy are used to assess the final classification performance based on the chosen trade-off.

Practical Applications in Medicine

Prognostic Models in Oncology:
In cancer survival prediction, binary classification (alive/dead) is often insufficient; clinicians require the survival probability to weigh treatment risks. Steyerberg et al. [27] emphasise that a model can have high Accuracy but poor calibration (e.g., consistently predicting 60% risk for patients who actually have a 40% risk). In such cases, the Brier Score is the superior metric because it quantifies the distance between the predicted probability and the actual outcome. A lower Brier score is associated with more reliable risk estimates, which are crucial for deciding whether to administer toxic chemotherapy.

Cardiovascular Risk Scoring:
Similarly, in predicting 10-year cardiovascular event risk (e.g., the Framingham Risk Score), Log Loss is widely used to penalise confident but incorrect predictions. If a model predicts a 99% chance of 'no heart attack' for a patient who subsequently suffers one, Log Loss applies a heavy penalty, forcing the algorithm to be more cautious and realistic in its probability estimates during training."

By incorporating Brier Score and Log Loss into the evaluation process, researchers and clinicians can ensure that their models provide not only accurate classifications but also reliable probability estimates, ultimately improving patient outcomes.

Implementations in Practice

Implementations of these measures are less widely available than other performance metrics, such as Accuracy or Precision. This can make it challenging to use these metrics in some programming environments. It is important to emphasise that Brier Score and Log Loss have strengths and weaknesses, and the choice of appropriate metrics depends on the specifics of the task (see **Table 5**).

Depending on the software used, the values obtained for these measures may vary minimally. In our examples (L-selectin and P-selectin for psoriasis detection, and heart rate for coronary disease detection), we first fitted simple logistic regression models with the respective variable as the predictor, yielding an individual predicted probability of the outcome for each patient. Brier Score was then calculated as the mean squared difference between these case-wise predicted probabilities and the observed outcomes, and Log Loss was computed using the exact individual probabilities. The results for calculations performed in R and in Python are presented in **Table 6**.

Validation

The data used to build and assess a classifier's quality is called learning or training data. In the following steps, each classifier, whether simple (based on a single variable) or complex, like a logistic regression model, neural network, or decision tree, should be validated with independent data, called test or validation data. Repeated testing of the same classifiers or models on

Table 5. Advantages and disadvantages of the chosen Probability Metrics.

Probability Metrics	Advantages	Disadvantages
Brier score	**Interpretability:** Brier Score is easier to interpret than Log Loss. It repre sents the mean squared difference between the predicted probabilities and the actual outcomes (0 or 1). A lower Brier Score indicates better model performance, making it intuitive to understand how well the model performs on average. **Robustness to calibration issues:** Even if a model's predicted probabilities are not perfectly aligned with the actual outcomes, Brier Score can still provide a reasonable assessment of performance.	**No penalty for overconfidence:** does not penalize models for being too confident but incorrect in their predictions. This can lead to preferring models that predict outstanding high/low values, even if they are wrong. In the case of extremely imbalanced data, where one class accounts for less than 1% of observations, **may not reflect the true effectiveness of the model.**
Log Loss	**Suitable for training models:** Minimizing Log Loss during training encourages the model to learn accurate probability estimates. **More sensitive to probability differences:** Log Loss is more sensitive to differences in predicted probabilities compared to Brier Score. This allows it to better distinguish between models that make subtle but significant improvements in probability estimation. **Discourages overconfidence:** Log Loss heavily penalizes models that are overly confident in their wrong predictions. This can be beneficial for tasks where assigning the correct probabilities is crucial.	**Limited interpretability:** Log loss is difficult to interpret in real-world terms. The metric is based on logarithmic values, which makes it difficult to interpret intuitively. **Assumes well-calibrated probabilities:** Log loss works best when a model predicts probabilities with high Accuracy. If a model is inaccurate or poorly calibrated, log loss may not be a reliable measure of performance. This can lead to preferring models that underpredict positive class probabilities, even if they better identify actual outcomes. In the case of extremely imbalanced data, where one class accounts for less than 1% of observations, **may not reflect the true effectiveness of the model.**

Table 6. Brier Score and Log Loss for balanced data for: no-skill classifier using the example of L-selectin in psoriasis detection, partially-skilled classifier using the example of P-selectin in psoriasis detection, skilful classifier using the example of heart rate in coronary disease detection; and for imbalanced data for a partially skilful classifier using the example of heart rate in coronary disease detection. Calculations were performed using individual probabilities obtained from a logistic regression model trained on the input variable and a null model (without variables). Results were presented using the LogLoss function from the MLmetrics package in R, and based on the LogisticRegression, log_loss, and NumPy functions from the scikit-learn and NumPy packages in Python. The data used for these calculations, along with the functions necessary to perform them, are available at https://github.com/marpatra/Metrics.Selectins-HR

	Classifier	Brier Score		Log Loss	
		Python	R	Python	R
L-selectin in psoriasis detection	no-skill	0.23050299261287152	0.230503	0.692726390827548	0.6927264
P-selectin in psoriasis detection	partially skillful	0.23457753518070662	0.2345775	0.656469675228019	0.6564697
heart rate in coronary disease detection	skillful	0.22963327943711911	0.2296331	0.6517807033032978	0.65178
heart rate in coronary disease detection	partially skillful	0.22963327943711911	0.2296331	0.65178	0.65178

***The R** and Python implementations produce similar outputs with slight differences due to numerical precision, optimisation algorithms, library settings, and convergence criteria.

a new dataset will indicate how well the original predictive model and its classifiers perform on new, unseen data. All listed metrics can be calculated on both the training set, to assess the current quality of classification and prediction, and on new validation and test datasets, to generalise this quality to future data on which it may be used.

Model Validation and Overfitting

Model validation is a critical step in ensuring the reliability and generalizability of classification models. Overfitting occurs when a model performs exceptionally well on the training data but fails to generalise to new, unseen data. This typically happens when the model learns noise or specific patterns in the training data that do not apply to the broader population.

Techniques to Prevent Overfitting

k-Fold Cross-Validation: In k-fold cross-validation, the dataset is divided into *k* subsets (folds). The model is trained on *k-1* folds and validated on the remaining fold. This process is repeated *k* times, with each fold used exactly once as the validation set. The results are averaged to provide an estimate of model performance. For instance, k-fold cross-validation is essential for obtaining a stable estimate of AUC(ROC) for the Heart Rate classification model, ensuring that the reported performance is not specific to a single data split. **Leave-One-Out Cross-Validation (LOOCV)**: LOOCV is a special case of k-fold cross-validation in which *k* equals the number of samples in the dataset. Each sample is used once as a validation set, while the remaining samples form the training set. This method is beneficial for small datasets, as it maximises the use of available data. This technique could be employed to rigorously estimate the Sensitivity and Specificity of the BI-RADS scale classifier (e.g., at the >4 threshold) when validating its performance in small, limited patient cohorts.

Bootstrap Methods: Bootstrap resampling is a resampling technique used to assess the variability and internal stability of model performance within the same underlying population (e.g., the same dataset). By repeatedly drawing samples with replacement from the original data and refitting or re-evaluating the model, bootstrap methods provide estimates of the uncertainty and optimism of performance measures (e.g. AUC, Brier Score). However, bootstrap resampling cannot replace evaluation in genuinely different patient populations. Assessment of model performance across settings or populations requires external validation on separate datasets, rather than resampling from a single cohort. Bootstrapping is highly useful for assessing the stability and confidence intervals of the AUC(PR) and Brier Score values reported for the selectin and heart rate models, providing a measure of how much these metrics might vary across different potential patient samples, and thus detecting overfitting.

Implications for classification Metrics: practical examples

Validation techniques are not merely procedural steps; they often reveal critical flaws in metric interpretation that theoretical calculations on training data miss.

Consider a study that employs high-dimensional genomic data to predict cancer subtypes (e.g., the MAQC-II study [28]). A classifier might achieve an Accuracy of 98% and an AUC of 0.99 on the training set due to the model memorising noise (overfitting). However, when subjected to 10-fold cross-validation, the AUC might drop drastically to 0.60. This discrepancy serves as a red flag that the initial high metrics were illusory.

Similarly, in radiomics studies for COVID-19 detection [2], models often show high Sensitivity on training cohorts. However, external validation on data from a different hospital typically indicates a significant drop in Specificity, resulting in a high number of False Positives. This happens because the model may learn scanner-specific artefacts rather than disease pathology. In such cases, relying solely on training F1-scores would be misleading; cross-validation highlights the need for metrics such as the Matthews Correlation Coefficient (MCC), which is more robust to such shifts in confusion-matrix distributions than the F1-score or Accuracy.

Practical Considerations in Medical Applications

In medical applications, where the stakes are high, ensuring that a model generalises well to new data is crucial. Overfitting can lead to overly optimistic performance estimates, poten-

tially resulting in the deployment of unreliable diagnostic tools. Proper validation techniques help mitigate this risk, ensuring that the model's performance is consistent across different datasets and populations. For example, a model trained to predict sepsis must be validated on diverse patient cohorts to ensure its reliability in real-world clinical settings.

Discussion

Summary of Metric Strengths and Weaknesses within the Clinical Context

No single metric is universally optimal. Their selection must reflect the clinical context, priorities, and data characteristics.

Holistic Metrics: Accuracy (ACC) and Error Rate are useful when data are balanced, and misclassification costs are symmetric. Cohen's Kappa corrects these scores for chance agreement [27]. Matthews Correlation Coefficient (MCC) combines holistic assessment with robustness to class imbalance, making it a recommended metric in projects regulated by the U.S. FDA [29,30]. The Diagnostic Odds Ratio (DOR) summarises the overall discriminatory effectiveness of a test. The ROC curve and the area under it (AUC) are used to assess a test's ability to distinguish between groups across various thresholds [31-33]. The PR (Precision-Recall) curve is more informative than the ROC curve for detecting rare events [34]. The selection of an optimal cutoff point can be performed using Youden's index or, better adapted to clinical realities, the tangent (cost) method, which explicitly incorporates the relative costs of FP and FN and prevalence.

Class-Oriented Metrics: As indicated above, Sensitivity, Specificity, PPV, and NPV are fundamental, and their relevance depends on the clinical objective. The F1 score (and its variants) is instrumental in settings with class imbalance, as it combines precision (PPV) and sensitivity (recall).

