Understanding, evaluation and translation of health-related research findings from science to end-users

Bralić, Nensi

UNIVERSITY OF SPLIT
SCHOOL OF MEDICINE

NENSI BRALIĆ, DMD

UNDERSTANDING, EVALUATION AND TRANSLATION OF HEALTH-RELATED RESEARCH FINDINGS FROM SCIENCE TO END-USERS

DOCTORAL DISSERTATION

MENTOR:
Asst. Prof. IVAN BULJAN

SPLIT, 2024

The research described in this dissertation was conducted at the Department of Research in Biomedicine and Health of the University of Split School of Medicine and as part of a Croatian Science Foundation project titled “Professionalism in Healthcare: Decision-Making in Practice and Science – ProDeM” (Grant agreement No. IP-2019-04-4882).

Published scientific paper on which the doctoral dissertation is based:

1) **Bralić N**, Buljan I. The association between research design and the perceived treatment effectiveness: a cross-sectional study. *Front Med (Lausanne)*. 2023 Dec 22;10:1220999. doi: 10.3389/fmed.2023.1220999 (Impact factor 5.058)

2) **Bralić N**, Mijatović A, Marušić A, Buljan I. Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study. *Sci Rep*. 2024 Mar 12;14(1):6016. doi: 10.1038/s41598-024-56727-6 (Impact factor 4.562)

ACKNOWLEDGEMENTS

I want to convey my heartfelt gratitude to my mentor, Ivan Buljan, for his unwavering support, guidance, and precious insights during the entire process. His expertise, encouragement, and innovative suggestions have been instrumental in deciding the direction and quality of my doctoral dissertation.

My sincerest appreciation goes to Professor Ana Marušić, whose constant support and encouraging work environment have been important in my professional and academic growth. Her guidance and opportunities provided by her leadership have been critical to the success of this project.

I am eternally grateful to my husband, Jakša, who has been a constant source of support, patience, and understanding for me. His unwavering confidence in my abilities has made my academic path rewarding and fulfilling.

My deepest appreciation goes to my family for their unwavering love, understanding, and encouragement. Their love and encouragement have provided me with the emotional support I needed to face the obstacles of academia.

TABLE OF CONTENTS

List of Abbreviations

3ie	International Initiative for Impact Evaluation
BERT	Bidirectional Encoder Representations from Transformers
CI	Confidence interval
CME	Continuing Medical Education
DMD	Doctor of dental medicine
EBM	Evidence-based medicine
IQR	Interquartile range
LIWC	Linguistic Inquiry and Word Count
MD	Medical doctor
Md	Median
NIPH	Norwegian Institute of Public Health
NLP	Natural Language Processing
NRC	National Research Council
OR	Odds ratio
PhD	Doctor of Philosophy
PLS	Plain language summary
RCT	Randomized controlled trial
SMOG	Simple Measure of Gobbledygook
SR	Systematic review
WHO	World Health Organization

1 INTRODUCTION

1.1 The Importance of Health-Related Research Findings

Health-related research findings play an essential role in understanding numerous elements of health and well-being. As such, they contribute significantly to advancing medical knowledge by abetting the discovery of disease mechanisms, identifying risk factors, and investigating new treatment options to improve treatment outcomes. Fundamentally, health-related information functions as an educational tool for their end-users, raising awareness towards various medical topics (1). With the expansion of the evidence-based medicine movement, patient inclusion, and open science initiatives, individuals now have better access to reliable health-related information, allowing them to take control of their health by making educated choices and implement particular preventative steps to reduce health risks (2,3). Access to reliable health-related information improves patient's health literacy, allowing them to fully comprehend health-related research findings and engage in meaningful patient-provider interactions (4). Furthermore, communicating health-related research findings with a larger audience is critical for addressing crucial public health concerns, increasing community well-being, and reacting to global health challenges (5).

1.2 Evidence-Based Medicine and Informed Decision-Making

The founder of Cochrane, Archibald Cochrane, pointed out once that throughout time, almost all patient-related medical decisions were made solely by a healthcare provider who based their decision either on the opinions of experienced senior colleagues, a random selection of information from variable literature or simply through trial and error (6). It was not until the 1990s that a new movement emerged with the aim of developing guidelines for clinical practice grounded on empirical data while placing a focus on the critical evaluation of the existing evidence (7). Evidence-based medicine (EBM) integrates three aspects to improve healthcare decisions. Currently best available evidence, along with the healthcare worker's clinical expertise and the values of the patient, are combined to make the best decision for the patient's health. Implementing EBM into practice entails five steps. One must first outline a clinically relevant question before searching for the best available evidence. Furthermore, the evidence must be critically appraised to determine if the evidence relates to the patient in question. After finding the best-known and relevant evidence, healthcare workers can consult with the patient and apply the evidence to treat them (8).

1.3 Levels of Evidence

In the context of evidence-based medicine, not all evidence is created equal. To be more precise, not all study designs provide an equal strength of evidence. The way a study is designed can inadvertently lead to the introduction of bias when interpreting the results, making them weaker and less reliable. Bias in studies can compromise validity, promote misinterpretation, and restrict the generalizability of the results (9). In the late 1970s, the Canadian Task Force on the Periodic Health Examination proposed a hierarchy of evidence presented as a "pyramid of evidence" (10). At the top of the pyramid were systematic reviews (SRs) with meta-analyses and randomized controlled trials (RCTs), which provide the highest level of evidence as they were designed to minimize the introduction of bias (by randomizing and blinding the participants, controlling conditions, and synthesizing evidence). In the middle of the pyramid are longitudinal and cross-sectional observational studies, while case studies and case series were placed at the bottom (**Figure 1**) (11).

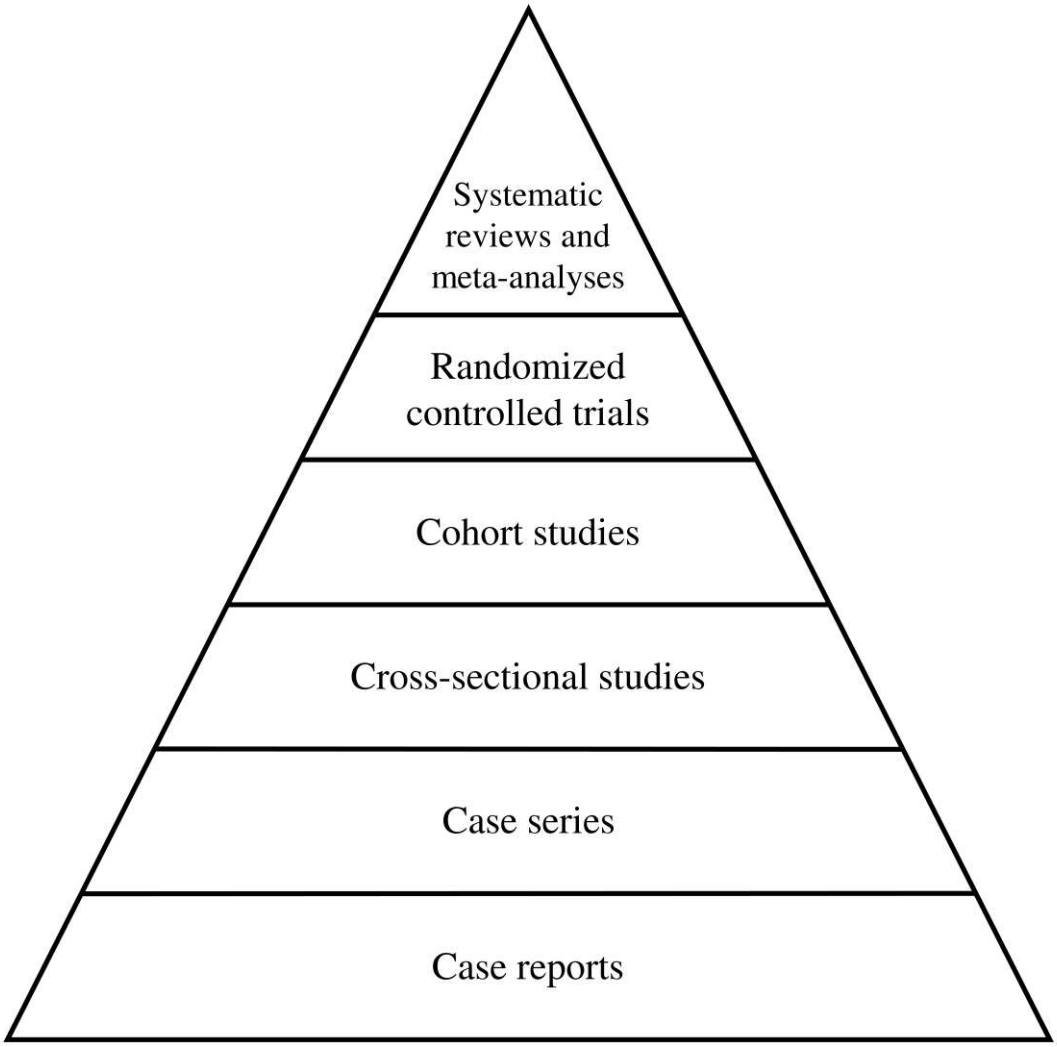

Figure 1. Pyramid of evidence.

1.4 Challenges in Interpreting Health-Related Research Findings

Interpreting medical research findings can be challenging for multiple reasons, which may affect their validity, reliability, and generalizability. Some of the main reasons are listed below.

1) **Low readability:** Health-related research findings are among the most complicated texts a person can come upon (12). Several studies assessing the readability of online health information aimed at patients have shown that more than 12 to 15 years of education was required for a layperson to be able to understand the text easily (13,14), and the readability of scientific texts has only decreased over time (15).

2) **Conflicting evidence:** Furthermore, with an exponential rise in the number of new studies published daily, it became common to have conflicting evidence regarding a particular topic. It can happen for different reasons, such as differences in study designs, sample sizes, inclusion criteria, measurement tools, and other factors. It could become problematic for any stakeholder in the healthcare system to draw definitive conclusions from such studies (16).

3) **Statistical vs. clinical significance:** Another challenge in interpreting health-related research findings is distinguishing between statistical and clinical significance (17). A great example would be a hypothetical study comparing a new hypertension medication to the gold standard. The data may show that the new drug lowers systolic blood pressure by 2 mmHg compared to the control, which might be statistically significant. However, a 2 mmHg drop in systolic blood pressure is unlikely to result in a meaningful improvement in the patient's symptoms, which makes it clinically insignificant.

4) **Generalizability:** Health-related research findings cannot always be generalized to all populations, settings, and contexts. The underlying principle behind any study conducted on people is to draw conclusions about the population based on data from a smaller sample of participants, as sampling the whole population is often impractical. However, in order to generalize the findings to the wider population, the study sample must be representative of the target population in every characteristic that might influence the outcome (18). For example, if only adult male participants were included in the study, its results could not be generalized to women or children, as

those groups differ in developmental, biological, cognitive, and behavioral characteristics.

5) **Publication bias:** Another important challenge to interpreting research findings and making informed decisions stems from publication bias. Publication bias occurs when it is easier for a researcher to publish studies showing positive and statistically significant results than those having negative or null findings. This bias can lead to exaggerating treatment effectiveness and effect sizes and misrepresenting the evidence base for a certain topic, leading to poor clinical decisions and compromised patient care (19).

6) **Complexity of health outcomes:** When assessing the effectiveness of a particular intervention, researchers must first choose an appropriate outcome measure. Choosing an outcome can sometimes be challenging as some interventions can be assessed using several different outcome measures, each with its strengths and limitations (20). Furthermore, different stakeholders in the healthcare system may prioritize different outcome measures based on their unique perspectives and needs (21). Finally, identifying patient-centered outcome measures relevant to the patient's choices and quality of life is critical (22).

1.5 Plain Language Communication

As the EBM movement gained popularity, it was accompanied by a rise in initiatives aiming to empower patients to make informed decisions about their health in partnership with the medical provider. Such initiatives included the development of training opportunities for medical workers (23), improving patient involvement in research and workshops (24), integrating patient values into clinical guidelines (25), and developing patient decision aids (26). However, to make informed decisions about their health, patients must first be able to critically evaluate the quality of the evidence, understand the benefits and risks associated with a specific treatment option, and set realistic expectations for the treatment's effectiveness (27). Because the readability of scientific papers was too high for laypeople to understand, it became necessary to convey the research findings to patients using plain language. Plain language is a style of communication specifically adapted for lay audiences. It is characterized by clarity, simplicity, and accessibility of information. The clarity and simplicity of plain language communication are reflected in presenting the information in a straightforward manner, as opposed to being ambiguous in the statements. Both can be

achieved by avoiding jargon and complex technical terms (28). To improve the accessibility and transparency of research findings and engage and empower patients to make informed decisions about their health, several evidence-producing organizations decided to publish plain language summaries (PLS) accompanying their scientific publications. Organizations such as Cochrane, Campbell Collaboration, and International Initiative for Impact Evaluation (3ie) either require or recommend their authors to provide PLSs with their conducted SRs (29–31).

1.6 Conclusiveness and Readability of Plain Language Summaries

Scientific papers can often be challenging to comprehend due to their complex and specialized terminology. PLSs play an important role in transferring knowledge gained through scientific methods to the lay audience by clarifying and simplifying the terms and making them more accessible (28). However, while those PLSs aimed to demystify science and research findings for the general population, a study has shown that the readability of those PLSs was not much better than the readability of their scientific abstracts and their readability was still above the recommended readability level (32).

1.7 Linguistic and Textual Characteristics of Scientific Texts

How a piece of information is presented within a text can have a significant impact on the retention of it. There are various characteristics of written text that play a role in communicating research findings to end users. Such characteristics include clarity and precision, use of technical terminology, conciseness, logical structure, or coherence (33). Other important characteristics of scientific texts are emotions and attitudes expressed by the authors, as well as sentiment and different elements of linguistic style within it, which have been shown to influence how individuals perceive the text. Studies have shown that the way a text is written can induce certain emotions within a reader, affecting whether they like it, believe in it, or view it as correct (34,35). There are several tools available for analyzing linguistic and psychological dimensions of the text, such as Linguistic Inquiry and Word Count (LIWC) software (36), Natural Language Processing (NLP) tools (37,38), sentiment analysis tools (39,40), different content analysis software (41), word embedding models (42), and use of psycholinguistic databases (43–45).

1) **Dictionary-based approach**

LIWC software is a textual analysis tool that calculates word count and the frequency of words that fit different dictionary categories in a text (36). The four dictionary categories are Analytic (46), Clout (47), Authentic (48), and Emotional Tone (49), which were created based on previous research conducted by the team. The Analytic category includes words that indicate logical, formal thinking, and cognitive processing. Examples of words that fit the Analytic category are '*evaluate*,' '*assess*,' '*strategy*,' and '*method*.' The Clout category reflects the language of leadership and status, with higher scores indicating assertiveness and dominance in communication and a lower score indicating a more hesitant text tone. Examples of words matching the Clout category are '*dominate*,' '*control*,' '*lead*,' and '*direct*.' Words fitting to the Authentic category reflect the honesty and sincerity of the expression within the text, and such words include '*genuine*,' '*real*,' '*transparent*,' and '*true*.' Finally, the Emotional tone category shows how well the text reflects positive or negative emotions such as joy, fear, and others. Words that fit the Emotional tone category are, for example, '*delighted*,' '*afraid*,' '*anxious*,' and '*hostile*.'

2) **Sentiment analysis**

Sentiment analysis, also referred to as opinion mining, is a computational approach that explores textual data to identify the emotional tone or sentiment contained within it. A commonly used approach for sentiment analysis involves using different sentiment lexicons or dictionaries in which words have already been pre-assigned to emotional categories (50). A researcher from the National Research Council in Canada developed one such lexicon called the NRC emotion lexicon (45). By conducting sentiment analysis using the NRC lexicon, we can calculate the number of words within a text that fit into different emotional categories and two valences (positive and negative). Emotional categories include Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise and Trust.

Identifying different aspects of language associated with clarity and understanding of health-related research findings can act as a basis for evidence-based clinical practice guidelines to enhance the ease of understanding and use of these texts, advancing the overall goal of promoting patient involvement in informed decision-making.

2 AIMS OF RESEARCH AND HYPOTHESES

2.1 The association between research design and the perceived treatment effectiveness: a cross-sectional study

There is limited knowledge regarding the degree to which different stakeholders in the healthcare system understand and utilize data from various scientific sources, as well as their understanding of the hierarchical structure of study designs. Therefore, the objective of this study was to evaluate how treatment effectiveness is perceived in relation to research design, as reported by researchers, healthcare workers, and consumers in Croatia.

The hypothesis of this study was:

1) The majority of healthcare workers and researchers will regard RCT and SR as credible forms of evidence, while consumers will view any form of study as reliable evidence; hence, we anticipate that consumers' choices will not vary considerably across different scenarios.

2.2 Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study

With this study, we aimed to assess the differences in readability, conclusiveness, and textual characteristics between PLSs of SRs published by medical and non-medical organizations, effectively addressing the gap in the existing body of research. The objective was to obtain a deeper understanding of the process by which scientific information is communicated to various individuals and how this influences their perception, involvement, and decision-making.

The hypothesis of this study was:

1) Compared to PLSs published by medical organizations, non-medical PLSs will provide less precise conclusions and have lower levels of readability.

3 METHODS

3.1 The association between research design and the perceived treatment effectiveness: a cross-sectional study

3.1.1 Study design and setting

A quantitative, cross-sectional, questionnaire-based study design was implemented to assess the association between the presented study design and the perceived effectiveness of the treatment. The data was collected online over a span of three months, from November 2021 to February 2022. Prior to the study's commencement, we preregistered it on the Open Science Framework (https://osf.io/t7xmf).

3.1.2 Study outcomes

The primary outcome was the perceived adequacy of the evidence presented in each scenario in order to determine the effectiveness of the treatment (evidence sufficiency rating). This was assessed using a scale ranging from 1 to 10 for each scenario, where a score of "1" indicated a complete absence of evidence supporting the treatment's effectiveness, and a score of "10" indicated sufficient evidence. The secondary outcome was a quantitative assessment of treatment effectiveness (treatment effectiveness assessment), ranging from 1 to 10, where a score of "1" indicated ineffectiveness and a score of "10" indicated total effectiveness. Additionally, we collected data about the participants' age, gender, level of evidence, and occupation. The participants' levels of education were categorized as follows: primary school, secondary school, college, undergraduate, graduate, postgraduate, and current university students. In the Croatian educational system, primary school lasts eight years, after which students proceed to secondary school for three to five years. Before the introduction of the Act on Academic and Professional Titles and Academic Degrees (NN 123/2023) in 2023, those students who finished up to three years of undergraduate school were considered to have finished college education. However, that degree has now been made equal to undergraduate school (bachelor's degree). After undergraduate school, a person can finish graduate school to get a master's degree, followed by a postgraduate school, acquiring a PhD.

3.1.3 Data collection

An online questionnaire was created on the Survey Monkey® platform (https://www.surveymonkey.co.uk/). The questionnaire comprised demographic questions and six scenarios illustrating six studies about a particular treatment for a certain condition, each with a different study design (case study, case series, cross-sectional study, cohort study, RCT, and SR). The participants were presented with these scenarios in a random order to avoid order bias. This was accomplished by utilizing the 'Page randomization' feature under the 'Design survey' section of the SurveyMonkey® platform for every participant. The study authors developed the questionnaire and the accompanying scenarios. The authors had expertise in medical and psychology education as well as research methodology. The scenarios were subsequently evaluated by two impartial researchers who evaluated their reliability, brevity, and applicability and provided recommendations for change. The collective expertise of the research group guaranteed that research designs were appropriately portrayed. The questionnaire was disseminated through openly accessible email addresses and Facebook posts. Reminders were sent within a time frame of two weeks to one month following the initial invitation. An English version of the questionnaire can be found in **Appendix 1**.

3.1.4 Participants

Sampling

The sampling strategy employed was a combination of convenience and snowballing sampling.

Inclusion criteria

Participants were researchers, healthcare workers, and consumers over 18 years of age. Researchers have been included due to the novel findings and advances they provide. Their contributions serve to broaden the foundation of scientific knowledge regarding medical procedures and therapies. Healthcare workers utilize this knowledge in order to diagnose, treat, and manage conditions in their patients. Consumers are the ultimate recipients of healthcare services and active contributors to shared decision-making.

A researcher who held a doctoral degree in the field of biomedicine and health or had a minimum of one publication within the previous year and held positions as scientific and teaching personnel of a faculty was eligible for inclusion.

Healthcare workers who held professional titles of medical doctor (MD) or doctor of medicine in dentistry (DMD), retired or still active, were included in the sample.

Patients not affiliated with the faculty, lacking a PhD in the field of biomedicine and health, and not practicing medicine or dentistry were considered consumers.

3.1.5 Sample size calculation

We conducted a sample size calculation to calculate the minimum number of participants in each group required to observe a hypothesized difference between groups for the primary outcome. For that, we utilized an online sample size calculator (https://select-statistics.co.uk/calculators/sample-size-calculator-two-means/). The parameters used for the calculation were:

- A hypothesized difference in the evidence sufficiency rating of 1 point on a scale of one to ten, and the anticipated standard deviation of 2,
- Alpha error rate of 5%,
- Power of 80%.

We calculated that, in order to observe such a difference, we needed a minimum of 63 participants per group.

3.1.6 Statistical analysis

We included only those responses that were completed in full. Frequencies and percentages were used to describe demographic data collected from the participants. The Shapiro-Wilk test was used to assess the data distribution of numerical variables. As the data was asymmetrical, we used medians and interquartile range (IQR) when presenting data for the entire sample. We used medians and 95% confidence intervals (95% CI) to present numerical data for each group individually. We employed the Kruskal-Wallis test to assess the differences between the groups' scores for primary and secondary outcomes. Furthermore, to

test the differences in scores each group had for every scenario, we employed the Friedman test. Conover post hoc test was used to conduct multiple pairwise comparisons. Software used to conduct all analyses were JASP, version 0.16.1 (JASP Team, 2022) and MedCalc, version 20.027 (MedCalc Software, Ostend, Belgium). The significance level was established at a p-value of less than 0.05.

3.2 Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study

3.2.1 Study design and setting

This quantitative, cross-sectional, methodological, research-on-research study was implemented to analyze different characteristics of PLSs published by medical and non-medical organizations.

3.2.2 Study outcomes

Conclusiveness

We categorized all PLSs into one of three conclusiveness categories: conclusive, inconclusive, and unclear (**Table 1**). For medical PLSs, this was done using a fine-tuned large language model grounded on SciBERT (51). SciBERT is a pre-trained natural language processing model specifically trained to understand and work with scientific text (52). Additionally, our model was trained on more than 4,400 pre-classified PLSs to categorize new PLSs into three distinct conclusiveness categories. By using a model, we were able to optimize operations and guarantee uniformity through the reduction of human error. We intended to use the model to classify non-medical PLSs like we did for medical ones. However, upon further examination of the model outputs related to non-medical PLSs, we determined that the classification of said PLSs was inaccurate. As a result, we opted to classify non-medical PLSs by hand. One author assessed the conclusiveness of the non-medical PLSs, whereas the other author verified this assessment.

Table 1. Description of the conclusiveness categories

Conclusiveness category	**Description**
Conclusive	
Positive	Moderate to high-quality evidence supports the safety or efficacy of the product.
Negative	Moderate to high-quality evidence suggests that the intervention may have adverse effects or is ineffective.
Equal	Effectiveness and safety were comparable across all evaluated interventions.
Inconclusive	
Positive	While the evidence does indicate potential safety or efficacy, it is deemed inconclusive and of poor quality, and the authors propose that further investigation is warranted.
Negative	The intervention was probably unsafe or inefficient, according to the available evidence. The authors advised against using the intervention or comparison. The available evidence is of low quality or inconclusive. The authors also emphasized the need for further research.
Equal	The interventions seem to have comparable effectiveness and safety. However, the evidence is of lower quality or inconclusive, prompting the authors to recommend further research.
Unclear	
No evidence	Empty reviews. Since the search did not yield any RCTs, the reviews contain no empirical data.
No opinion	No opinion or judgment was offered by the authors.
Unclear	The authors failed to provide a definitive conclusion.

Readability assessment

The assessment of readability was conducted by utilizing the Simple Measure of Gobbledygook (SMOG) index readability score (53). The SMOG index quantifies the years of education required for an average individual to understand a given text. It was used as it has been shown to be the most appropriate readability measure for healthcare contexts (54). To calculate the SMOG index for every PLS, we used the *quanteda* and *quanteda.textstats* (55) packages in R. It was recommended by the American Medical Association that written materials for patients containing health information adhere to an educational level of no more than the sixth grade (56).