Calibration Assessment Metrics: Brier Score and Log Loss assess the accuracy of estimated probabilities. The Brier Score is easier to interpret, whereas log loss is more sensitive to minor errors and is commonly used in machine learning.

Guidelines for Metric Selection and Their Clinical Rationale

The prevalence of a disease has a fundamental impact on the interpretation of diagnostic test

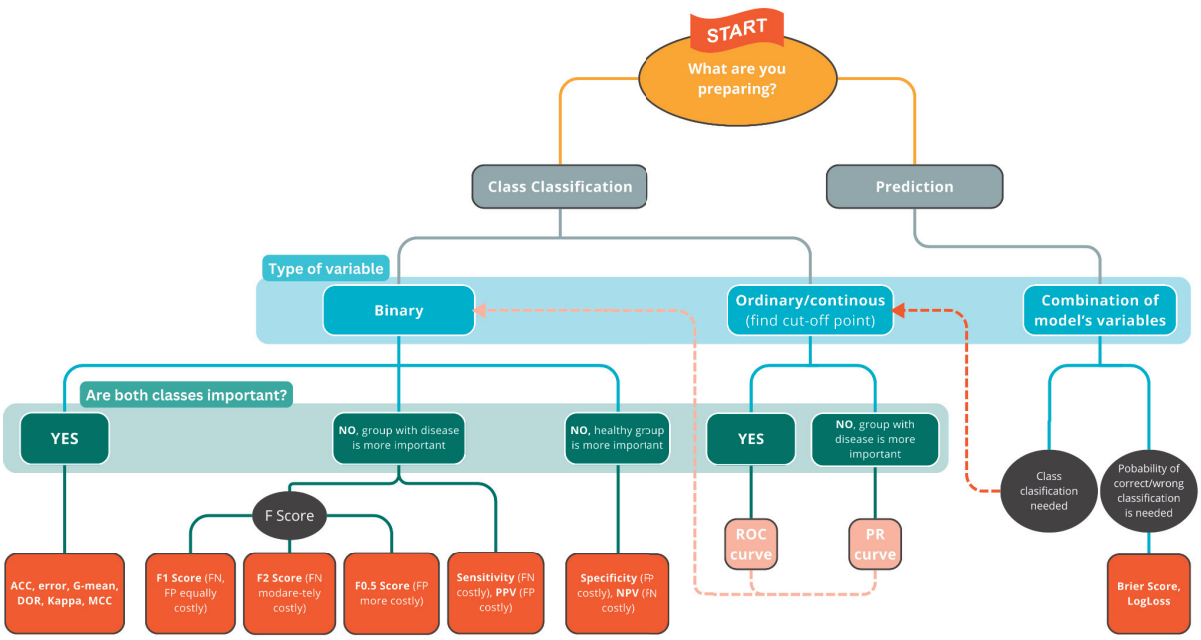


Figure 6. Graph facilitating the selection of a metric depending on the purpose for which it is determined (assignment to classes, prediction of probability of assignment to classes), type of data, validity or balancing of classes of events (sick) and no events (healthy). While this graph helps establish the primary objective, in practice, it is common to use a combination of indicators to evaluate the model's performance fully.

results and the choice of evaluation metrics. A structured selection framework, presented as a decision tree (see **Figure 6**), comprises three steps: the problem type (classification vs. prediction), the type of target variable, and the relative importance of classes for a given task.

The decision in the final step in medicine is critically determined by the asymmetry of misclassification costs, which varies with the stage of the diagnostic process. This is vividly illustrated in **Figure 7**, which presents the trade-off in cancer screening. At the screening stage, the primary goal is to rule out the disease, making a False Negative (FN) – missing a sick patient – the most critical error, as it delays potentially life-saving treatment. Consequently, screening tests are optimised for high Sensitivity and Negative Predictive Value (NPV). It is crucial to note, however, that a positive screening result typically triggers further confirmatory steps rather than immediate aggressive therapy. At this subsequent confirmatory diagnostic stage, the cost of a False Positive (FP) – subjecting a healthy person to invasive procedures and psychological distress – becomes predominant. Therefore, confirmatory tests must exhibit high Specificity and Positive Predictive Value (PPV) to ensure that treatment is administered only to those who genuinely need it. In practice, minimising one type of error often increases the other, and a common compromise is to use balanced metrics such as the G-mean or metrics from the F-score family.

A Critical Overview and Future Directions

A conscious selection of metrics – often involving a combination of several – is essential for reliable evaluation. The metrics discussed thus far represent established, mathematically rigorous approaches to model assessment. However, the field continues to evolve, driven by the need for more intuitive, actionable, and human-centric evaluation tools.

Classical metrics, together with considerations of disease prevalence and the asymmetric clinical costs of errors at different stages of the patient pathway, enable the selection of measures that ensure a clinically sound and accurate evaluation of diagnostic and predictive models. The future of model evaluation in medicine, however, lies in the synergy between these traditional foundations and the development of new techniques necessary for deploying AI systems (e.g., Shapley Additive exPlanations [35], U-smile [36]. New methods are constantly being introduced, and it was not possible to discuss and present them in a single summary.

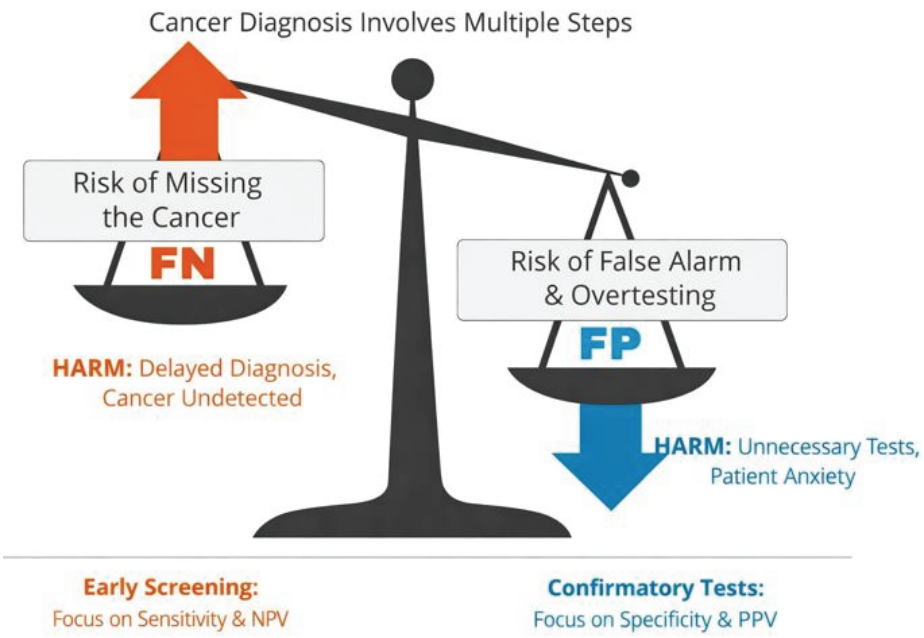


Figure 7. The harms caused by false negative and false positive prediction errors in the BI-RADS scale. A False Negative (FN) is typically weighted more heavily than a False Positive (FP), as the primary goal is to identify all potential cases to avoid the risk of disease progression.

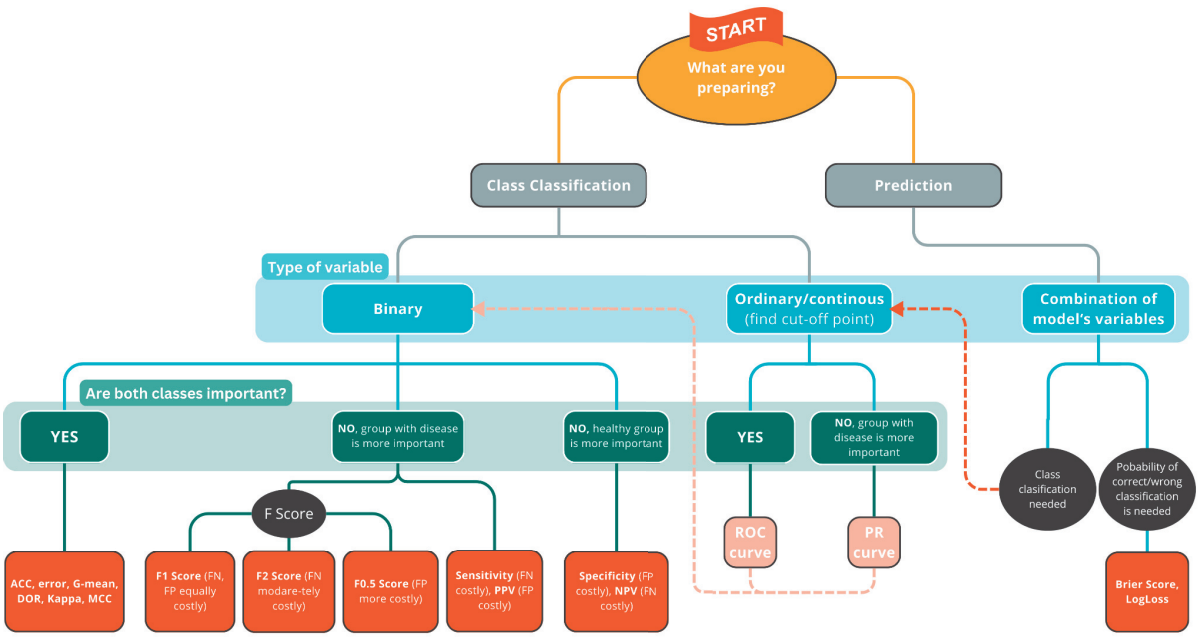


Figure 8. Algorithm for statistical decision-making.

It is important to note that this tutorial focuses on binary classification. In contrast, multi-class problems—along with their corresponding metrics, interpretability estimation methods, and associated evaluation challenges—constitute an essential and natural direction for future work in this rapidly evolving field.

Acknowledgements

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

Funding sources
There are no sources of funding to declare.