Language characteristics

We analyzed the textual properties of PLSs using the Linguistic Inquiry and Word Count (LIWC) software (36). In addition to determining the total number of words in each PLS, LIWC also determines what proportion of those words fall into each of four distinct dictionary categories: Analytical tone, Clout, Authenticity, and Emotional tone. A higher score on the analytical tone variable suggests a more formal, logical, and hierarchical style of writing, which is indicative of an objective tone. Lower scores on the clout variable indicate a more cautious text tone; this variable is connected with confidence and assertiveness. Increased usage of honesty-related words, personal pronouns, and expressions related to sincerity in communication is indicative of a higher level of authenticity. Furthermore, the emotional tone score increases as the text conveys a greater range of positive emotions. LIWC software provides numerical values ranging from 0 to 100, representing the proportion of words in a particular PLS that fall into each LIWC category.

Furthermore, we also analyzed the PLSs for sentiment using R's *syuzhet* (40) package. The *syuzhet* package's *get_NRC_sentiment* function is a sentiment analysis tool that uses a dictionary to score the text on eight distinct emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) and determine its positive or negative valence. In the end, the function tracks the number of terms in each category that are present in the NRC lexicon.

3.2.3 Materials

All PLSs from the most recent English-language SRs published by the Cochrane and the Norwegian Institute of Public Health (NIPH), as well as non-medical PLSs from the Campbell Collaboration and the International Initiative for Impact Evaluation (3ie) from inception until November 10, 2022, were eligible to be included. Cochrane is internationally recognized for its comprehensive SRs and is considered a gold standard in evidence synthesis within the healthcare sector. NIPH is a nationally and internationally recognized public health research and evidence synthesis authority. Similarly, Campbell Collaboration and 3ie are respected organizations known for their contributions to evidence synthesis and impact evaluation in the field of social sciences. Campbell Collaboration focuses on social and educational interventions, whereas 3ie evaluates the efficacy of development programs and policies. We did not include SRs that stated they followed the Cochrane approach but were not published as Cochrane SRs in the Cochrane Library. We had intended to incorporate PLSs published by two additional organizations – the Evidence for Policy and Practice Information and Co-ordinating Centre and the Joanna Briggs Institute. However, we discovered that they did not adopt PLSs as a tool for communicating the results of their SRs to the laypeople.

3.2.4 Data collection

A request was made via email to the delegates of all six organizations, requesting access to their PLSs. Joanna Briggs Institute clarified that they do not release PLSs alongside their SRs. Additionally, in their reply, Cochrane spokesperson advised us to fill out the online data-request form that is offered on their website. The web-scraping technique was employed using the *rvest* (57) and *tidyverse* (58) packages in R software version 4.2.1 (R Core Team, 2020) to extract titles, links, and publication dates of SRs from the websites of the Campbell Collaboration (https://www.campbellcollaboration.org/), NIPH (https://www.fhi.no/), and 3ie (https://www.3ieimpact.org/). NIPH and 3ie PLSs had to be manually extracted, while the ones from the Campbell Collaboration could be obtained by web-scraping. Using the "*Export selected citation(s)*" option in the Cochrane Library, we retrieved all of the Cochrane PLSs. Then, we used the R package *stringr* (59) to extract the titles, links, publication dates, and PLSs from each citation. All data was entered into a Microsoft Excel spreadsheet.

3.2.5 Statistical analysis

We used percentages and frequencies to display the conclusiveness data. Shapiro-Wilk test was used to determine the distribution of the continuous variables. The medians with 95% confidence intervals were used to describe all numerical data. In order to compare the conclusiveness data both within and across groups, a chi-squared test was employed. Furthermore, to compare the groups' scores on the SMOG index, LIWC variables, and sentiment analysis variables, the Mann-Whitney U test was used. The discrepancies in scores across the three categories of conclusiveness were tested using the Kruskal-Walls test.

A stepwise logistic regression model was built, with the significant variables serving as criteria predictors (medical or non-medical). Findings are displayed using odds ratios (ORs), with 95% confidence intervals (CI), as well as McFadden R^2. All analyses were carried out with MedCalc software, version 20.027 (MedCalc Software, Ostend, Belgium).

4 RESULTS

4.1 The association between research design and the perceived treatment effectiveness: a cross-sectional study

Researchers and healthcare workers with 1,816 publicly accessible email addresses were invited to take part in the study. Automated confirmation of message delivery was received from 1,437 addresses, accounting for 79.13 % of the total. Two hundred fifty-four addresses (13.99%) did not receive the message, and for 125 (6.88%) addresses, we did not receive an automated response regarding the delivery status. In an attempt to reach the consumers, we sent out invitations to the email addresses of 171 Croatian citizens' associations, requesting that they share the invitation with their members. We received the automated response for all addresses, and 156 (91.23 %) addresses received the message. Furthermore, the link to the questionnaire was shared on personal profiles and groups on Facebook, Meta Platforms, Inc.

We excluded 783 respondents out of a total of 1,389 who entered the questionnaire link due to missing information. Twenty-two respondents were excluded because they did not fulfill the inclusion requirements. The responses from eleven researchers who were either scientific or teaching personnel were excluded from the analysis. Finally, 584 participants were included in the analysis. Ninety-seven of them were researchers, 201 were healthcare workers, and 286 were consumers (**Figure 2**).

Most respondents were women (74.3%), with a median age of 43.5 years (IQR 33-52). Researchers and healthcare workers mostly finished graduate or postgraduate educational programs, whereas consumers primarily finished secondary and graduate schools (**Table 2**).

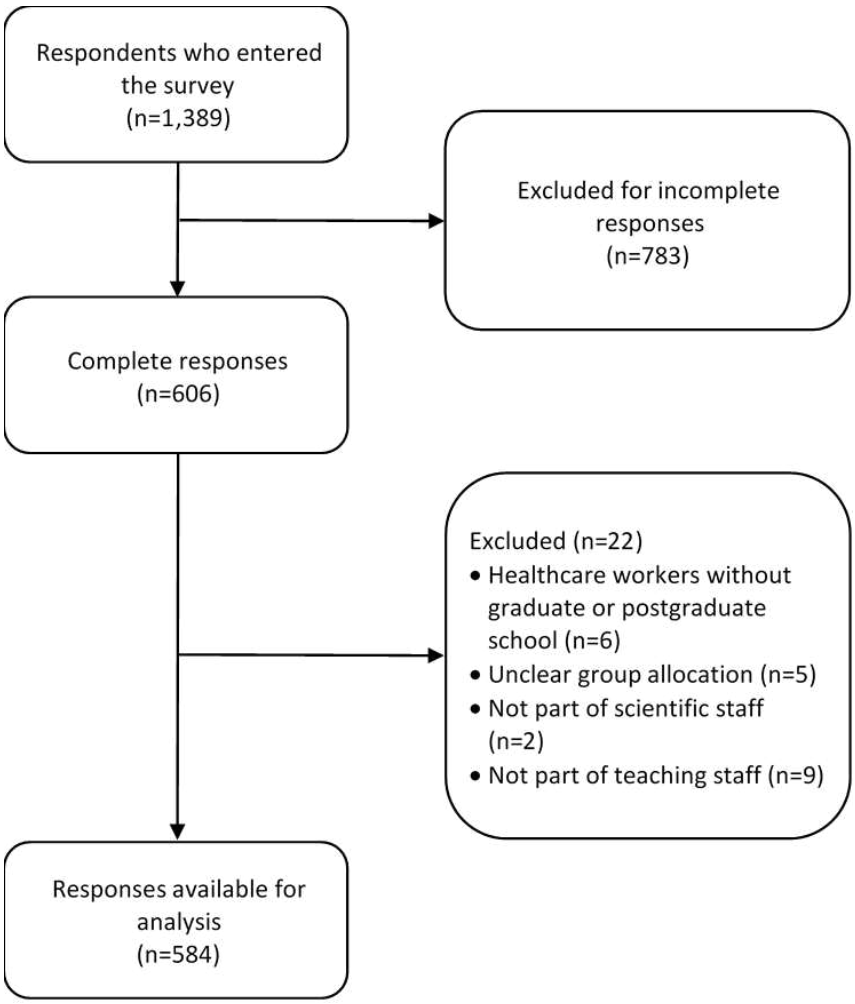

Figure 2. Participant flow (Reproduced under CC BY 4 license from (60)).

Table 2. Participant characteristics (Reproduced under CC BY 4 license from (60))

Variables	Total	Researchers	Healthcare workers	Consumers	P value
	n=584	n=97	n=201	n=286	
Women (n, %)	434 (74.3)	65 (67)	144 (71.6)	225 (78.7)	0.043*
Age (Md, IQR)	43.5 (33 to 52)	46 (38 to 53)	42 (31 to 51)	44 (34 to 52)	0.061†
Education (n, %)					
Primary school	3 (0.5)	0	0	3 (1)	<0.001*
Secondary school	113 (19.3)	0	0	113 (39.5)	
College	23 (3,9)	0	0	23 (8)	
Undergraduate school	22 (3.8)	0	0	22 (7.7)	
Graduate school	191 (32.7)	9 (9.3)	103 (51.2)	79 (27.6)	
Postgraduate school	208 (35.6)	88 (90.7)	98 (48.8)	22 (7.7)	
University student	24 (4.1)	0	0	24 (8.4)	

Md median, IQR interquartile range
*Chi-square test
†Kruskal-Wallis test

4.1.1 Evidence sufficiency rating

All participants' evidence-sufficiency scores improved as the evidence level increased, with statistically significant differences between every research design, with the exception of RCT and SR.

For researchers, there were no differences in ratings between case study and case series or between RCT and SR, respectively. Scores differed between researchers and consumers for every research design, with the exception of the RCT. Upon comparing the findings of researchers to healthcare workers, there were discrepancies in case series and cross-sectional study ratings.

Scores for healthcare workers did not differ significantly between case study and case series, cross-sectional study and cohort research, or RCT and SR, respectively. Scores for case study, case series, RCT, and SR differed between healthcare workers and consumers.

When compared to researchers and healthcare workers, consumers gave higher ratings for case study and case series, and lower for SR. There were no statistically significant differences in consumer ratings between the cross-sectional study and cohort study, the cohort study and cross-sectional study, or the RCT and SR, respectively. **Figure 3** highlights the scores for each group across all study designs. The score differences for each scenario across the groups and a comparison of designs for every group are provided in **Table 3**.

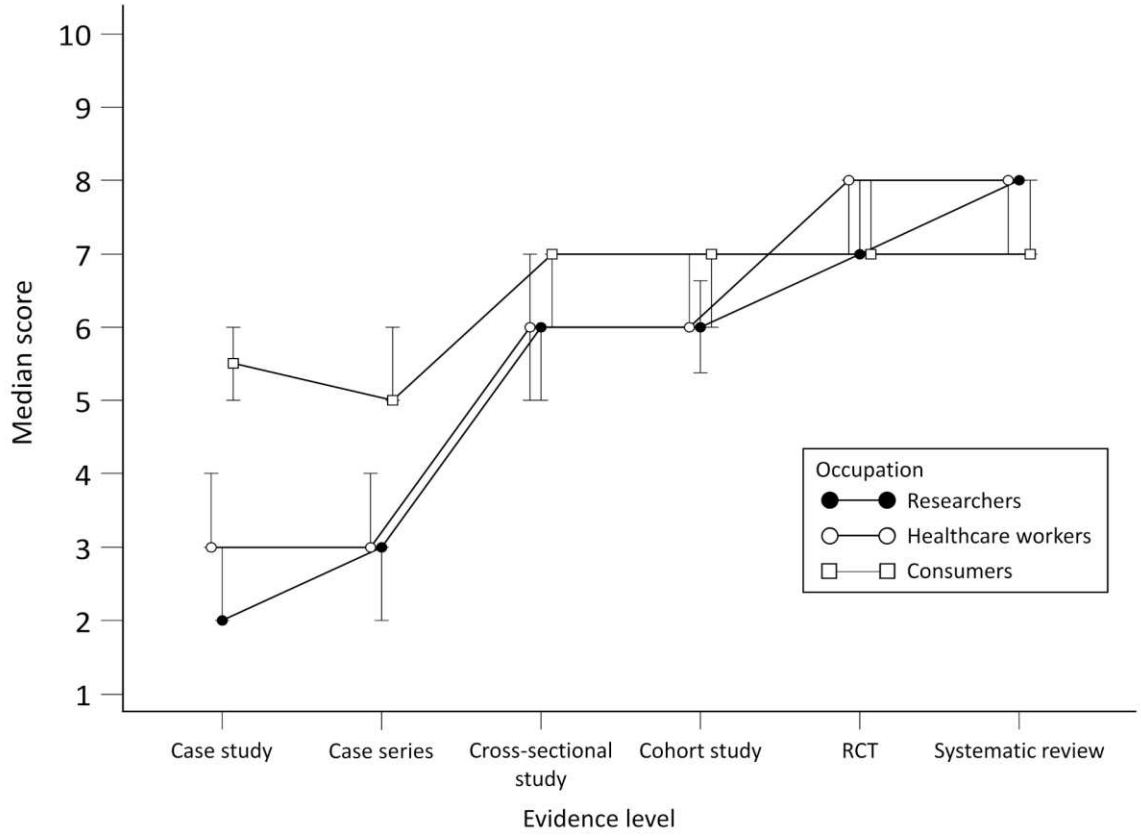

Figure 3. Median evidence sufficiency scores between the three groups. Reproduced under CC BY 4 license from (60).

Table 3. Differences in scores for evidence sufficiency for each scenario across the groups and a comparison of research designs for every group (Reproduced under CC BY 4 license from (60))

Variables (Md, 95% CI)	Researchers	Healthcare workers	Consumers	P*
	n=97	n=201	n=286	
Case study	2 (2 to 3)[#]	3 (3 to 4)[#]	5.5 (5 to 6)[‡]	<0.001
Case series	3 (2 to 3)	3 (3 to 4)	5 (5 to 6)	<0.001[§]
Cross-sectional study	6 (5 to 6)[‡]	6 (5 to 7)[b]	7 (6 to 7)[b]	0.003
Cohort study	6 (5,37 to 6,63)	6 (6 to 7)[l]	7 (6 to 7)[a]	0.151
RCT	7 (7 to 8)[l]	8 (7 to 8)	7 (7 to 8)	0.003
Systematic review	8 (8 to 8)[a]	8 (7 to 8)[a]	7 (7 to 8)[‡a]	0.004
P†	<0.001	<0.001	<0.001	

Md median, CI confidence interval

[*]Kruskal-Wallis test.

[†]Friedman test.

[‡]Different from the other two groups.

[§]All groups are different

[l]Not different from the other two groups

[#]Not different from the case series

[a]Not different from RCT

[b]Not different from the cohort study

4.1.2 Treatment effectiveness assessment

For all three groups, the ratings for the perceived treatment effectiveness in the scenarios increased as the evidence level increased. There were differences in scores across every study design, with the exception of RCT and SR.

There were no differences in scores for researchers between case study and cross-sectional study, cohort study and RCT, and RCT and SR, respectively. Researchers and consumers had differing scores for every research design, with the exception of cohort study and RCT. Compared to healthcare workers, researchers gave significantly lower scores for treatment effectiveness assessment for case series and RCT.

There were differences in scores for healthcare workers across every research design except RCT and SR. Their scores differed from those of consumers in every research design, with the exception of the cohort study.

Consumers scored higher for case study and case series but lower for SR than researchers and healthcare workers. There were no statistically significant differences in consumer scores between the case study and cross-sectional study, the cross-sectional study and cohort study, or the RCT and the SR, respectively.

Figure 4 shows the scores per group across all study designs. The differences in scores for each scenario across the groups, as well as a comparison of research designs for each group, are provided in **Table 4**.

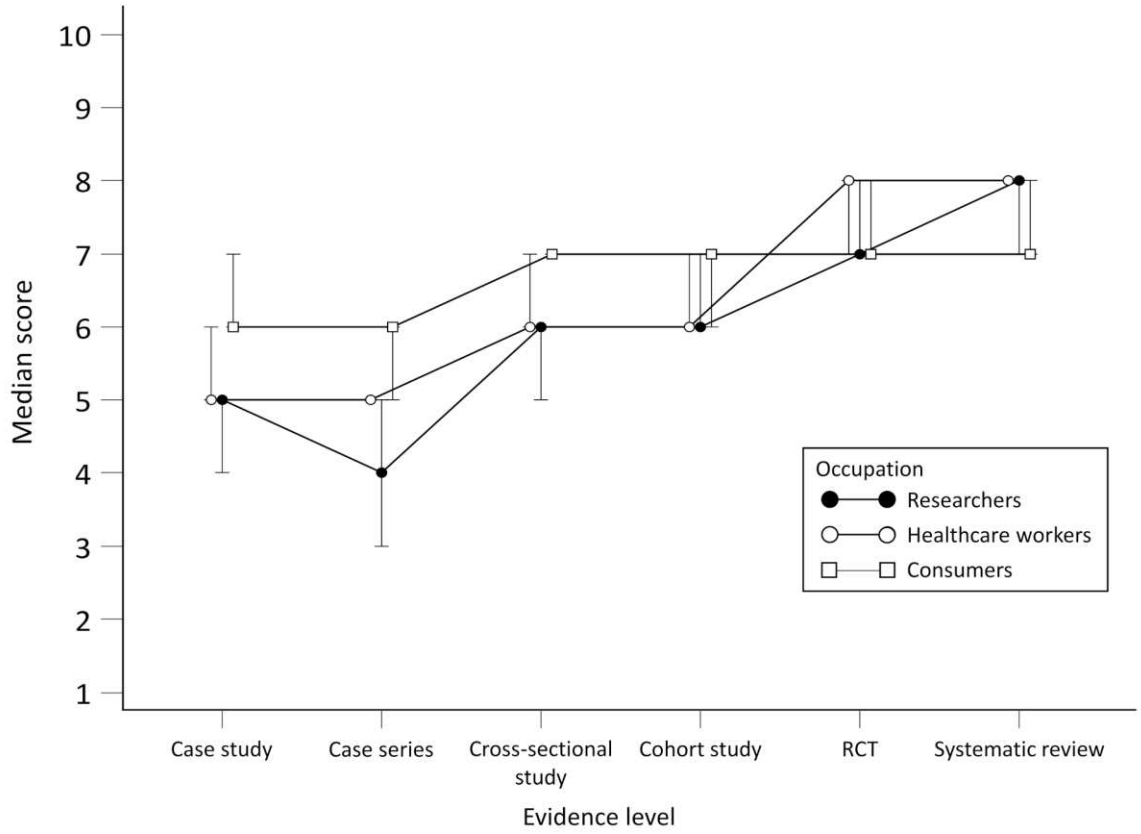

Figure 4. Treatment effectiveness rating between the three groups. Reproduced under CC BY 4 license from (60).

Table 4. Differences in scores for treatment effectiveness for every scenario across the groups and a comparison of research designs for each group (Reproduced under CC BY 4 license from (60))

Variables (Md, 95% CI)	Researchers	Healthcare workers	Consumers	P*
	n=97	n=201	n=274	
Case study	5 (4 to 5)[#]	5 (5 to 6)	6 (6 to 7)[‡#]	<0.001
Case series	4 (3 to 5)	5 (5 to 5)	5 (5 to 6)	<0.001[§]
Cross-sectional study	6 (5 to 6)	6 (6 to 7)	7 (7 to 7)[‡]	<0.001
Cohort study	6 (6 to 7)[a]	7 (6 to 7)	7 (6 to 7)[#]	0.296[I]
RCT	7 (7 to 8)	8 (7 to 8)[‡]	7 (7 to 8)	0.018
Systematic review	8 (7 to 8)[a]	8 (8 to 8)[a]	7 (7 to 8)[‡a]	0.002
P[†]	<0.001	<0.001	<0.001	

Md median, CI confidence interval

[*]Kruskal-Wallis test.

[†]Friedman test.

[‡]Different from the other two groups.

[§]All groups are different

[I]No differences between groups

[#]Not different from a cross-sectional study

[a]Not different from an RCT

4.2 Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study

Collectively, the examined organizations published 9,476 SRs. Among these, 9,209 were from medical organizations, with 8,928 from Cochrane and 281 from NIPH. The remaining 267 reviews were from non-medical organizations, with 220 from Campbell Collaboration and 47 from 3ie.

Of the 8,928 Cochrane SRs, 29 lacked PLSs, and 425 had been withdrawn. This resulted in 8,474 Cochrane PLSs that were analyzed in this study. NIPH released a total of 281 SRs. Out of these, only 32 were written in English. Among the English reviews, only one had a PLS. However, this PLS was not included in the analysis due to significant differences in sample sizes and the potential for sample size bias. The Campbell Collaboration has published 220 SRs, of which 68 did not have a PLS. Therefore, we included the remaining 152 PLSs. 3ie released a total of 47 SRs, of which 11 had PLSs considered for the analysis. Ultimately, our analysis included a total of 8,637 PLSs, with 8,474 originating from medical organizations and 163 from non-medical ones (**Figure 5**).

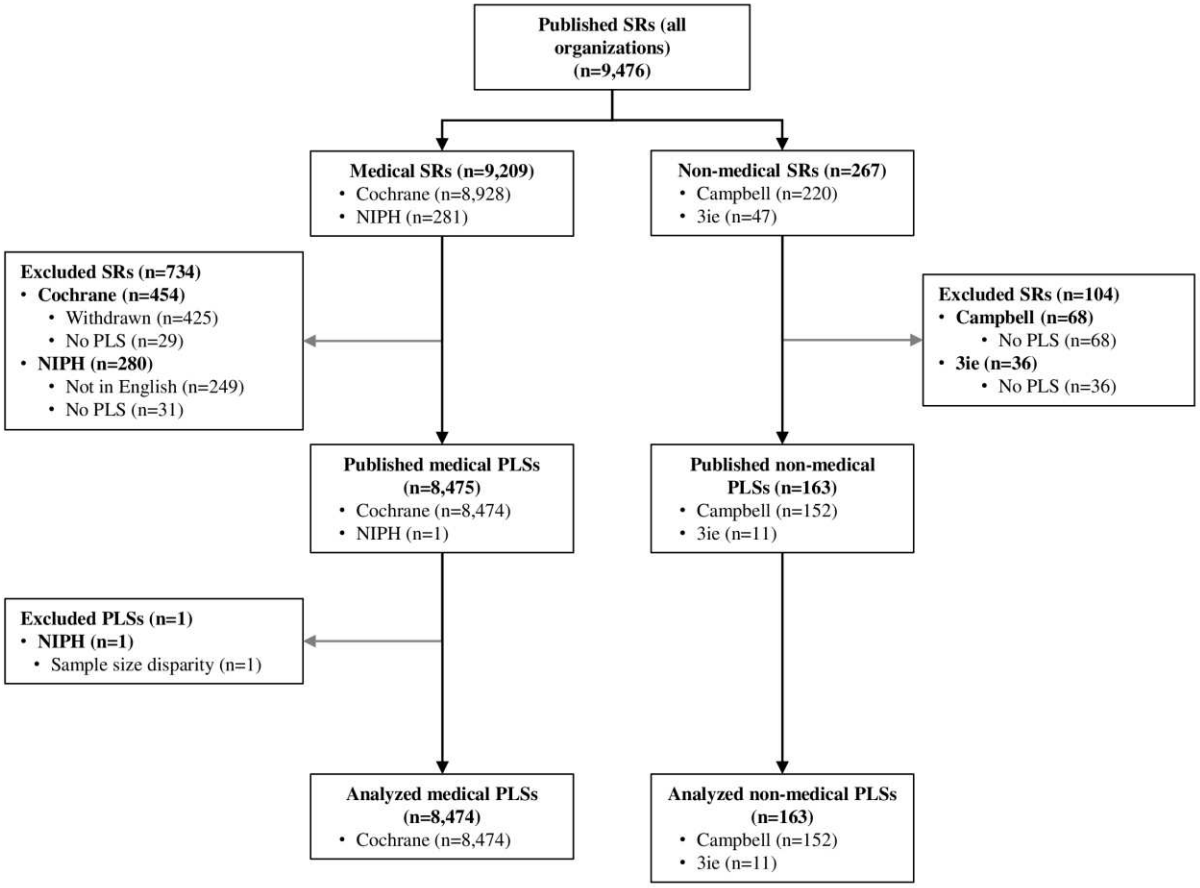

Figure 5. Flowchart of the selection of PLSs. Reproduced under CC BY 4 license from (61).