References

1. Rose G. Sick individuals and sick populations. Int J Epidemiol. 2001;30(3):427–432. https://doi.org/10.1093/ije/30.3.427.
2. Shandhi MMH, Cho PJ, Roghanizad AR, Singh K, Wang W, Enache OM, et al. A method for intelligent allocation of diagnostic testing by leveraging data from commercial wearable devices: a case study on COVID-19. NPJ Digit Med. 2022;5(1):130. https://doi.org/10.1038/s41746-022-00672-z.
3. Xi Y, Ding Y, Cheng Y, Zhao J, Zhou M, Qin S. Evaluation of the medical resource allocation: evidence from China. Healthcare (Basel). 2023;11(6):829. https://doi.org/10.3390/healthcare11060829.
4. Guzik P, Więckowska B. Data distribution analysis – a preliminary approach to quantitative data in biomedical research. J Med Sci. 2023;92(2):e869. https://doi.org/10.20883/medical.e869.
5. George DB, Taylor W, Shaman J, Rivers C, Paul B, O'Toole T, et al. Technology to advance infectious disease forecasting for outbreak management. Nat Commun. 2019;10(1):3932. https://doi.org/10.1038/s41467-019-11901-7.
6. Myers A, Johnston N, Rathore S, Kwon D, Kline J, Jehi L, et al. Electronic health record-based prediction model for acute kidney injury in patients undergoing major gastrointestinal surgery: a pilot study. J Pers Med. 2020;10(1):21. https://doi.org/10.3390/jpm10010021.
7. Flaks-Manov N, Topaz M, Hoshen M, Balicer RD, Shadmi E. Identifying patients at highest-risk: the best timing to apply a readmission predictive model. BMC Med Inform Decis Mak. 2019;19:118. https://doi.org/10.1186/s12911-019-0836-6.
8. Skov Benthien K, Kart Jacobsen R, Hjarnaa L, Mehl Virenfeldt G, Rasmussen K, Toft U. Predicting individual risk of emergency hospital admissions – a retrospective validation study. Risk Manag Healthc Policy. 2021;14:3865–3872. https://doi.org/10.2147/RMHP.S314588.
9. Berchialla P, Lanera C, Sciannameo V, Gregori D, Baldi I. Prediction of treatment outcome in clinical trials under a personalized medicine perspective. Sci Rep. 2022;12(1):4115. https://doi.org/10.1038/s41598-022-07801-4.
10. Selby JV, Fireman BH. Building predictive models for clinical care—where to build and what to predict? JAMA. 2015;313(17):1705–1706. https://doi.org/10.1001/jama.2015.3680.
11. Battineni G, Sagaro GG, Chintalapudi N, Amenta F. Applications of machine learning predictive models in the chronic disease diagnosis. J Pers Med. 2020;10(2):21. https://doi.org/10.3390/jpm10020021.

12. Ghaffar Nia A, Kaplan M, Khelifi A, et al. Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discov Artif Intell. 2023;3:5. https://doi.org/10.1007/s44163-023-00049-5.
13. Toma I, Wei Y. The application of artificial intelligence methods to public health data. Encyclopedia. 2023;3(2):590–601. https://doi.org/10.3390/encyclopedia3020042.
14. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ. 2023;23:689. https://doi.org/10.1186/s12909-023-04698-z.
15. Mansouri A, Mencattini A, Salmeri M, et al. A HUME approach for causal effects estimation in presence of unmeasured confounding. Bioinformatics. 2018;34(24):4274–4283. https://doi.org/10.1093/bioinformatics/bty490.
16. Ferner RE, Aronson JK. Susceptibility to adverse drug reactions. Br J Clin Pharmacol. 2019;85(10):2205–2212. https://doi.org/10.1111/bcp.14017.
17. Twick I, de Vetten JH, ten Berge M, et al. Performance measures for machine learning in a neonatal intensive care unit: a systematic review. J Am Med Inform Assoc. 2022;29(6):1064–1074. https://doi.org/10.1093/jamia/ocac036.
18. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. BMJ. 2015;351:h3868. https://doi.org/10.1136/bmj.h3868.
19. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Proc 23rd Int Conf Mach Learn (ICML). New York: ACM; 2006. p.233–240. https://doi.org/10.1145/1143844.1143874.
20. Canbek G, Taskaya Temizel T, Sagiroglu S. PToPI: a comprehensive review, analysis, and knowledge representation of binary classification performance measures/metrics. SN Comput Sci. 2022;4(1):13. https://doi.org/10.1007/s42979-022-01409-1.
21. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) is more informative than Cohen's kappa and Brier score in binary classification assessment. IEEE Access. 2023;16:4. https://doi.org/10.1109/ACCESS.2023.3301604.
22. Ferri C, Hernández-Orallo J, Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognit Lett. 2009;30(1):27–38. https://doi.org/10.1016/j.patrec.2008.08.010.
23. Choi HY, Park HE, Seo H, et al. Fasting plasma glucose and glycated hemoglobin cutoffs for predicting diabetes and prediabetes: the Korean genome and epidemiology study. J Korean Med Sci. 2018;33(50):e93. https://doi.org/10.3346/jkms.2018.33.e93.
24. Wai JH, Turner RM, Koeman J, et al. Outcome analysis for BI-RADS category 3 in the national mammography database. Diagn Interv Imaging. 2017;98(3):179–190. https://doi.org/10.1016/j.diii.2017.01.003.
25. He H, Garcia EA. Learning from imbalanced data. IEEE Trans Knowl Data Eng. 2009;21(9):1263–1284. https://doi.org/10.1109/TKDE.2008.239.
26. Hassanzad F, Fryback D, Hosmer D, et al. Optimal cut-point analysis: an updated review of methods in medical research. J Appl Stat. 2024;51(4):1222–1242. https://doi.org/10.1080/02664763.2022.2130717.
27. Naulaerts S, et al. The impact of feature selection on the performance of imbalanced binary classification problems. Oncotarget. 2017;8:109343–109353. (DOI could not be verified online; recommend confirming source.)
28. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.
29. Shi L, Reid LH, Jones WD, et al. The microarray quality control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. Nat Biotechnol. 2010;28(8):827–838. https://doi.org/10.1038/nbt.1665.
30. Su Z, Łabaj PP, Li S, et al. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the SEQC/MAQC-III consortium. Nat Biotechnol. 2014;32(9):903–914. https://doi.org/10.1038/nbt.2957.
31. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–35. https://doi.org/10.1002/1097-0142-(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
32. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 1993;39(4):561–577. https://doi.org/10.1093/clinchem/39.4.561.
33. Obuchowski NA. ROC analysis. AJR Am J Roentgenol. 2005;184(2):364–372. https://doi.org/10.2214/ajr.184.2.01840364.
34. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432. https://doi.org/10.1371/journal.pone.0118432.
35. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, von Luxburg U, Bengio S, et al., editors. Advances in Neural Information Processing Systems 30. Red Hook (NY): Curran Associates; 2017. p.4765–4774.
36. Więckowska B, Kubiak KB, Guzik P. Evaluating the three-level approach of the U-smile method for imbalanced binary classification. PLoS One. 2025;20(4):e0321661. https://doi.org/10.1371/journal.pone.0321661.

JMS Journal of Medical Science

Pulmonary Resistance Modification under treatment with Sacubitril/valsartan in Patients with Heart Failure with reduced ejection fraction (PRESENT-HF) – Study protocol and rationale for a multicentre randomised controlled trial

Ewa Straburzyńska-Migaj
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0002-0545-3370

Magdalena Dudek
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0001-6550-6182
Corresponding author: magdalena.dudek@usk.poznan.pl

Jacek Migaj
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0002-7962-3934

Agnieszka Bartczak-Rutkowska
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0002-6607-9191

Jarosław Hiczkiewicz
Clinical Department of Cardiology, Nowa Sól Multidisciplinary Hospital, Nowa Sól, Poland
https://orcid.org/0000-0002-1111-3372

Piotr Cygański
Department of Internal Medicine and Cardiology, City Hospital, Olsztyn, Poland
https://orcid.org/0000-0003-0698-6651

Agnieszka Tycińska
Department of Cardiology, Medical University of Bialystok, Białystok, Poland
https://orcid.org/0000-0001-7855-7261

Marek Gierlotka
Department of Cardiology, University Hospital, Institute of Medical Sciences, University of Opole, Opole, Poland
https://orcid.org/0000-0001-5639-2128

Ewa Kozielska-Nowalany
II Department of Cardiology in Zabrze, Faculty of Medical Sciences in Zabrze, Medical University of Silesia, Katowice, Poland
https://orcid.org/0000-0002-5977-513X

Jacek Kubica
Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
https://orcid.org/0000-0001-8250-754X

Małgorzata Tomaszewska
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0009-0006-7343-1401

Piotr Sawiński
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
–

Lech Paluszkiewicz
Clinic for Thoracic and Cardiovascular Surgery, Heart and Diabetes Centre North-Rhine Westphalia, Ruhr-University of Bochum, Bad Oeynhausen, Germany
https://orcid.org/0000-0002-9095-1277

Maciej Lesiak
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0003-2630-5016

Marta Kałużna-Oleksy
1st Department of Cardiology, Poznan University of Medical Sciences, Poznań, Poland
Clinical Hospital in Poznań, Poland
https://orcid.org/0000-0003-4048-6247

https://doi.org/10.20883/medical.e1332

Keywords: HFrEF, ARNI, sacubitril/valsartan, pulmonary hypertension, pulmonary hypertension due to left heart disease

Received 2025-07-01
Accepted 2025-10-10
Published 2025-12-31

How to Cite: Straburzyńska-Migaj E, Dudek M, Migaj J, Bartczak-Rutkowska A, Hiczkiewicz J, Cygański P, Tycińska A, Gierlotka M, Kozielska-Nowalany E, Kubica J, Tomaszewska M, Sawiński P, Paluszkiewicz L, Lesiak M, Kałużna-Oleksy M. Pulmonary Resistance Modification under treatment with Sacubitril/valsartan in Patients with Heart Failure with reduced ejection fraction (PRESENT-HF) – Study protocol and rationale for a multicentre randomised controlled trial. Journal of Medical Science. 2025 December;94(4);e1332. https://doi.org/10.20883/medical.e1332

© 2025 by the Author(s). This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) licencse. Published by Poznan University of Medical Sciences

ABSTRACT

Introduction. Heart failure with reduced ejection fraction (HFrEF) remains a major clinical challenge, partly due to suboptimal implementation of guideline-directed medical therapy and complications such as pulmonary hypertension (PH) that worsen symptoms and prognosis.