4.2.1 Conclusiveness

PLSs from medical organizations mostly had unclear conclusions (62.06 %), and only 8.4 % had conclusive findings. On the other hand, PLSs from non-medical organizations were mainly inconclusive (56.44 %). Furthermore, a more significant percentage of non-medical PLSs had conclusive findings compared to medical PLSs (**Table 5**).

Table 5. Comparison of distributions of PLSs among groups depending on the conclusiveness categories (Reproduced under CC BY 4 license from (61))

Variables N(%)	Overall sample	Medical	Non-medical	P*
	n=8,637	n=8,474	n=163	
Conclusive	739 (8.56)	710 (8.4)	29 (17.79)	<0.001
Inconclusive	2,597 (30.07)	2,505 (29.56)	92 (56.44)	<0.001
Unclear	5,301 (61.38)	5,259 (62.06)	42 (25.77)	<0.001
P*	<0.001	<0.001	<0.001	

*Chi-squared test

4.2.2 Readability

There was a statistically significant difference between medical and non-medical PLSs regarding readability, as measured by high SMOG ratings (P=0.010). Medical PLSs had a median SMOG index score of 15.51 (95% CI 15.47 to 15.58), while non-medical PLSs had a median score of 15.22 (95% CI 14.94 to 15.50).

4.2.3 Language characteristics

The length of non-medical PLSs was considerably greater than that of medical PLSs. In addition, non-medical PLSs had higher scores for clout (P=0.041) and emotional tone (P<0.001). No significant variations were observed between the groups in terms of the analytical tone. However, medical PLSs with unclear conclusiveness obtained lower scores compared to inconclusive and unclear medical PLSs. Non-medical PLSs that were conclusive scored higher for 'clout' (P=0.041) and 'authenticity' (P=0.010) compared to conclusive PLSs published by medical organizations. Furthermore, inconclusive PLSs published by medical organizations had a greater degree of clout, while conclusive medical PLSs displayed a higher level of authenticity than other medical PLSs (**Figure 6**).

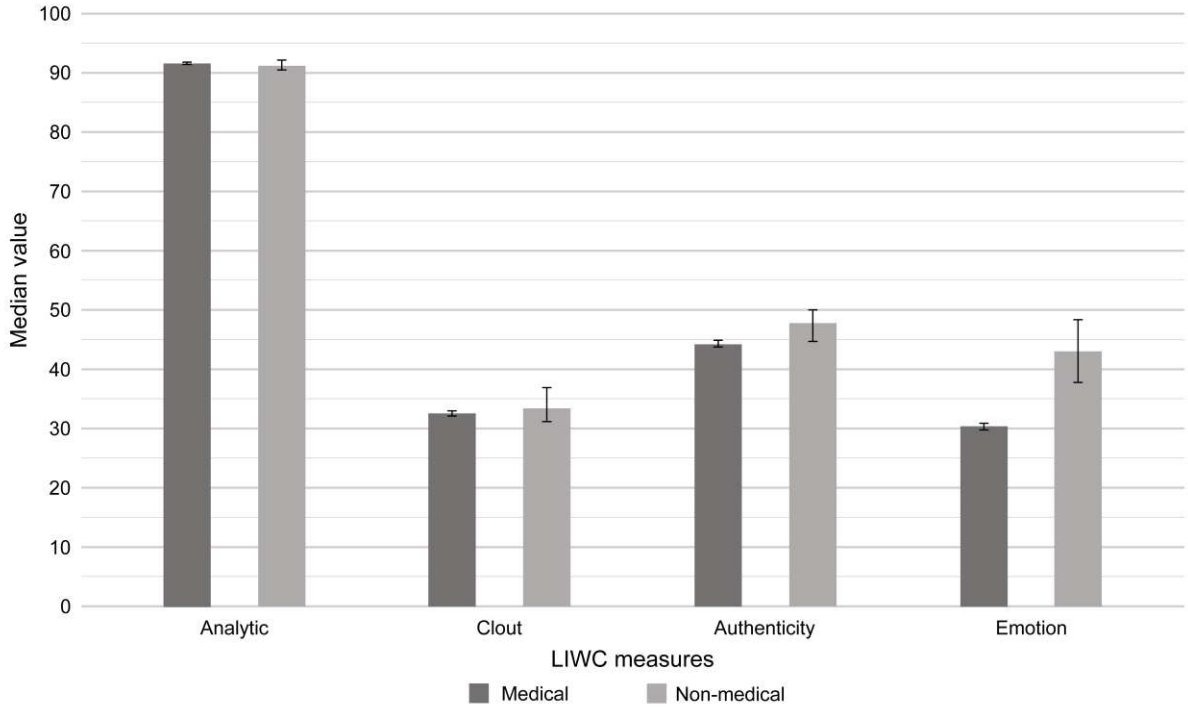

Figure 6. Percentage of linguistic traits observed in the PLSs for each group. Error bars represent the 95% confidence intervals. Reproduced under CC BY 4 license from (61).

4.2.4 Sentiment analysis

Non-medical PLSs exhibited greater scores for positive and negative valences than medical PLSs. Furthermore, PLSs published by non-medical organizations had a greater prevalence of words depicting emotions like joy, anger, trust, anticipation, and surprise, whereas PLSs published by medical organizations demonstrated higher levels of disgust and fear (**Figure 7**). Ten most frequently used words from each of the eight emotional categories and two valences for both groups are presented in **Appendix 2**.

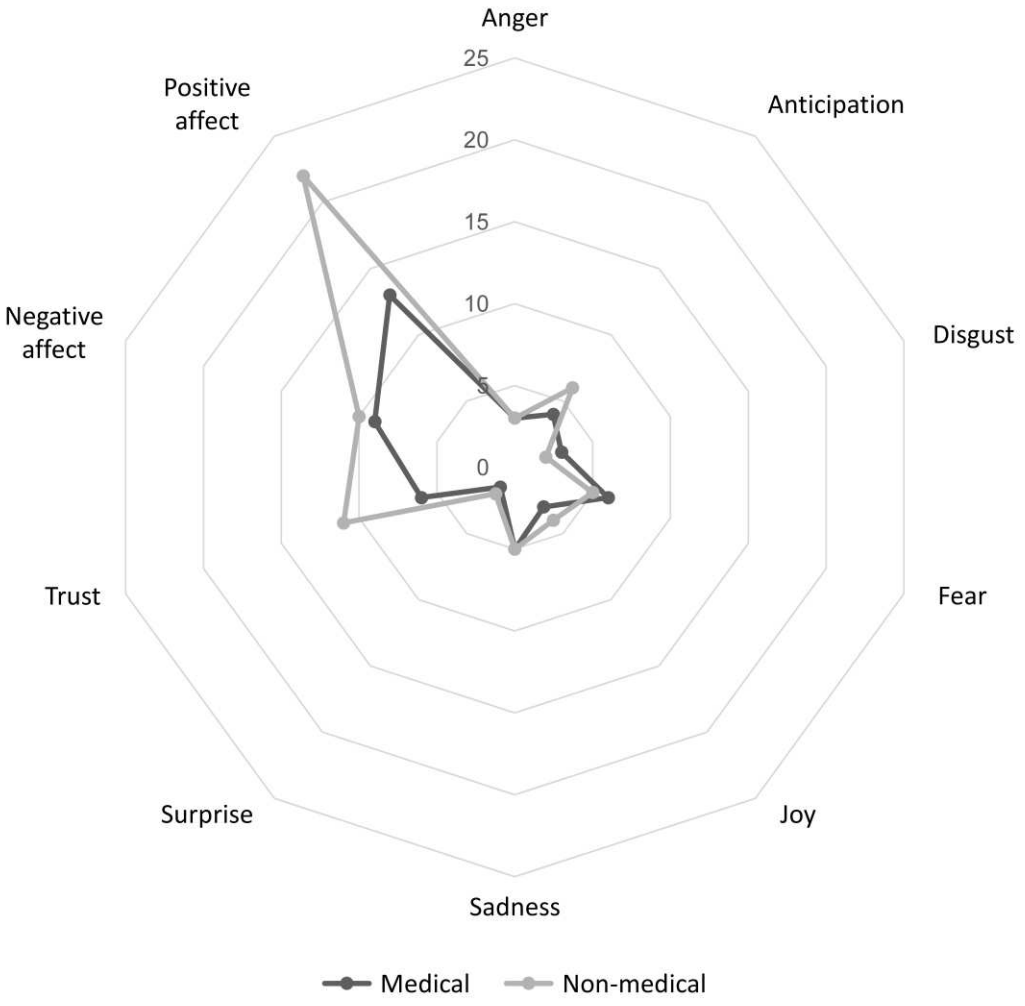

Figure 7. Evaluation of the PLSs' sentiments in the medical and non-medical domains. Reproduced under CC BY 4 license from (61).

4.2.5 Logistic regression analysis

Several variables were identified by the logistic model as potential indicators of medical or non-medical PLS status (**Table 6**). Both PLSs with a lower SMOG index (OR=0.808; 95% CI 0.721 to 0.905; P<0.001) and PLSs with a greater word count (OR=1.005; 95% CI 1.003 to 1.006; P<0.001) were more likely to be from a non-medical organization. However, the difference was practically irrelevant. Furthermore, medical PLSs were predicted by greater disgust, fear, and joy scores, though PLSs published by non-medical organizations were associated with greater ratings for analytical tone and emotions like anger, trust, and positive sentiment. The logistic regression included nine variables that explained 36.9% of the variance.

Table 6. Logistic regression model of variables predicting if the PLS was from a medical or non-medical organization

Variables	OR (95% CI)	P-value
SMOG index	0.808 (0.721 to 0.905)	<.001
Word Count	1.005 (1.003 to 1.006)	<.001
Analytic tone	1.074 (1.040 to 1.109)	<.001
Anger	1.673 (1.477 to 1.895)	<.001
Disgust	0.698 (0.609 to 0.800)	<.001
Fear	0.617 (0.562 to 0.677)	<.001
Joy	0.873 (0.789 to 0.966)	.009
Trust	1.204 (1.120 to 1.293)	<.001
Positive sentiment	1.102 (1.047 to 1.161)	<.001

Note. Type level 'Non-medical' coded as class 1, and 'Medical' as class 0.

OR odds ratio, CI confidence interval

McFadden R^2 - 0.369

5 DISCUSSION

5.1 The association between research design and the perceived treatment effectiveness: a cross-sectional study

5.1.1 Summary of the main findings

This cross-sectional, questionnaire-based study found notable variations in the perceptions and understanding of the hierarchy of evidence between different stakeholders in the Croatian healthcare system. Significant differences were observed in both outcomes between experts and consumers. Consumers had difficulty discerning across study designs as there were minimal variations in their perception of the sufficiency of evidence or the effectiveness of therapy across different scenarios. In comparison to other participant groups, consumers consistently assigned higher scores to scenarios that presented lower levels of evidence and lower scores to scenarios that depicted higher levels of evidence.

We aimed to evaluate the concept of scientific evidence, particularly in light of its increased emphasis in recent years and in the context of the coronavirus pandemic (62). The vast majority of media platforms had become overloaded with health-related information, containing accurate information and misinformation. It became crucial to discern the reliable information amongst it all (63–65). Identifying credible sources of health information, particularly regarding possible treatments, transmission, and effects, became significantly more challenging during this "infodemic," a term created by the Director General of the World Health Organization (WHO) (66). As a result of the infodemic, false information circulated rapidly through social media, and consumers had problems locating and comprehending resources regarding the disease's symptoms, treatment, and prevention (67,68). Researching the comprehension of the hierarchy of evidence as presented within different study designs may, therefore, be useful for knowledge translation to a broader audience. Our findings align with a previous study on healthcare workers in Canada that revealed a widespread inadequacy in critical appraisal skills among young medical doctors in that country (69). However, what sets our study apart is the incorporation of other population groups. As anticipated, researchers demonstrated the most comprehensive understanding of the topic, assigning higher ratings to RCT and SR and lower to the case study and case series than other stakeholders. On the other hand, consumers had the lowest ability to distinguish across study designs and assigned comparable ratings to all. We found that healthcare

workers in our study had difficulty distinguishing between observational study designs, experimental study designs, and SR. This may also have ramifications for how doctors comprehend the scientific evidence. Further research could determine whether or not the comprehension of the evidence hierarchy has any ramifications for the practice of medicine in clinical settings. Both healthcare workers and researchers undergo education in scientific methods as part of their studies and professional training. However, consumers lack formal training in subjects related to methods of conducting scientific research, which is not included in the curriculum for the earlier stages of education. Nevertheless, education on health and medical research has become more vital than it has ever been before as patient empowerment became a key component in the process of making healthcare decisions in the modern context. Moreover, the most worrisome aspect is to the healthcare workers' ratings regarding the sufficiency of evidence for making an informed decision and evaluating the effectiveness of treatments. In order to deliver optimal care to their patients, healthcare workers must have the ability to find and rigorously evaluate relevant evidence within their area of expertise. Due to their design, observational studies are unable to provide evidence on the effectiveness of treatment. Therefore, data obtained from a cross-sectional study should not be considered as equally sufficient to the evidence from an RCT or an SR. We identified the importance of introducing educational programs on the essential principles of conducting a research in primary or secondary schools. Mandatory courses on research methodology can also be implemented into the Continuing Medical Education (CME) programmes for healthcare workers.

5.1.2 Limitations of the study

Our study has several limitations which should be considered when interpreting the results. An important factor to consider is the fact that it was carried out on the Croatian population. There may be differences between the educational systems in Croatia and those in other countries when it comes to healthcare stakeholders' access to education and training in research methods, as well as education in general. Furthermore, researchers, healthcare workers, and consumers who participated in our study may have diverse experiences and backgrounds. This might lead to different perceptions of research designs, which should be taken into account when generalizing these results. In addition, because we employed a number of methods to disseminate the questionnaire, it was not possible to provide an accurate estimate of the overall response rate for this study. Only individuals who were

directly sent the invitation by electronic mail could be considered for the response rate calculation. It was not possible to accurately determine the number of individuals who came into contact with the posts on social media, citizen associations' members to whom the link to the questionnaire was sent, or how many individuals were included using the snowball sample method. Some respondents may have relied on variables unrelated to the scenarios themselves while answering the questions, despite the fact that we made every effort to structure the scenarios in the most effective manner possible to exclude the likelihood of response bias.

5.1.3 Strengths of the study

However, our study has several important strengths. For the sake of avoiding order bias, which may have been introduced by displaying the scenarios all at once or in incremental order, the questionnaire presented the scenarios to each participant in a random sequence. Also, participants were not provided with the specific details of the design used in each scenario in order to prevent the order of study designs from being determined using outside sources. Additionally, we designed the questionnaire to be completed in less than 10 minutes in order to guarantee the acquisition of high-quality responses from those taking part.

5.1.4 Suggestions for future research

A suggestion for further research is to carry out a more thorough validation of our questionnaire using the pre-existing methodological approaches (70). Additional multinational studies are required to validate these findings and offer a more profound understanding of the issue. Furthermore, future studies could examine the effectiveness of training programs in earlier stages of education on methods of conducting scientific research, as well as their inclusion in CME programs for various members of the healthcare system.

5.2 Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study

5.2.1 Summary of the main findings

This cross-sectional, research-on-research study revealed significant discrepancies between medical and non-medical PLSs in terms of their readability, conclusiveness, and textual characteristics. The readability of medical and non-medical PLSs was very low, as readers require more than 15 years of education to understand the texts, even though it was recommended that those texts be written so that individuals with as little as six years of education can easily comprehend them. Non-medical PLSs were found to be more conclusive, and they contained more words.

Given that the medical profession frequently handles choices that might mean the difference between life and death, and the biological variation in treatment response, illness progression, and genetics, the objective differences between the two research domains may be the source of the difference in conclusiveness. As a result, researchers in medical disciplines might exercise greater caution and reservation when drawing conclusions regarding a specific treatment option. There is also the possibility that a lower level of conclusiveness in medical PLSs is just a fair depiction of the current state of medical evidence, implying that it should not always be deemed a weakness. These findings are consistent with those of other studies that reported inconclusiveness of Cochrane SRs (71–74) and PLSs (32).

The difference in word count that was found between the two groups of PLSs may be an indication of distinct communication methods. Non-medical PLSs are more likely to prioritize exhaustive clarifications and use words to ensure they are clear and comprehensive. Medical PLSs, on the other hand, may place an emphasis on presenting concise conclusions instead of overwhelming the reader. Also, compared to medical PLSs, non-medical PLSs may have been more heterogeneous or measured a greater number of outcomes. In addition, there was a disparity between the guidelines proposed by Cochrane and Campbell concerning the word limits that were advised for PLSs (30,75). Cochrane recommends that PLSs be kept in the range of 400 to 850 words. Conversely, Campbell advises that the objective length should be between 600 and 750 words.

Our analysis of the textual characteristics identified that non-medical PLSs scored higher for the emotional categories of anger, emotional tone, clout, anticipation, surprise, and joy. In contrast, PLSs published by medical organizations scored higher on disgust and fear. The findings suggest that non-medical PLSs have a greater number of words and expressions that project self-assurance and social influence. On the other hand, there was a greater percentage of emotional and affective words and phrases in non-medical PLSs that indicated either positive or negative emotions. In addition, non-medical PLSs had a greater number of words and expressions that projected a hopeful perspective toward the future, and a sense of wonder and belief in particular individuals or entities. In contrast, medical PLSs displayed a greater number of terms that are related to feelings of worry and terror and expressions that provide a strong sense of dislike. That may be because medical PLSs have a greater number of words connected with the terms pain and disease, which may be a contributing factor to the varied levels of negative emotions among the two groups of PLSs. However, those terms do not account for the higher occurrence of positive words.

PLSs serve as a gateway to complex concepts for readers without the background knowledge to fully understand a research paper (76). It is crucial to keep the readability level at or below the recommended level in order to include as many readers as possible. The results of this study are consistent with the conclusions of Banić et al. (32) and Karačić et al. (77), who similarly concluded that PLSs published by Cochrane require over 15 years of education to understand with ease. The idea of having consumers involved in writing PLSs has gained attention in recent years to improve the readability of such texts. Many organizations and research groups have advocated for patient involvement throughout the research process, which includes knowledge dissemination and translation (78,79). Moreover, while Cochrane has recently revised its guidelines for writing PLSs to include a suggestion for seeking consumer feedback on potential improvements to the PLSs (75), such practice has still not been implemented fully.

5.2.2 Limitations of the study

It is important to consider several limitations when interpreting these results. The primary issue is the discrepancy in the number of PLSs published by medical and non-medical organizations. Cochrane is the leading organization that publishes health-related PLSs. To be included in the Cochrane Database of Systematic Reviews, a PLS is necessary for every SR

(80). This practice has not been systematically adopted by other scientific domains. Given that the medical field is currently the leading producer of SRs, this may explain why the total number of published PLSs is so drastically different between the organizations. An additional limitation is that we employed different methods for assessing conclusiveness in both groups. Even though we initially planned to use the machine learning tool for all PLSs, careful examination of the results revealed that the tool lacks accuracy when applied to non-medical PLSs. This may be due to the fact that the tool was trained specifically on medical literature, which could result in its lack of precision when classifying non-medical PLSs. We, therefore, manually categorized and validated non-medical PLSs. On top of that, we divided PLSs into two groups exclusively according to the publishing organization. It is important to note that this may not be completely accurate, as both the Campbell Collaboration and 3ie periodically provide SRs related to health topics. The majority of non-medical organizations' content consists of social and behavioral programs, societal concerns and regulations, occupational safety and productivity, training programs, well-being and welfare services, and legal systems and implementation of the law.

5.2.3 Strengths of the study

This study has several strengths which improve its reliability. While previous studies analyzing different characteristics of PLSs have only included those pertaining to a single medical specialty, we included all the latest versions of PLSs of SRs published by Cochrane, Campbell Collaboration, and 3ie, irrespective of the specialty or subcategories of topics within the databases, thus ensuring a robust selection process. Furthermore, using a machine-learning technique for medical PLSs' classification has various advantages over manual categorization. They include increased efficiency in handling large volumes of data quickly, guaranteeing consistency and precision in making predictions (thus reducing human error and bias), and the ability to adapt to specific research aims, thereby enhancing the overall quality of a study. Additionally, there is a lack of studies assessing these outcomes in PLSs published by non-medical organizations, which was addressed in our study.

5.2.4 Suggestions for future research

Qualitative studies should be conducted to identify the factors that led to the disparity in the number of words between the two groups of PLSs, as well as to make firmer conclusions on

the reasons behind the disparity in the emotional characteristics between the groups. Furthermore, more research is required to decide on the optimal tone and simplification of the PLSs to engage the readers. The possibility for advancement may lie in the utilization of language models like ChatGPT (https://chat.openai.com/) or BERT (https://doi.org/10.48550/arXiv.1810.04805), which are expected to become more integrated into daily life in the future. Additionally, writers need to exercise caution when simplifying language in PLSs to ensure that the findings of the PLSs are not altered. Therefore, further research is required to determine whether reducing the reading levels affects the message quality from a PLS. Moreover, in order to save time and resources while improving the readability of messages for the public, future research might investigate the application of large language models.

5.3 Integrating Findings, Implications, and Future Directions

Our results highlight the significance of evidence-based decision-making in clinical practice. The healthcare workers' and consumers' consideration of different study designs as equally adequate for evaluating treatment effectiveness suggests a possible misinterpretation or lack of knowledge regarding the evidence hierarchy. This underscores the critical need for targeted interventions to improve scientific literacy among healthcare workers, particularly those responsible for making decisions about treatment, as well as patients involved in the shared decision-making. Possible interventions may consist of training programs, workshops, or online resources to boost comprehension of research methods, evidence hierarchy, and critical appraisal skills. Future interventions should aim to increase healthcare workers' knowledge of evidence hierarchies and promote the use of evidence-based guidelines to improve patient care outcomes. The second study focuses on the quality of PLSs published by medical and non-medical organizations. While both types of PLSs exceeded the recommended readability level, PLSs published by non-medical organizations were found to be more conclusive and easier for readers to understand. This suggests that non-medical organizations could hold valuable insights into effectively communicating complex concepts to various target groups.

Furthermore, PLSs published by medical organizations contained more words that conveyed negative emotions, like disgust and fear, whereas PLSs published by non-medical organizations communicated more words associated with positive emotions. These findings highlight the importance of emotional tone in health communication and suggest that using

more positive language may improve the accessibility and reception of medical information. Considering the significance of PLSs in communicating research findings to various audiences, it is essential to establish a uniform format and content for such. Ensuring PLSs are clear and understandable across different fields and organizations may require developing standards or templates for this purpose. Organizations may improve health literacy among diverse populations by building on each other's strengths and improving the effectiveness of their communication efforts.

Finally, the results from these studies highlight the complexity of knowledge translation and communication, as well as the importance of raising awareness about scientific literacy, improving health communication, and encouraging healthcare workers to base their decisions on evidence.

6 CONCLUSION

When determining the perceived sufficiency of scientific evidence for the effectiveness of an intervention, researchers could differentiate among the types of research designs more effectively than healthcare workers or patients.

Healthcare workers regarded observational study designs as equivalent to experimental study designs and SRs of RCTs for deciding about treatment effectiveness.

Consumers had the lowest ability to differentiate between the six study designs. This could be explained by the lack of education on research methodology at lower levels of education in Croatia.

The medical and non-medical PLSs differ in terms of conclusiveness, readability, word count, and textual characteristics. However, the underlying explanation for those disparities remains unexplained. Nevertheless, these differences might be explained by variations in publication methods or differences in the relevant disciplines.

Both non-medical and medical PLSs remain to be written below the recommended level of readability, which can lead to misinterpretation of the study results, the spread of false information, and the exclusion of certain groups of readers. An important issue is whether the suggested degree of readability can be achieved without compromising the findings.

Overall, our study may help PLS readers better understand and engage with the scientific information, which, in turn, improves their decision-making abilities and may have significant implications for the field.

7 SAŽETAK

Ciljevi: Ovo istraživanje imalo je za cilj procijeniti i ispitati različite aspekte dokaza potrebnih za donošenje dobro informiranih zdravstvenih odluka, kao i kvalitete laičkih sažetaka (PLS) koje objavljuju medicinske i nemedicinske organizacije. Cilj je bio empirijskim istraživanjem pridonijeti razumijevanju ove teme i ponuditi vrijedne uvide u područja zdravstvene komunikacije i odlučivanja temeljenog na dokazima.

Metode: Proveli smo dva presječna istraživanja (anketno i metodološko). Uključili smo različite dionike zdravstvenog sustava (znanstvenike, liječnike i pacijente) i analizirali sve PLS-ove iz najnovijih verzija sustavnih pregleda koje su objavili Cochrane (medicinski), Campbell Collaboration i International Initiative for Impact Evaluation (nemedicinski). Procijenili smo razumijevanje dionika o hijerarhiji dokaza te zaključivast, čitljivosti i tekstualne karakteristike PLS-a.