Material and methods. PRESENT-HF (ClinicalTrials.gov Identifier: NCT05487261, https://clinicaltrials.gov/ct2/show/NCT05487261) is a multicenter, randomised, double-blind, comparator-controlled clinical trial investigating whether sacubitril/valsartan, compared to enalapril, lowers pulmonary artery pressure and pulmonary vascular resistance as measured by right heart catheterisation in patients with HFrEF and secondary PH. The study will recruit approximately 230 patients, with a 1:1 randomisation to either sacubitril/valsartan or enalapril, and a 52-week follow-up that includes both in-hospital and out-of-hospital treatment phases.

Results. The primary endpoint is the change from baseline in mean pulmonary artery pressure and pulmonary vascular resistance; secondary outcomes include clinical events, quality of life, functional capacity, and safety.

Conclusions. This trial will offer new insights into targeted strategies for improving outcomes in patients with HFrEF complicated by PH, potentially expanding treatment options and informing future guidelines.

Introduction

Despite massive progress in the treatment of heart failure with reduced ejection fraction (HFrEF), prognosis and quality of life (QoL) are still poor. The reasons for this situation are multiple, including poor implementation of guideline-directed medical therapy (GDMT) and multimorbidity in many patients. A special problem in many patients is pulmonary hypertension (PH) associated with heart failure (HF), the true prevalence of which is unknown. It is estimated primarily based on indirect echocardiographic assessment [1,2]. Pulmonary hypertension related to left heart disease (PH-LHD) is the most prevalent among patients with PH and is related to a worse prognosis. A critical issue is PH in patients being evaluated for heart transplantation, as high pulmonary vascular resistance (PVR) is a contraindication for the procedure. Pulmonary hypertension in HFrEF has garnered the interest of numerous researchers investigating its pathophysiology and potential treatment. It is known that specific pharmacological agents used for Pulmonary Arterial Hypertension are contraindicated in the PH-LHD study, which aims to assess the effects of angiotensin receptor-neprilysin inhibitor (ARNI) treatment on RHC parameters in patients with HFrEF and secondary PH.

Methods and analysis

Trial design

PRESENT-HF is a multicenter, randomised, double-blind, comparator-controlled clinical trial that will include patients with chronic HFrEF (ClinicalTrials.gov Identifier: NCT05487261). Approximately 260 participants will be screened, resulting in around 230 patients being randomised across five centres. Subjects will be randomised to receive either an investigational medicinal product – sacubitril/valsartan or a comparator, enalapril.

The study will consist of 5 phases:

- Screening (Vs),
- Randomisation (V0 – day 0) – the eligible patients will be randomly assigned (1:1) to the intervention or comparator arms,
- In-hospital initiation of therapy (with wash-out strategy) (from day 0 to the day of discharge),
- Out-of-hospital phase – increasing the dose of the drug (up-titration phase),
- Out-of-hospital and hospital phase – follow-up on the target dose.

The observation period will span 52 weeks, consisting of a 4-week active up-titration phase followed by 48 weeks of follow-up (see **Figure 1**). Functional capacity is assessed by 6-minute walk distance, New York Heart Association

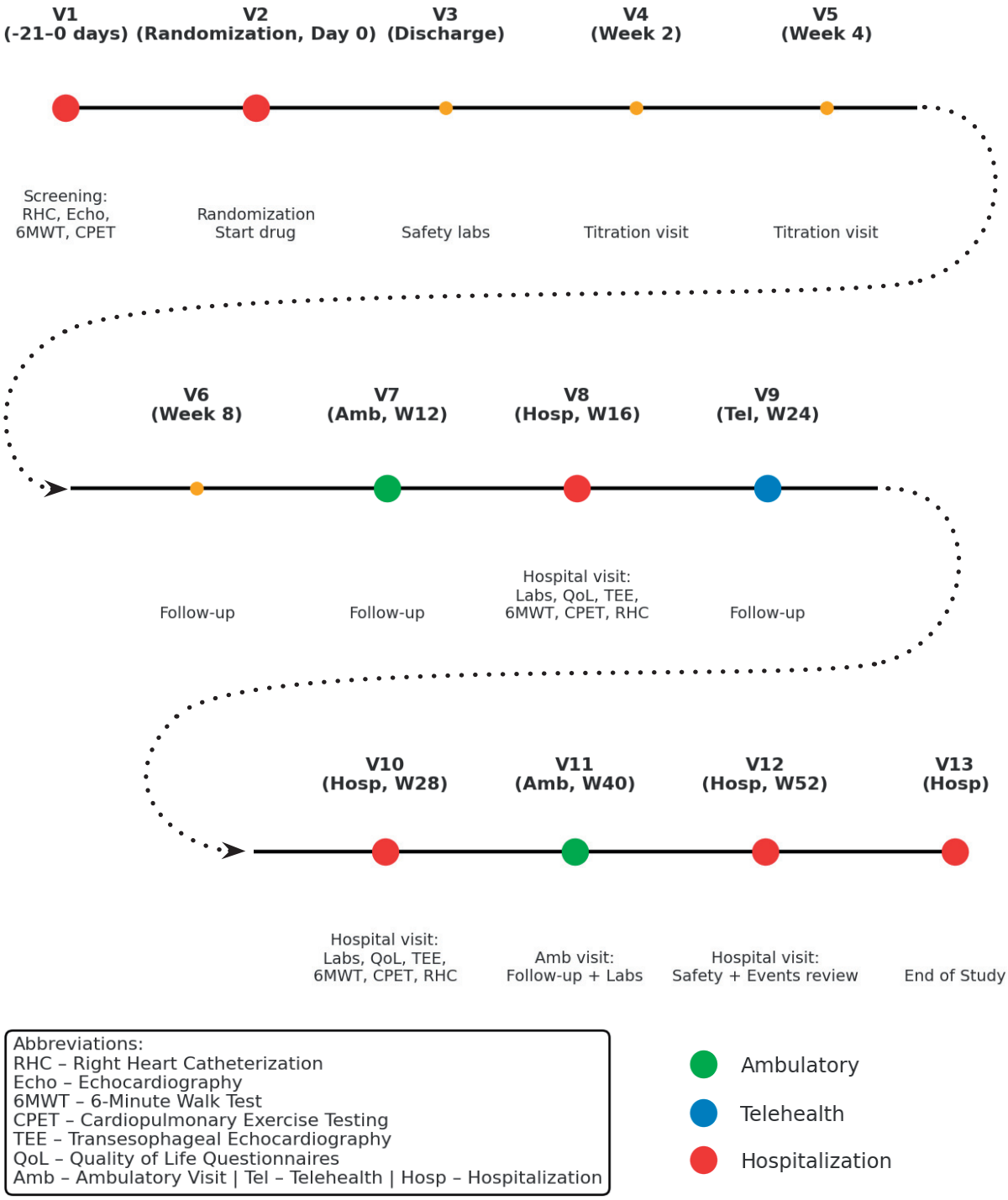


Figure 1. Present-HF: Visit Schedule (V1-V13) with Abbreviations Legend.

(NYHA) class (secondary endpoints) and cardiopulmonary exercise testing (CPET) as exploratory outcomes, with centralised analysis.

Primary and secondary objectives

The trial's objective is to assess the effect of sacubitril/valsartan therapy in comparison to angiotensin-converting enzyme inhibitors (ACE--I) – enalapril, on the parameters of right heart catheterisation in terms of reduction of pulmonary artery pressure (PAP) and PVR in patients with developed PH due to HfrEF.

Secondary objectives are to assess the treatment efficacy as evaluated by the incidence

of major adverse cardiac and cerebrovascular events (MACCE) as a composite endpoint, and separately, to determine the quality of life, and treatment safety and tolerance – assessment of Adverse Events (AE) and Serious Adverse Events (SAE) and their assessment in terms of severity and relationship to the tested substance. Additionally, as part of the exploratory part of the study, the influence of therapy on the selected biomarkers will be analysed.

The primary endpoint for efficacy assessment will be the change from baseline in mean pulmonary artery pressure (mPAP) and PVR.

Secondary endpoints:

1. Change from baseline in pulmonary wedge pressure (PWP) – measured by RHC.
2. Change from baseline in the diastolic pressure gradient (DPG; where DPG = diastolic mPAP – mean PWP).
3. Change meters from baseline in the 6-minute walk test (6MWT).
4. Evaluation of the parameters of the cardiopulmonary exercise test – peak VO_2, VE/VCO_2 slope.
5. Assessment of echocardiographic parameters (LVEF, LV volumes, TAPSE, RV function indices, estimated sPAP).
6. The incidence of the composite endpoint of MACCEs, such as death from all causes, cardiac death, hospitalisation due to worsening/decompensation of HF, heart transplantation, and the need for a left ventricular or biventricular assist device.
7. Hospitalisation or an unplanned visit to an Emergency Department or an unplanned outpatient visit related to HF.
8. The need for an unplanned intravenous administration of diuretics and/or an unplanned hospitalisation, an outpatient visit due to the need to administer intravenous diuretics or requiring an increase in the dose of diuretics >50% from baseline.
9. Quality of life assessment will be conducted between 0-52 weeks – Kansas City Cardiomyopathy Questionnaire (KCCQ-12), WHO (WHOQOL-BREF), SF-36 questionnaire, EQ-5D -3L questionnaire.
10. Assessment of the New York Heart Association (NYHA) classes – change from baseline.

Sample size calculation

Estimating the proportion of patients meeting the study criteria in the general population is difficult. According to the 2016 report of the Heart Failure Section of the Polish Cardiac Society, there are about 1.2 million patients with HF in Poland (2.6% of the population), with 220,000 new cases annually. Pulmonary hypertension occurs in 12–38% of HF patients, nearly half due to HfrEF, giving an estimated study-eligible population of up to 200,000 (0.5–1% of the general population).

Sample size calculations were based primarily on the primary endpoint, PVR (expected reduction from 2 to 1 Wood unit, SD 1), giving a minimum of 23 patients. Additional calculations were performed for 6MWT (n = 86 for a 50 m improvement) and for MACCE analysis using survival models (n = 89 for survival 0.90 vs. 0.70, expected deaths = 35). Taking all endpoints and potential data loss into account, the required sample size was set at 230 patients (including 5% dropout)

Study population – inclusion and exclusion criteria

The study population will consist of patients of both sexes aged 18 years or older with HFrEF in NYHA class II-IV functional classes, who have been diagnosed and treated for at least 3 months before inclusion in the study, and in whom the suspicion of PH has been confirmed based on right heart catheterisation. Participants will be recruited from five tertiary referral cardiology centres in Poland. Eligible patients are identified during routine care or planned hospitalisations for HF diagnostic evaluation. Before randomisation, all participants must undergo right heart catheterisation (RHC) to confirm pulmonary hypertension (PH). Recruitment will be coordinated centrally to ensure standardised screening, consent, and data collection procedures across sites. Patients will meet all inclusion criteria and do not meet any exclusion criteria (see **Supplementary Table 1**). The original definition of PH used a mean pulmonary artery pressure (mPAP) ≥25 mmHg; hence, according to the ESC 2022 guidelines, PH is diagnosed when mPAP is >20 [3].

Treatment protocol – interventional methods

Randomisation process will be performed centrally via an electronic automatic system IVRS as a part of eCRF. Numbered study treatment packs

will contain either IMP (S/V) or comparator (enalapril) according to a computer-generated randomisation plan. Each eligible subject will be randomly assigned to either active treatment (S/V) or comparator (enalapril) using a validated, centralised procedure via eCRF, which automates the random assignment of treatment groups to randomisation numbers (see **Figure 2**). Subject randomisation (1:1) will be determined by an algorithm including a random variable and accounting for the following stratification factors: sex, age (<70 years/≥70 years), HF aetiology (ischemic/ non-ischemic), HF duration (newly diagnosed at Index hospitalisation/known documented HF before Index hospitalisation). Eligible participants are randomised 1:1 to receive either sacubitril/valsartan (S/V) or enalapril. For patients previously treated with ACE inhibitors, a 36-hour washout is mandatory before the first S/V dose (see **Figure 2**). After the patient's last dose of the current therapy, randomisation occurs at Visit 2 (V2). A 36-hour washout period is required before the patient takes the first dose of the new investigational product (IP). Discharge is possible at least six hours after the first IP administration.

Dose adjustments are based on blood pressure, renal function, and potassium levels, which are assessed at each visit. Temporary down-titration is permitted for symptomatic hypotension, worsening renal function, or hyperkalemia.

Administration

Study medications are given orally, twice daily, in the morning and evening, preferably at the exact times each day. Patients are instructed to take doses with water, with or without food.

Follow-Up Schedule

- **In-hospital phase:** Initiation under supervision, with discharge allowed ≥6 hours after first dose if stable.
- **Outpatient phase:** Visits at 2 and 4 weeks for dose titration, then every 3 months until week 52.

Background therapy

All enrolled patients will be treated in accordance with the current ESC guidelines and receive standard chronic HF treatment. GDMT includes beta-blockers, diuretics, ACE-I or ARB, mineralocor-

Dosing and Up-Titration

Study Arm	Starting Dose	Titration Schedule	Target Dose
Sacubitril/Valsartan	24/26 mg twice daily (or 49/51 mg twice daily if baseline SBP ≥110 mmHg and eGFR ≥60 mL/min/1.73m²)	Dose doubled every 2 weeks as tolerated	97/103 mg twice daily
Enalapril	2.5–5 mg twice daily	Dose doubled every 2 weeks as tolerated	10 mg twice daily

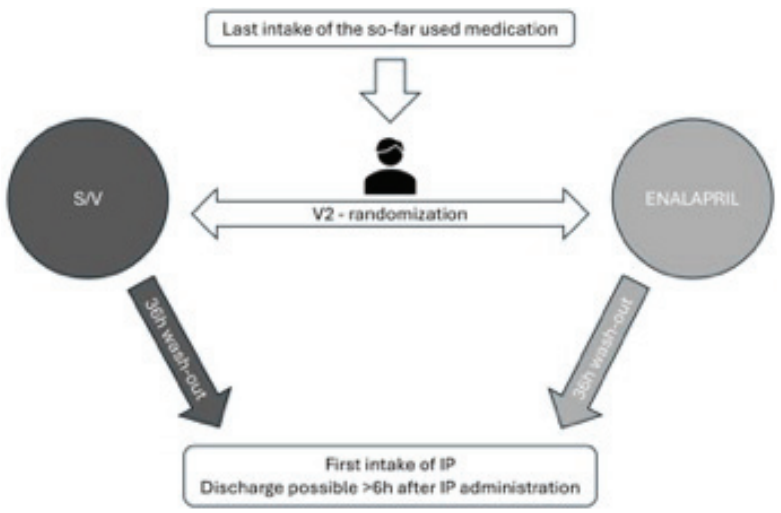


Figure 2.

ticoid receptor antagonists (MRA), and sodium--glucose cotransporter-2 inhibitors (SGLT2-I), if those above are not contraindicated. In patients with atrial fibrillation and/or coronary artery disease, the current standard therapy will be continued.

Ethical issues

This clinical trial received approval from the Ethics Committee of the Medical University of Poznań, Poland (approval number: 77/21, 4 Feb 2021).

The study will be conducted in accordance with the International Conference on Harmonisation Guidelines for Good Clinical Practice, the Declaration of Helsinki, European Directive 2001/20/EC, the Act of 6 September 2001 on Pharmaceutical Law in Poland, and applicable local health laws and authorities, as well as the Ethics Committee requirements.

Results

Statistical analysis

All data collected will be analysed descriptively. Standard descriptive statistical methods, including the number of patients, arithmetic mean, standard deviation, upper and lower quartiles, minimum, median and maximum, will be applied. For categorical variables, tables of frequencies (both absolute and relative) will be presented. The number of patients for whom the data are missing will be provided where appropriate.

Reported adverse events and comorbidities will be coded using the MedDRA dictionary (a version current at the time of the study initiation), and all adverse event summaries will present preferred terms and System Organ Class.

All data up to the time of study completion or withdrawal will be included in the analysis, regardless of treatment duration.

Results will be compared between study arms. For continuous outcomes, distribution will be first tested for normality using the Shapiro-Wilk test. Then, Student's t-test will be used to compare the distribution of normally distributed variables between study arms, and the Mann-Whitney test will be used otherwise. For categorical outcomes, Fisher's exact test of chi-squared test will be used, depending on the expected sizes of the categories. Time-to-event data will be analysed using survival analysis methods. Specifically, survival curves will be estimated using the Kaplan-Meier method and compared between study arms using the log-rank test.

Additionally, the effect of treatment (study arm) on the risk of event occurrence will be assessed using Cox regression. The level of statistical significance will be set at 0.05. Two-sided tests will be used.

Safety monitoring

All patients involved in the study will be monitored according to the visit schedule. On each visit, the patient will have blood samples taken for laboratory tests, among others, for creatinine and electrolytes. The dose level received by the patient will depend on the mean values of the systolic blood pressure (SBP), which is assessed on each visit.

The administration of the study medication will be stopped when any of the following occurs: significant, symptomatic and persistent hypotension (SBP<=90 mmHg), worsening of renal function (defined as an increase in serum creatinine ≥0.5 mg/dL and/or worsening of eGFR ≥ 25%) or development of hyperkaliemia (defined as K ≥ 5.5mEq/L/L). All adverse events that could be potentially associated with the drug will be identified and reported in the form of AE and SAE reports, in accordance with the ICH guideline for Good Clinical Practice. Additionally, the study will be monitored by an Independent Data Monitoring Committee (IDMC), which will also provide a therapy safety assessment throughout the entire study duration.

Discussion

Pulmonary hypertension associated with HFrEF influences symptoms, QoL and prognosis. The original definition of PH used a mean pulmonary artery pressure (mPAP) ≥25 mmHg. Findings from extensive studies have shown, for the first time, that a significant increase in mortality and hospitalisation risk arises in patients with mPAP > 20 mmHg. Hence, according to the ESC 2022 guidelines, PH is diagnosed when mPAP is greater than 20 mmHg [7]. However, despite this, there is still no specific treatment for PH-LHD. One of

the promising drugs is sacubitril/valsartan. S/V combines the effects of an angiotensin receptor blocker with a neprilysin inhibitor, which enhances the activity of the natriuretic peptide system and other vasoactive peptides, thus strengthening the protective effect of endogenous natriuretic peptides in the treatment of HF, an ARNI, functions as an activator of particulate guanylyl cyclase, leading to increased levels of vasoactive peptides and belongs to the peptide signal cascade through cyclic guanosine monophosphate (cGMP), producing significant antimitogenic and vasodilatory effects [10,11].

The main clinical confirmation of the effectiveness of S/V was provided by the PARADIGM--HF study from 2014, which showed that S/V was superior to enalapril in reducing the risk of death and hospitalisation for heart failure. It received the first class of recommendation, hence it constitutes the basic treatment for patients with HFrEF. However, no study has confirmed the influence of ARNI on PH that can complicate the course of HFrEF. There is no specific treatment for PH in HF.

According to the Prospective comparison of an ARNI with an ACE-I to Determine the Impact on Global Mortality and morbidity in Heart Failure (PARADIGM-HF)[10] trial, the S/V in comparison with enalapril showed a robust 20% relative risk reduction in cardiovascular mortality and hospitalization due to worsening of HF among the ambulatory HF patients with a left ventricular ejection fraction (LVEF) ≤40% (i.e., changed to ≤35% by an amendment to the protocol midtrial) and New York Heart Association functional class II-IV symptoms.

The trial was stopped early, according to prespecified rules, after a median follow-up of 27 months, because the boundary for an overwhelming benefit with S/V had been crossed, resulting in a reduction in the primary outcome in the study group.

Scientific reports indicate that S/V, in the absence of targeted therapy for PH-LHD, appears to improve outcomes in these patients significantly. The beneficial effects of NPs occur through a complex signalling system, which induces direct vasodilation, increases glomerular filtration, promotes natriuresis, reduces renin secretion, and exhibits antihypertrophic and antifibrotic myocardial effects.