Rezultati: Znanstvenici su bili najsposobniji za razlikovanje ustroja istraživanja, dok pacijenti i liječnici nisu razlikovali opservacijska istraživanja, eksperimentalna ispitivanja i sustavne preglede. Nadalje, medicinski PLS-ovi bili su manje zaključivi i pokazali su slabiju čitljivost od nemedicinskih PLS-ova. Uz to, medicinski PLS-ovi sadržavali su više riječi koje pokazuju emocije gađenja i straha, dok su nemedicinski PLS-ovi prikazivali više pozitivnih emocija.

Zaključci: Tečajevi o osnovama znanstvene metodologije trebali bi biti obvezni za niže razine obrazovana te uključeni u trajno medicinsko usavršavanje za sve dionike u zdravstvenom sektoru. Općenito, naši rezultati mogu pomoći čitateljima PLS-ova da bolje razumiju i uključe se u znanstvene informacije, što, zauzvrat, poboljšava njihove sposobnosti donošenja odluka i može imati značajne implikacije za ovo područje.

8 SUMMARY

Aims: This research aimed to evaluate and examine various aspects of the evidence necessary for making well-informed health decisions, as well as the qualities of plain language summaries (PLSs) published by medical and non-medical organizations. The goal was to contribute to the comprehension of this topic through empirical research and offer valuable insights into the fields of health communication and evidence-based decision-making.

Methods: We conducted two cross-sectional studies (questionnaire-based and methodological). We included different stakeholders in the healthcare system (researchers, healthcare workers, and consumers) and analyzed all PLSs from the latest versions of SRs published by Cochrane (medical), Campbell Collaboration, and the International Initiative for Impact Evaluation (non-medical). We assessed the stakeholders' comprehension of the hierarchy of evidence and conclusiveness, readability, and textual characteristics of the PLS.

Results: Researchers were most capable of differentiating between study designs, while consumers and healthcare workers did not differentiate between observational studies, experimental studies, and SRs. Furthermore, medical PLSs were less conclusive and displayed lower readability than non-medical PLSs. Additionally, medical PLSs contained more words displaying emotions of disgust and fear, while non-medical PLSs displayed more positive emotions.

Conclusions: Courses on basic research methodology should be required for lower-level learners and included in Continuing Medical Education for all stakeholders in the healthcare sector. Overall, our results may help PLS readers better understand and engage with the scientific information, which, in turn, improves their decision-making abilities and may have significant implications for the field.

9 REFERENCES

1. Nass SJ, Levit LA, Gostin LO. "The Value, Importance, and Oversight of Health Research.," In: Institute of Medicine (U.S.), editor. *Beyond the HIPAA privacy rule: enhancing privacy, improving health through research*. Washington, D.C: National Academies Press (2009). p. 111–52

2. Tennant JP, Waldner F, Jacques DC, Masuzzo P, Collister LB, Hartgerink ChrisHJ. The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Research* (2016) 5:632. doi: 10.12688/f1000research.8460.3

3. Bellomo R, Bagshaw SM. Evidence-based medicine: Classifying the evidence from clinical trials – the need to consider other dimensions. *Crit Care* (2006) 10:232. doi: 10.1186/cc5045

4. Benjamin RM. Improving Health by Improving Health Literacy. *Public Health Rep* (2010) 125:784–785. doi: 10.1177/003335491012500602

5. Orritt R, Powell P. Getting the word out: how to talk to the public about your research. *Breathe* (2020) 16:200008. doi: 10.1183/20734735.0008-2020

6. Cochrane AL. *Effectiveness and efficiency: random reflections on health services*. London: Nuffield Provincial Hospitals Trust. (1972). 92 p.

7. Guyatt G. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA* (1992) 268:2420. doi: 10.1001/jama.1992.03490170092032

8. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ* (1995) 310:1122–1126. doi: 10.1136/bmj.310.6987.1122

9. Norvell D. Study types and bias—Don't judge a study by the abstract's conclusion alone. *Evid-Based Spine-Care J* (2010) 1:7–10. doi: 10.1055/s-0028-1100908

10. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J* (1979) 121:1193–1254.

11. Wallace SS, Barak G, Truong G, Parker MW. Hierarchy of Evidence Within the Medical Literature. *Hosp Pediatr* (2022) 12:745–750. doi: 10.1542/hpeds.2022-006690

12. Hayes DP. The growing inaccessibility of science. *Nature* (1992) 356:739–740. doi: 10.1038/356739a0

13. Pearson K, Ngo S, Ekpo E, Sarraju A, Baird G, Knowles J, Rodriguez F. Online Patient Education Materials Related to Lipoprotein(a): Readability Assessment. *J Med Internet Res* (2022) 24:e31284. doi: 10.2196/31284

14. Shneyderman M, Snow GE, Davis R, Best S, Akst LM. Readability of Online Materials Related to Vocal Cord Leukoplakia. *OTO Open* (2021) 5:2473974X2110326. doi: 10.1177/2473974X211032644

15. Plavén-Sigray P, Matheson GJ, Schiffler BC, Thompson WH. The readability of scientific texts is decreasing over time. *eLife* (2017) 6:e27725. doi: 10.7554/eLife.27725

16. Carpenter DM, Geryk LL, Chen AT, Nagler RH, Dieckmann NF, Han PKJ. Conflicting health information: a critical research need. *Health Expect* (2016) 19:1173–1182. doi: 10.1111/hex.12438

17. Sharma H. Statistical significance or clinical significance? A researcher's dilemma for appropriate interpretation of research results. *Saudi J Anaesth* (2021) 15:431. doi: 10.4103/sja.sja_158_21

18. Kamper SJ. Generalizability: Linking Evidence to Practice. *J Orthop Sports Phys Ther* (2020) 50:45–46. doi: 10.2519/jospt.2020.0701

19. Thornton A. Publication bias in meta-analysis its causes and consequences. *J Clin Epidemiol* (2000) 53:207–216. doi: 10.1016/S0895-4356(99)00161-4

20. Coster WJ. Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies. *Am J Occup Ther* (2013) 67:162–170. doi: 10.5014/ajot.2013.006015

21. Abdalla R, Pavlova M, Groot W. Prioritizing Outcome Measures for Value-Based Healthcare: Physicians' Perspectives in Saudi Arabia. *Value Health Reg Issues* (2023) 37:62–70. doi: 10.1016/j.vhri.2023.05.002

22. Bajgain KT, Mendoza J, Naqvi F, Aghajafari F, Tang K, Zwicker J, Santana M-J. Prioritizing Patient Reported Outcome Measures (PROMs) to use in the clinical care of youth living with mental health concerns: a nominal group technique study. *J Patient-Rep Outcomes* (2024) 8:20. doi: 10.1186/s41687-024-00694-z

23. Nembaware V, African Genomic Medicine Training Initiative, Mulder N. The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. *Front Genet* (2019) 10:1209. doi: 10.3389/fgene.2019.01209

24. Young B, Bagley H. Including patients in core outcome set development: issues to consider based on three workshops with around 100 international delegates. *Res Involv Engagem* (2016) 2:25. doi: 10.1186/s40900-016-0039-6

25. Zhang Y, Coello PA, Brożek J, Wiercioch W, Etxeandia-Ikobaltzeta I, Akl EA, Meerpohl JJ, Alhazzani W, Carrasco-Labra A, Morgan RL, et al. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. *Health Qual Life Outcomes* (2017) 15:52. doi: 10.1186/s12955-017-0621-0

26. Stacey D, Volk RJ, for the IPDAS Evidence Update Leads (Hilary Bekker, Karina Dahl Steffensen, Tammy C. Hoffmann, Kirsten McCaffery, Rachel Thompson, Richard Thomson, Lyndal Trevena, Trudy van der Weijden, and Holly Witteman). The International Patient Decision Aid Standards (IPDAS) Collaboration: Evidence Update 2.0. *Med Decis Making* (2021) 41:729–733. doi: 10.1177/0272989X211035681

27. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient Involvement in Health Care Decision Making: A Review. *Iran Red Crescent Med J* (2014) 16: doi: 10.5812/ircmj.12454

28. Salita JT. Writing for lay audiences: A challenge for scientists. *Med Writ* (2015) 24:183–189. doi: 10.1179/2047480615Z.000000000320

29. The Cochrane Collaboration. Standards for the reporting of Plain language summaries in new Cochrane Intervention Reviews 2013. (2013) https://consumers.cochrane.org/sites/consumers.cochrane.org/files/uploads/pleacs_2019.pdf [Accessed January 4, 2022]

30. The Steering Group of the Campbell Collaboration. How to write a plain language summary for a Campbell systematic review. The Campbell Collaboration. (2016). https://core.ac.uk/download/pdf/52568807.pdf

31. Briefs. *Int Initiat Impact Eval* https://www.3ieimpact.org/evidence-hub/publications/briefs [Accessed January 10, 2022]

32. Banić A, Fidahić M, Šuto J, Roje R, Vuka I, Puljak L, Buljan I. Conclusiveness, linguistic characteristics and readability of Cochrane plain language summaries of intervention reviews: a cross-sectional study. *BMC Med Res Methodol* (2022) 22:240. doi: 10.1186/s12874-022-01721-7

33. Tausczik Y, Pennebaker J. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J Lang Soc Psychol* (2010) 29:24–54. doi: 10.1177/0261927X09351676

34. Alter AL, Oppenheimer DM. Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personal Soc Psychol Rev* (2009) 13:219–235. doi: 10.1177/1088868309341564

35. Denecke K, Deng Y. Sentiment analysis in medical settings: New opportunities and challenges. *Artif Intell Med* (2015) 64:17–27. doi: 10.1016/j.artmed.2015.03.006

36. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The Development and Psychometric Properties of LIWC-22. Austin, TX: University of Texas at Austin. (2022). https://www.liwc.app

37. Bird S, Klein E, Loper E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc. (2009). https://www.nltk.org/book/

38. Řehůřek R, Sojka P. Software Framework for Topic Modelling with Large Corpora. University of Malta, Valletta, Malta: University of Malta (2010). p. 45–50 http://is.muni.cz/publication/884893/en

39. Hutto CJ, Gilbert EE. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International Conference on Weblogs and Social Media (ICWSM-14).* Ann Arbor, MI (2014)

40. Jockers ML. Syuzhet: Extract Sentiment and Plot Arcs from Text. (2015) https://github.com/mjockers/syuzhet [Accessed January 13, 2022]

41. Jackson K. *Qualitative data analysis with NVivo*. 3rd edition. Thousand Oaks, CA: SAGE Publications. (2019).

42. Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics (2014). p. 1532–1543 doi: 10.3115/v1/D14-1162

43. Coltheart M. The MRC Psycholinguistic Database. *Q J Exp Psychol Sect A* (1981) 33:497–505. doi: 10.1080/14640748108400805

44. Peti-Stantić A, Anđel M, Gnjidić V, Keresteš G, Ljubešić N, Masnikosa I, Tonković M, Tušek J, Willer-Gold J, Stanojević M-M. The Croatian psycholinguistic database: Estimates for 6000 nouns, verbs, adjectives and adverbs. *Behav Res Methods* (2021) 53:1799–1816. doi: 10.3758/s13428-020-01533-x

45. Mohammad SM, Turney PD. Crowdsourcing a Word-Emotion Association Lexicon. (2013) doi: 10.48550/ARXIV.1308.6297

46. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI. When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLoS ONE* (2014) 9:e115844. doi: 10.1371/journal.pone.0115844

47. Kacewicz E, Pennebaker JW, Davis M, Jeon M, Graesser AC. Pronoun Use Reflects Standings in Social Hierarchies. *J Lang Soc Psychol* (2014) 33:125–143. doi: 10.1177/0261927X13502654

48. Newman ML, Pennebaker JW, Berry DS, Richards JM. Lying Words: Predicting Deception from Linguistic Styles. *Pers Soc Psychol Bull* (2003) 29:665–675. doi: 10.1177/0146167203029005010

49. Cohn MA, Mehl MR, Pennebaker JW. Linguistic Markers of Psychological Change Surrounding September 11, 2001. *Psychol Sci* (2004) 15:687–693. doi: 10.1111/j.0956-7976.2004.00741.x

50. Liu B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 1st ed. Cambridge University Press. (2015). doi: 10.1017/CBO9781139084789