Most of the evidence supporting the effects of S/V on PH derives from experimental and echocardiographic studies, mostly retrospective. In experimental studies of pulmonary hypertension in rats, treatment with S/V, but not valsartan, resulted in a significant reduction in right ventricular (RV) pressure (62 ± 4 vs. 46 ± 5 mmHg) compared with the placebo. This was associated with reduced pulmonary vascular wall thickness, increased lung levels of atrial natriuretic peptide, brain natriuretic peptide, and cGMP, and decreased plasma endothelin-1 levels. S/V reduces pulmonary pressures, vascular remodelling, and RV hypertrophy in a rat model of PH. It is suggested that it may be appropriate for the treatment of pulmonary hypertension and RV dysfunction [5].

In another experimental study in which rats were treated with monocrotaline or a hypoxic environment for 14 days to induce PH, S/V was given thereafter for the next 14 days. This study indicated that S/V reduced pulmonary pressure, adverse vascular remodelling and right ventricular hypertrophy in rats [14].

In the retrospective observational study of 93 patients with HFrEF with RV dysfunction based on echocardiographic parameters, it was found that S/V treatment was associated with significant improvements in the following RV function indicators: tricuspid annular plane systolic excursion, tri, including cuspid annular s' peak velocity (S'), RV fractional area change, and pulmonary artery systolic pr[15]. The authors suggested that this improvement may be independent of left heart dysfunction, and in patients with RV dysfunction and HFrEF, S/V may improve RV remodelling.

The systematic review and meta-analysis by Zhang et al. confirmed that S/V significantly enhances proper ventricular function and reduces PH in patients with HFrEF [16]. Across 10 observational studies involving 875 patients (mean age 62.2 years, 74% men), S/V led to significant improvements in key parameters: tricuspid annular plane systolic excursion increased by 1.26 mm (95% CI, 0.33–2.18 mm; P = 0.008), tricuspid annular peak systolic velocity rise by 0.85 cm/s (95% CI, 0.25–1.45 cm/s; P = 0.005), and estimated systolic pulmonary arterial pressure decreased by 7.21 mm Hg (95% CI, 5.38–9.03 mm Hg; P < 0.001). These improvements suggest that S/V positively affects proper ventricular performance and pulmonary pressures, effects that

are not solely dependent on left ventricular reverse remodelling.

The prospective observational study by Polito et al. evaluated the effects of S/V on clinical outcomes and echocardiographic parameters in a real-world cohort of 90 patients with HFrEF [17]. Over a 12-month follow-up, S/V treatment significantly reduced the risk of the composite primary outcome (cardiac death and HF rehospitalisation) with a hazard ratio (HR) of 0.31 (95% CI, 0.11–0.83; P = 0.019). S/V also lowered the risk of HF rehospitalisation (HR: 0.27; 95% CI, 0.08–0.94; P = 0.039) compared to a control group of patients receiving standard medical therapy. Additionally, improvements were observed in NYHA class, LVEF and systolic PAP at 6 months. Importantly, S/V preserved renal function and reduced the need for furosemide at 6 and 12 months. These findings suggest that S/V offers significant clinical benefits in reducing adverse outcomes and improving functional and echocardiographic parameters in patients with HFrEF.

Another prospective study by Yamaguchi et al., based on echocardiographic assessments after the 6MWT, examined the effects of ARNI on pulmonary circulation in 39 HF patients [18]. Significant improvements were observed in LV volume (160.7 ± 49.6 mL vs 136.0 ± 54.3 mL, P < 0.001) and LVEF (37.6 ± 11.3% vs 44.9 ± 11.5%, P < 0.001) for the 31 patients who completed both baseline and follow-up assessments. Follow-up 6MWT stress echocardiography revealed an increase in the 6MWT distance from 380 m to 430 m (P = 0.003). Significantly, the ratio of mean PAP to cardiac output decreased with ARNI treatment (6.9 mmHg/L/min vs 2.8 mmHg/L/min, P = 0.002), indicating improved pulmonary pressure-flow dynamics. These findings suggest that ARNI treatment not only enhances LV function and exercise capacity but also optimises the pulmonary pressure-flow relationship in heart failure patients.

In the two cases reported by De Simone et al., the use of ARNI was noted in patients with advanced HFrEF. This has resulted in an improvement in pH, as measured during RHC, and, consequently, in both clinical status and prognosis. These two cases demonstrate that natriuresis and vasodilation may play a crucial role in reducing pulmonary pressures [19].

ARNIMEMS-HFpEF was a single-arm interventional study with a short observation period (7 days), investigating patients with heart failure with preserved ejection fraction (HFpEF), [20]. The sample size consisted of 14 patients. The results indicated a significant reduction in mPAP by 4.14 ± 5.7 mmHg after 7 days of S/V treatment (P = 0.019), as assessed using the CardioMEMS device. Furthermore, a notable improvement in functional capacity was observed in the 6MWT, with distances increasing from 270.6 ± 101.3 meters to 298.3 ± 88.4 meters (P < 0.001) following the ARNI therapy.

Another retrospective case series involved 18 patients with HFrEF and pulmonary pressures as measured using the CardioMEMS device over 5 days after S/V initiation [7]. The study showed a significant reduction in median (interquartile range) mPAP by −3.6 mmHg (−9.8, −0.7) post-ARNI initiation. Additionally, the baseline median (interquartile range) PVR was significantly higher at 357 (320, 548) dyn/s/cm^5 compared to 137 (73, 172) dyn/s/cm^5 (P = 0.001), suggesting a marked reduction in pulmonary resistance following treatment. These studies emphasise the early hemodynamic benefits of S/V, including cuts in pulmonary pressures and improvements in functional capacity, which contribute to a broader understanding of S/V therapeutic effects in both HFpEF and HFrEF populations.

Conclusions

Pulmonary hypertension secondary to HFrEF remains a major contributor to symptom burden, hospitalisations, and adverse prognosis, yet no specific pharmacological treatment has been proven to modify its course. Sacubitril/valsartan, a cornerstone therapy for HFrEF, has demonstrated robust clinical benefits in reducing mortality and hospitalisations for heart failure. Evidence from preclinical models, echocardiographic studies, and limited invasive hemodynamic data suggests that sacubitril/valsartan may also reduce pulmonary pressures, improve proper ventricular function, and favorably remodel the pulmonary vasculature.

The PRESENT-HF trial is the first randomised, double-blind, controlled study to prospectively investigate the effect of sacubitril/valsartan on invasively measured pulmonary artery pressures and pulmonary vascular resistance in patients

with HFrEF and confirmed PH. By combining gold-standard hemodynamic assessment with clinical, functional, and quality-of-life endpoints, this trial aims to address a critical knowledge gap and expand the therapeutic role of ARNI therapy to include the management of PH-LHD.

Declarations

Conflict of interest

Ewa Straburzyńska-Migaj received consultancy fees and/or honoraria from Novartis, Boehringer Ingelheim, Pfizer, Bayer, Abbott, Servier, AstraZeneca, and Bausch Health; Agnieszka Tycińska – consultancy fees and/or honoraria from Abiomed, Boehringer Ingelheim, Pfizer, Bayer, Servier, AstraZeneca, and Bausch Health; Marta Kałużna-Oleksy's education and lecture collaboration with Novartis; Magdalena Dudek, Jacek Migaj, Agnieszka Bartczak-Rutkowska, Jarosław Hiczkiewicz, Piotr Cygański, Marek Gierlotka, Ewa Kozielska-Nowalany, Jacek Kubica, Małgorzata Tomaszewska, Piotr Sawiński, Lech Paluszkiewicz, and Maciej Lesiak declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors contribiution

Conceptualization, E.S.-M., M.K.-O., M.D.; methodology, E.S.-M., M.K.-O., J.M., M.D.; software, J.M., validation, M.D., M.K.-O.; formal analysis, M.D., M.K.-O.; investigation, E.S.-M., M.D, A.T, M.K.-O, A.B.-R., J. H., P.C, M.G., E. K.-N., J.K., M.T., P.S, L.P., M.L.; resources, M.D.; data curation, M.D. and J.M.; writing—original draft preparation, E.S.-M., M.K.-O, M.D.; writing—review and editing, E.S.-M., M.D, A.T, M.K.-O, A.B.-R., J. H., P.C, M.G., E. K.-N., J.K., M.T., P.S, L.P., M.L; visualization, E.S.-M., M.K.-O., M.D.; supervision, E.S.-M. and M.L.; project administration, E.S-M., M.K.-O. All authors have approved the final version of the manuscript.

Funding

This research received financial support from the Medical Research Agency, Poland (project no. 2019/ABM/01/00078-00).

References

1. Aras MA, Psotka MA, De Marco T. Pulmonary Hypertension Due to Left Heart Disease: an Update. Curr Cardiol Rep. 2019 May 27;21(7):62. https://doi.org/10.1007/s11886-019-1149-1
2. Ameri P, Mercurio V, Pollesello P, Anker MS, Backs J, Bayes-Genis A, et al. A roadmap for therapeutic discovery in pulmonary hypertension associated with left heart failure. A scientific statement of the Heart Failure Association (HFA) of the ESC and the ESC Working Group on Pulmonary Circulation & Right Ventricular Function. Eur J Heart Fail. 2024 Apr;26(4):707–29. https://doi.org/10.1002/ejhf.3236
3. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Eur Heart J. 2022 Oct 11;43(38):3618–731. https://doi.org/10.1093/eurheartj/ehac237
4. Guazzi M, Ghio S, Adir Y. Pulmonary Hypertension in HFpEF and HFrEF: JACC Review Topic of the Week. J Am Coll Cardiol. 2020 Sep 1;76(9):1102–11. https://doi.org/10.1016/j.jacc.2020.06.069
5. Clements RT, Vang A, Fernandez-Nicolas A, Kue NR, Mancini TJ, Morrison AR, et al. Treatment of Pulmonary Hypertension With Angiotensin II Receptor Blocker and Neprilysin Inhibitor Sacubitril/Valsartan. Circ Heart Fail. 2019 Nov;12(11):e005819. https://doi.org/10.1161/CIRCHEARTFAILURE.119.005819
6. Landolfo M, Piani F, Esposti DD, Cosentino E, Bacchelli S, Dormi A, et al. Effects of sacubitril valsartan on clinical and echocardiographic parameters of outpatients with heart failure and reduced ejection fraction. IJC Heart & Vasculature. 2020 Dec 1;31:100656. https://doi.org/10.1016/j.ijcha.2020.100656
7. Tran JS, Havakuk O, McLeod JM, Hwang J, Kwong HY, Shavelle D, et al. Acute pulmonary pressure change after transition to sacubitril/valsartan in patients with heart failure reduced ejection fraction. ESC Heart Fail. 2021 Apr;8(2):1706–10. https://doi.org/10.1002/ehf2.13225
8. Galiè N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Respir J. 2015 Oct;46(4):903–75. https://doi.org/10.1093/eurheartj/ehv317
9. Xu Y, Yang B, Hui J, Zhang C, Bian X, Tao M, et al. The emerging role of sacubitril/valsartan in pulmonary hypertension with heart failure. Front Cardiovasc Med. 2023;10:1125014. https://doi.org/10.3389/fcvm.2023.1125014
10. McMurray JJV, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med. 2014 Sep 11;371(11):993–1004. https://doi.org/10.1056/NEJMoa1409077
11. Kusaka H, Sueta D, Koibuchi N, Hasegawa Y, Nakagawa T, Lin B, et al. LCZ696, Angiotensin II Receptor-Neprilysin Inhibitor, Ameliorates High-Salt-Induced Hypertension and Cardiovascular Injury More Than Valsar-

tan Alone. Am J Hypertens. 2015 Dec;28(12):1409–17. https://doi.org/10.1093/ajh/hpv015
12. Solomon SD, Claggett B, Desai AS, Packer M, Zile M, Swedberg K, et al. Influence of Ejection Fraction on Outcomes and Efficacy of Sacubitril/Valsartan (LCZ696) in Heart Failure with Reduced Ejection Fraction: The Prospective Comparison of ARNI with ACEI to Determine Impact on Global Mortality and Morbidity in Heart Failure (PARADIGM-HF) Trial. Circ Heart Fail. 2016 Mar;9(3):e002744. https://doi.org/10.1161/CIRCHEARTFAILURE.115.002744
13. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021 Sep 21;42(36):3599–726. https://doi.org/10.1002/ejhf.2333
14. Liu S, Wang Y, Lu S, Hu J, Zeng X, Liu W, et al. Sacubitril/valsartan treatment relieved the progression of established pulmonary hypertension in rat model and its mechanism. Life Sci. 2021 Feb 1;266:118877. https://doi.org/10.1016/j.lfs.2020.118877
15. Yang Y, Shen C, Lu J, Fu G, Xiong C. Sacubitril/Valsartan in the Treatment of Right Ventricular Dysfunction in Patients With Heart Failure With Reduced Ejection Fraction: A Real-world Study. J Cardiovasc Pharmacol. 2022 Feb 1;79(2):177–82. https://doi.org/10.1097/FJC.0000000000001162
16. Zhang J, Du L, Qin X, Guo X. Effect of Sacubitril/Valsartan on the Right Ventricular Function and Pulmonary Hypertension in Patients With Heart Failure With Reduced Ejection Fraction: A Systematic Review and Meta-Analysis of Observational Studies. J Am Heart Assoc. 2022 May 3;11(9):e024449. https://doi.org/10.1161/JAHA.121.024449
17. Polito MV, Silverio A, Rispoli A, Vitulano G, Auria FD, De Angelis E, et al. Clinical and echocardiographic benefit of Sacubitril/Valsartan in a real-world population with HF with reduced ejection fraction. Sci Rep. 2020 Apr 20;10(1):6665. https://doi.org/10.1038/s41598-020-63801-2
18. Yamaguchi N, Hirata Y, Nishio S, Takahashi T, Saijo Y, Kadota M, et al. Pulmonary pressure-flow responses to exercise in heart failure treated with angiotensin receptor neprilysin inhibitor. Int J Cardiol. 2024 Apr 1;400:131789. https://doi.org/10.1016/j.ijcard.2024.131789
19. De Simone V, Guarise P, Zanotto G, Morando G. Reduction in pulmonary artery pressures with use of sacubitril/valsartan. Journal of Cardiology Cases. 2019 Nov 1;20(5):187–90. https://doi.org/10.1016/j.jccase.2019.08.006
20. Codina P, Domingo M, Barceló E, Gastelurrutia P, Casquete D, Vila J, et al. Sacubitril/valsartan affects pulmonary arterial pressure in heart failure with preserved ejection fraction and pulmonary hypertension. ESC Heart Fail. 2022 Aug;9(4):2170–80. https://doi.org/10.1002/ehf2.13952

Supplementary Table 1. Inclusion and exclusion criteria for the PRESENT-HF study.

Inclusion criteria	Exclusion criteria
1. Age ≥18 years of age who are able to complete and sign the informed consent form. 2. HF patients in NYHA II-IV in whom RHC reveals post-capillary pulmonary hypertension; both IpcPH as well as CpcPH(Humbert et al., 2022) 3. Stable patients haemodynamics, which is defined as no change in diuretic use for at least 4 weeks prior to study entry. 4. HF during optimal treatment with ACE-I/ ARB, beta-blocker, MRA, SGLT2-I except in cases where the above-mentioned treatment was contraindicated or not tolerated. 5. Understanding and acceptance of the research assumptions and methods, and signing the informed consent by the patient.	1. Current treatment with S/V. 2. Cardiogenic shock. 3. Current treatment with sildenafil. 4. Patients ineligible or contraindicated for treatment with S/V. 5. Patients with a history of angioedema. 6. Patients who have had a heart transplant or have had a circulatory support device. 7. Patient on the urgent list for heart transplantation. 8. Isolated right HF secondary to lung disease. 9. Documented untreated significant ventricular arrhythmia with syncope within the previous 3 months. 10. Symptomatic bradycardia or second or third degree atrioventricular block not protected by a pacemaker. 11. Factors that prevent RHC testing (e.g. very serious condition of the patient that makes it impossible to lie down, cardiogenic shock, allergy to contrast agents, etc.). 12. Pregnant or lactating women. 13. Women of childbearing age, defined as the physiological possibility of becoming pregnant, unless using two methods of contraception. 14. Acute coronary syndrome, including myocardial infarction (STEMI, NSTEMI), a ondition with carotid revascularization or major cardiovascular surgery in the last 30 days. 15. Stroke or TIA within the last 3 months. 16. Previous CRT implantation in the last 3 months or planning for CRT implantation. 17. Life expectancy < 6 months. 18. Severe renal failure, eGFR < 30 ml / min / 1.73 m2 (calculated according to the MDRD formula). 19. Serum potassium > 5.2 mEqL. 20. Liver failure or elevated liver transaminases (total bilirubin > 3 mg / dL and/or ALT and/or AST ≥3x ULN). 21. A major surgery planned within 6 months of randomization. 22. Planned coronary angioplasty or pacemaker/ICD/CRT implantation within the next 6 months. 23. Severe primary valve disease (NOT secondary mitral regurgitation) or obstructive hypertrophic cardiomyopathy. 24. The presence of a malignant neoplasm of any organ system, i.e. clinical signs or no stable remission for at least 3 years after the end of the last treatment, with the exception of non-invasive basal cell carcinoma, squamous cell carcinoma of the skin or cervical epithelial dysplasia. 25. Diseases that significantly reduce physical performance: – a) severe COPD putting off oxygen therapy, – b) severe asthma, – c) morbid obesity (BMI>40 kg/m^2), – d) significant lower limb atherosclerosis with intense intermittent claudication. 26. Uncontrolled hypertension (SBP>170 mmHg and / or DBP>100 mmHg). 27. Symptomatic hypotension (SPB < 90 mmHg). 28. Any situation that may make it impossible to perform the research in accordance with the protocol or express written consent in the opinion of the researcher, including abuse of alcohol, drugs or other psychoactive substances. 29. Participation in a study with a device or medicinal product within 3 months prior to randomization or 5 half-lives, whichever is longer, prior to the screening visit.

Abbreviations: ACE-I – Angiotensin-Converting Enzyme Inhibitors; ALT – Alanine Aminotransferase; ARB – Angiotensin Receptor Blockers; AST – Aspartate Aminotransferase; BMI – Body Mass Index; CpcPH – Combined Post- and Pre-capillary Pulmonary Hypertension; COPD – Chronic Obstructive Pulmonary Disease; CRT – Cardiac Resynchronization Therapy; DBP – Diastolic Blood Pressure; eGFR – Estimated Glomerular Filtration Rate; ESC – European Society of Cardiology; HF – Heart Failure; ICD – Implantable Cardioverter-Defibrillator; IpcPH – Isolated Post-capillary Pulmonary Hypertension; MDRD – Modification of Diet in Renal Disease; MRA – Mineralocorticoid Receptor Antagonists; NSTEMI – Non-ST-Elevation Myocardial Infarction; NYHA – New York Heart Association; mPAP – Mean Pulmonary Artery Pressure; PAWP – Pulmonary Artery Wedge Pressure; PH – Pulmonary Hypertension; PVR – Pulmonary Vascular Resistance; RHC – Right Heart Catheterization; S/V – Sacubitril/Valsartan (sakubitryl/walsartan); SBP – Systolic Blood Pressure; SGLT2-I – Sodium-Glucose Cotransporter-2 Inhibitors; STEMI – ST-Elevation Myocardial Infarction; TIA – Transient Ischemic Attack; ULN – Upper Limit of Normal; WU – Wood Units.

Journal of Medical Science (JMS) is a PEER-REVIEWED, OPEN ACCESS journal that publishes original research articles and reviews which cover all aspects of clinical and basic science research. The journal particularly encourages submissions on the latest achievements of world medicine and related disciplines. JMS is published quarterly by Poznan University of Medical Sciences.

ONLINE SUBMISSION:
Manuscripts should be submitted to the Editorial Office by an e-mail attachment: nowinylekarskie@ump.edu.pl. You do not need to mail any paper copies of your manuscript.

All submissions should be prepared with the following files:
- Cover Letter
- Manuscript
- Tables
- Figures
- Supplementary Online Material

COVER LETTER: Manuscripts must be accompanied by a cover letter from the author who will be responsible for correspondence regarding the manuscript as well as for communications among authors regarding revisions and approval of proofs. The cover letter should contain the following elements: (1) the full title of the manuscript, (2) the category of the manuscript being submitted (e.g. Original Article, Brief Report), (3) the statement that the manuscript has not been published and is not under consideration for publication in any other journal, (4) the statement that all authors approved the manuscript and its submission to the journal, and (5) a list of at least two referees.

MANUSCRIPT: Journal of Medical Science publishes Original Articles, Brief Reports, Review articles, Mini-Reviews, Images in Clinical Medicine and The Rationale and Design and Methods of New Studies. From 2014, only articles in English will be considered for publication. They should be organized as follows: Title page, Abstract, Introduction, Materials and Methods, Results, Discussion, Acknowledgments, Conflict of Interest, References and Figure Legends. All manuscripts should be typed in Arial or Times New Roman font and double spaced with a 2,5 cm (1 inch) margin on all sides. They should be saved in DOC, DOCX, ODT, RTF or TXT format. Pages should be numbered consecutively, beginning with the title page.

Ethical Guidelines
Authors should follow the principles outlined in the Declaration of Helsinki of the World Medical Association (www.wma.net). The manuscript should contain a statement that the work has been approved by the relevant institutional review boards or ethics committees and that all human participants gave informed consent to the work. This statement should appear in the Material and Methods section. Identifying information, including patients' names, initials, or hospital numbers, should not be published in written descriptions, illustrations, and pedigrees. Studies involving experiments with animals must be conducted with approval by the local animal care committee and state that their care was in accordance with institution and international guidelines.

Authorship
According to the International Committee on Medical Journal Ethics (ICMJE), an author is defined as one who has made substantial contributions to the conception and development of a manuscript. Authorship should be based on all of the following: 1) substantial contributions to conception and design, data analysis and interpretation; 2) article drafting or critical advice for important intellectual content; and 3) final approval of the version to be published. All other contributors should be listed as acknowledgments. All submissions are expected to comply with the above definition.

Conflict of Interest
The manuscript should contain a conflict of interest statement from each author. Authors should disclose all financial and personal relationships that could influence their work or declare the absence of any conflict of interest. Author's conflict of interest should be included under Acknowledgements section.

Abbreviations
Abbreviations should be defined at first mention, by putting abbreviation between brackets after the full text. Ensure consistency of abbreviations throughout the article. Avoid using them in the title and abstract. Abbreviations may be used in tables and figures if they are defined in the table footnotes and figure legends.

Trade names
For products used in experiments or methods (particularly those referred to by a trade name), give the manufacturer's full name and location (in parentheses). When possible, use generic names of drugs.

Title page
The first page of the manuscript should contain the title of the article, authors' full names without degrees or titles, authors' institutional affiliations including city and country and a running title, not exceeding 40 letters and spaces. The first page should also include the full postal address, e-mail address, and telephone and fax numbers of the corresponding author.

Abstract
The abstract should not exceed 250 words and should be structured into separate sections: Background, Methods, Results and Conclusions. It should concisely state the significant findings without reference to the rest of the paper. The abstract should be followed by a list of 3 to 6 Key words. They should reflect the central topic of the article (avoid words already used in the title).

The following categories of articles can be proposed to the Journal of Medical Science:

ORIGINAL RESEARCH

Original articles: Manuscripts in this category describe the results of original research conducted in the broad area of life science and medicine. The manuscript should be presented in the format of Abstract (250-word limit), Keywords, Introduction, Material and Methods, Results, Discussion, Perspectives, Acknowledgments and References. In the Discussion section, statements regarding the importance and novelty of the study should be presented. In addition, the limitations of the study should be articulated. The abstract must be structured and include: Objectives, Material and Methods, Results and Conclusions. Manuscripts cannot exceed 3500 words in length (excluding title page, abstract and references) and contain no more than a combination of 8 tables and/or figures. The number of references should not exceed 45.

Brief Reports: Manuscripts in this category may present results of studies involving small sample sizes, introduce new methodologies, describe preliminary findings or replication studies. The manuscript must follow the same format requirements as full length manuscripts. Brief reports should be up to 2000 words (excluding title page, abstract and references) and can include up to 3 tables and/or figures. The number of references should not exceed 25.

REVIEW ARTICLES

Review articles: These articles should describe recent advances in areas within the Journal's scope. Review articles cannot exceed 5000 words length (excluding title page, abstract and references) and contain no more than a combination of 10 tables and/or figures. Authors are encouraged to restrict figures and tables to essential data that cannot be described in the text. The number of references should not exceed 80.

A THOUSAND WORDS ABOUT... is a form of Mini-Reviews. Manuscripts in this category should focus on latest achievements of life science and medicine. Manuscripts should be up to 1000 words in length (excluding title page, abstract and references) and contain up to 5 tables and/or figures and up to 25 most relevant references. The number of authors is limited to no more than 3.

OTHER SUBMISSIONS

Invited Editorials: Editorials are authoritative commentaries on topics of current interest or that relate to articles published in the same issue. Manuscripts should be up to 1500 words in length. The number of references should not exceed 10. The number of authors is limited to no more than 2.

Images in Clinical Medicine: Manuscripts in this category should contain one distinct image from life science or medicine. Only original and high-quality images are considered for publication. The description of the image (up to 250 words) should present relevant information like short description of the patient's history, clinical findings and course, imaging techniques or molecular biology techniques (e.g. blotting techniques or immunostaining). All labeled structures in the image should be described and explained in the legend. The number of references should not exceed 5. The number of authors is limited to no more than 5.

The Rationale, Design and Methods of New Studies: Manuscripts in this category should provide information regarding the grants awarded by different founding agencies, e.g. National Health Institute, European Union, National Science Center or National Center for Research and Development. The manuscript should be presented in the format of Research Project Objectives, Research Plan and Basic Concept, Research Methodology, Measurable Effects and Expected Results. The article should also contain general information about the grant: grant title, keywords (up to five), name of the principal investigator and co-investigators, founding source with the grant number, Ethical Committee permission number, code in clinical trials (if applicable). Only grant projects in the amount over 100,000 Euro can be presented. Manuscripts should be up to 2000 words in length (excluding references) and can include up to 5 tables and/or figures. The abstract should not exceed 150 words. The number of authors is limited to the Principal Investigator and Co-investigators.

Acknowledgements

Under acknowledgements please specify contributors to the article other than the authors accredited. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.). Also acknowledge all sources of support (grants from government agencies, private foundations, etc.). The names of funding organizations should be written in full.

References

All manuscripts should use the 'Elsevier – Vancouver (no "et al.")' style for references. References should be numbered consecutively in the order in which they appear in the text and listed at the end of the paper**.** References cited only in Figures/Tables should be listed in the end. Reference citations in the text should be identified by Arabic numbers in square brackets. Some examples:

This result was later contradicted by Smith and Murray [3].
Smith [8] has argued that...
Multiple clinical trials [4–6, 9] show...

Journal names should be abbreviated according to Index Medicus. If available alwaysprovide Digital Object Identifier (DOI) or PubMed Identifier (PMID) for every reference.

Some examples

Standard journal articles

1. Petrova NV, Kashirskaya NY, Vasilyeva TA, Kondratyeva EI, Marakhonov AV, Macek Jr M, Ginter EK, Kutsev SI, Zinchenko RA. Characteristics of the L138ins (p.Leu138dup) mutation in Russian cystic fibrosis patients. JMS [Internet]. 2020 Mar 31;89(1):e383. doi: 10.20883/medical.383.

Books

Personal author(s)

1. Rang HP, Dale MM, Ritter JM, Moore PK. Pharmacology. 5th ed. Edinburgh: Churchill Livingstone; 2003.

Editor(s) or compiler(s) as authors

2. Beers MH, Porter RS, Jones TV, Kaplan JL, Berkwits M (editors). The Merck manual of diagnosis and therapy. 18th ed. Whitehouse Station (NJ): Merck Research Laboratories; 2006.

Chapter in the book

1. Phillips SJ, Whisnant JP. Hypertension and stroke. In: Laragh JH, Brenner BM, editors. Hypertension: pathophysiology, diagnosis, and management. 2nd ed. New York: Raven Press; 1995. p. 465–478.

TABLES: Tables should be typed on sheets separate from the text (each table on a separate sheet). They should be numbered consecutively with Arabic numerals. Tables should always be cited in text (e.g. table 2) in consecutive numerical order. Each table should include a compulsory, concise explanatory title and an explanatory legend. Footnotes to tables should be typed below the table body and referred to by superscript lowercase letters. No vertical rules should be used. Tables should not duplicate results presented elsewhere in the manuscript (e.g. in figures).

FIGURES: All illustrations, graphs, drawings, or photographs are referred to as figures and must be uploaded as separate files when submitting a manuscript. Figures should be numbered in sequence with Arabic numerals. They should always be cited in text (e.g. figure 3) in consecutive numerical order. Figures for publication must only be submitted in high-resolution TIFF or EPS format (minimum 300 dpi resolution). Each figure should be self-explanatory without reference to the text and have a concise but descriptive legend. All symbols and abbreviations used in the figure must be defined, unless they are common abbreviations or have already been defined in the text. Figure Legends must be included after the reference section of the Main Text.

Color figures: Figures and photographs will be reproduced in full colour in the online edition of the journal. In the paper edition, all figures and photographs will be reproduced as black-and-white.

SUPPLEMENTARY ONLINE MATERIAL: Authors may submit supplementary material for their articles to be posted in the electronic version of the journal. To be accepted for posting, supplementary materials must be essential to the scientific integrity and excellence of the paper. The supplementary material is subject to the same editorial standards and peer-review procedures as the print publication.

Review Process

All manuscripts are reviewed by the Editor-in-Chief or one of the members of the Editorial Board, who may decide to reject the paper or send it for external peer review. Manuscripts accepted for peer review will be blind reviewed by at least two experts in the field. After peer review, the Editor-in-Chief will study the paper together with reviewer comments to make one of the following decisions: accept, accept pending minor revision, accept pending major revision, or reject. Authors will receive comments on the manuscript regardless of the decision. In the event that a manuscript is accepted pending revision, the author will be responsible for completing the revision within 60 days.

Copyright

The copyright to the submitted manuscript is held by the Author(s), who grants the Journal of Medical Science (JMS) a nonexclusive licence to use, reproduce, and distribute the work, including for commercial purposes.