51. Mijatović A, Ursić L, Buljan I, Marušić A. A Pretrained Language Model for Classification of Cochrane Plain Languages Summaries on Conclusiveness of Recommendations. *Abstracts accepted for the 27th Cochrane Colloquium, London, UK*. Cochrane Database of Systematic Reviews (2023). p. 36464 doi: 10.1002/14651858.CD202301

52. Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. In: Inui K, Jiang J, Ng V, Wan X, editors. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics (2019). p. 3615–3620 doi: 10.18653/v1/D19-1371

53. Mc Laughlin GH. SMOG Grading-a New Readability Formula. *J Read* (1969) 12:639–646.

54. Wang L-W, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Res Soc Adm Pharm* (2013) 9:503–516. doi: 10.1016/j.sapharm.2012.05.009

55. Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A. quanteda: An R package for the quantitative analysis of textual data. *J Open Source Softw* (2018) 3:774. doi: 10.21105/joss.00774

56. Weiss BD. *Health Literacy and Patient Safety: Help Patients Understand. Manual for Clinicians. (2nd ed.)*. American Medical Association Foundation. (2007). https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-clinicians-2nd-ed [Accessed January 13, 2022]

57. Wickham H. *rvest: Easily Harvest (Scrape) Web Pages*. (2022). https://rvest.tidyverse.org/, https://github.com/tidyverse/rvest

58. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. Welcome to the tidyverse. *J Open Source Softw* (2019) 4:1686. doi: 10.21105/joss.01686

59. Wickham H. *stringr: Simple, Consistent Wrappers for Common String Operations*. (2022). https://stringr.tidyverse.org, https://github.com/tidyverse/stringr

60. Bralić N, Buljan I. The association between research design and the perceived treatment effectiveness: a cross-sectional study. *Front Med* (2023) 10:1220999. doi: 10.3389/fmed.2023.1220999

61. Bralić N, Mijatović A, Marušić A, Buljan I. Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study. *Sci Rep* (2024) 14:6016. doi: 10.1038/s41598-024-56727-6

62. Djulbegovic B, Guyatt G. Evidence-based medicine in times of crisis. *J Clin Epidemiol* (2020) 126:164–166. doi: 10.1016/j.jclinepi.2020.07.002

63. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health Surveill* (2020) 6:e18444. doi: 10.2196/18444

64. Cuello-Garcia C, Pérez-Gaxiola G, van Amelsvoort L. Social media can have an impact on how we manage and investigate the COVID-19 pandemic. *J Clin Epidemiol* (2020) 127:198–201. doi: 10.1016/j.jclinepi.2020.06.028

65. Lwin MO, Lee SY, Panchapakesan C, Tandoc E. Mainstream News Media's Role in Public Health Communication During Crises: Assessment of Coverage and Correction of COVID-19 Misinformation. *Health Commun* (2023) 38:160–168. doi: 10.1080/10410236.2021.1937842

66. Simon F, Howard PN, Kleis Nielsen R. Types, sources, and claims of COVID-19 misinformation. *Reuters Inst Study Journal* (2020) https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation [Accessed October 26, 2022]

67. McCaffery K, Dodd R, Cvejic E, Ayre J, Isautier J, Copp T, Bonner C, Pickles K, Nickel B, Dakin T, et al. Health literacy and disparities in COVID-19–related knowledge, attitudes, beliefs and behaviours in Australia. *Public Health Res Pract* (2020) 30: doi: 10.17061/phrp30342012

68. Patil U, Kostareva U, Hadley M, Manganello JA, Okan O, Dadaczynski K, Massey PM, Agner J, Sentell T. Health Literacy, Digital Health Literacy, and COVID-19 Pandemic Attitudes and Behaviors in U.S. College Students: Implications for Interventions. *Int J Environ Res Public Health* (2021) 18:3301. doi: 10.3390/ijerph18063301

69. Godwin M, Seguin R. Critical appraisal skills of family physicians in Ontario, Canada. *BMC Med Educ* (2003) 3:10. doi: 10.1186/1472-6920-3-10

70. Dias AAL, Souza RS, Eduardo AHA, Felix AMDS, Figueiredo RMD. Validação de dois cenários de simulação clínica para ensino de prevenção e controle de infecções relacionadas à assistência à saúde. *Rev Eletrônica Enferm* (2022) 24:70072. doi: 10.5216/ree.v24.70072

71. Mandel D, Littner Y, Mimouni F, Lubetzky R. Conclusiveness of the Cochrane Neonatal Reviews: A systematic analysis. *Acta Paediatr* (2006) 95:1209–1212. doi: 10.1080/08035250600580537

72. Momosaki R, Tsuboi M, Yasufuku Y, Furudate K, Kamo T, Uda K, Tanaka Y, Abo M. Conclusiveness of Cochrane Reviews in physiotherapy: a systematic search and analytical review. *Int J Rehabil Res* (2019) 42:97–105. doi: 10.1097/MRR.0000000000000338

73. Cohen S, Mandel D, Mimouni FB, Marom R, Lubetzky R. Conclusiveness of the Cochrane Reviews in Nutrition: a systematic analysis. *Eur J Clin Nutr* (2014) 68:143–145. doi: 10.1038/ejcn.2013.252

74. Zhang X, Wu Z, Zhao H, Li C, Wu J, Dai G. Conclusiveness of the Cochrane Reviews in Palliative and Supportive Care for Cancer: A Systematic Analysis. *Am J Hosp Palliat Med* (2017) 34:53–56. doi: 10.1177/1049909115605058

75. Pitcher N, Mitchell D, Hughes C. Template and guidance for writing a Cochrane Plain language summary. (2022) https://training.cochrane.org/handbook/current/guidance-writing-cochrane-plain-language-summary.pdf

76. Santesso N, Glenton C, Lang B. Evidence That Patients can Understand and Use? *Z Für Evidenz Fortbild Qual Im Gesundheitswesen* (2008) 102:493–496. doi: 10.1016/j.zefq.2008.08.027

77. Karačić J, Dondio P, Buljan I, Hren D, Marušić A. Languages for different health information readers: multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats. *BMC Med Res Methodol* (2019) 19:75. doi: 10.1186/s12874-019-0716-x

78. About PCORI. *Patient-Centered Outcomes Res Inst* (2021) https://www.pcori.org/about/about-pcori [Accessed April 22, 2024]

79. Meissner P, Cottler LB, Eder M “Mickey”, Michener JL. Engagement science: The core of dissemination, implementation, and translational research science. *J Clin Transl Sci* (2020) 4:216–218. doi: 10.1017/cts.2020.8

80. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. *Cochrane handbook for systematic reviews of interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons. (2019).

10 APPENDICES

10.1 Appendix 1: English version of the questionnaire. (Reproduced under CC BY 4 license from (60))

Scenarios about scientific research

Dear Sir or Madam,

Before you is a survey used as a research instrument for the project " Professionalism in Health: Decision Making in Practice and Research – ProDeM" financed by the Croatian Science Foundation (IP-2019-04-4882), which is carried out at the Faculty of Medicine of the University in Split. In this study, we want to find out how you evaluate information about health and the ways in which you evaluate the effectiveness of health therapies.

Your answers are completely anonymous, they will be used only for research purposes and analyzed on a group level, and only researchers will have access to your data.

The survey is voluntary and intended for all adult citizens of the Republic of Croatia.

If you do not want to participate in the research, you are free to stop filling out the questionnaire at any time, and your data will not be taken into account during the analysis if you do not wish to do so. The study was approved by the Ethics Committee of the University of Split School of Medicine, Class 003-08/21-03/0003 No. 2181-198-03-04-21-0084.

For further information about the research, you are free to contact the research manager Nensi Ćaćić via email: nensi.cacic@mefst.hr.

Please read the offered scenarios and questions and answer them by choosing the offered answers in the provided place. We sincerely appreciate your feedback and thank you for your patience and affability.

1. I agree to participate in this research.
   - o Yes
   - o No

2. How old are you? *(please state only the number)*

_____

3. Gender:

- Male
- Female

4. Completed level of education:

- Primary school
- Secondary school
- College
- Undergraduate school
- Graduate school
- Postgraduate school
- Currently a university student

5. Occupation *(choose the one in which you spend **the most time**)*

- Healthcare worker (MD or DMD) – currently employed, unemployed or retired
- Researcher – had a PhD from the field of **Biomedicine and Health** in the last two years **OR** at least one scientific paper from the field of **Biomedicine and Health** published in the last year
- I do not have an occupation, I am currently **studying**
- Other (please state): _____

6. Position within the faculty: **only for researchers*

- Scientific and teaching staff
- Teaching staff
- Scientific staff

7. Do you have a PhD?

- Yes
- No

8. Have you published a scientific paper in the last year?

- Yes
- No

9. What year did you get your doctorate? *(please state only the number of the year)*

_____

On the following pages, you will be shown 6 scenarios of scientific research. Please read them carefully and answer the two questions below each scenario.

Scenario

This research aimed to evaluate the effectiveness of Drug X in relieving pain within a week after tooth extraction.

A participant is a 45-year-old male who came to the dental office complaining of pain in the lower jaw area on the right. In the medical history, he states that his lower first molar on that side has occasionally had spontaneous pain for the past month and that the day before he ate cherries and bit the pit with that tooth.

Examination of the mouth revealed a deep caries on the lower right first molar, and it was evident that the tooth had cracked from the crown to the root. An X-ray of the tooth showed inflammation under the root of that tooth and a tooth fracture that reached the bottom of the root. After that, the tooth was extracted, the alveolus was cleaned, and the patient was sent home.

As the occurrence of pain was expected for the next few days, the patient was recommended to drink painkiller Drug X twice a day (every 12 hours) for seven days and was asked to

record his pain level on a scale of 1 to 10 four times a day (1 hour before taking Drug X and 4 hours after).

After a week, the patient returned to the office for a check-up. The subject states that the pain significantly decreased within 4 hours after taking medicine and does not mention side effects related to taking the therapy.

10. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

11. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

Scenario

This research aimed to evaluate the effectiveness of Drug X in alleviating pain in the temporomandibular (chewing) joint area.

Four participants (two men and two women) participated in the research. All participants had pain symptoms in the temporomandibular (chewing) joint that occurs in the morning. The female participants were 43 and 29 years old, and the male participants were 64 and 32 years old. One female participant suffered from and was taking therapy for epilepsy, and the other had given birth 6 months prior and was currently breastfeeding. One male participant wore total dentures, and the other was infected with the hepatitis B virus. All the respondents were absent from work due to pain in the last month and felt pain a certain number of other days in the month. The participants were prescribed preventive medication of painkiller Drug X in

the evening before going to bed for a month and were instructed to monitor the number of days they felt joint pain when waking up and the number of days they were absent from work.

After a month, the participants returned for a check-up. The first participant stated that she missed significantly fewer days from work than the month before and felt somewhat less pain during the past month. The second participant stated that during the past month, he did not miss work due to pain and felt significantly less pain than in the previous month. The third participant stated that during the last month, he missed somewhat fewer days from work and felt somewhat less pain than in the previous month. The fourth participant stated that she missed significantly fewer days from work during the last month and felt somewhat less pain.

The participants did not mention side effects related to taking the therapy

12. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

13. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

Scenario

In this study, the authors wanted to evaluate the effectiveness of painkiller Drug X in adult patients suffering from periodontitis Stage 3, Grade B, who rated their pain as 4 or more on a

scale of 1 to 10, and who were taking painkiller Drug X as directed by periodontology specialists regularly in the last year.

Twenty-five periodontology specialists from the Republic of Croatia distributed the survey to their patients (who meet the criteria for inclusion in this study) when coming for a pre-arranged check-up. The survey consisted of questions related to health-related quality of life and pain assessment after using the assessed drug.

Finally, 267 participants completed the survey (125 men and 142 women).

The results show that 75% of people felt that Drug X helped.

14. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

15. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

Scenario

The purpose of this study was to assess the effectiveness of Drug X in relieving pain.

Participants in this study were students of the University of Split School of Medicine (320 respondents, 195 women and 125 men) who stated that they had a need to take painkillers at least once a month.

Students were divided into two groups (160 respondents per group). The participants themselves chose which drug and what quantity they wanted to take. When they felt the need to take a painkiller, participants in one group took Drug X (intervention group), and participants in the other group took the drug they usually take when they need a painkiller (control group).

The participants were followed for three years, and at the end, they filled out a survey in which they rated their overall satisfaction with the medicine they were taking, the effectiveness of the medicine, the duration of effect of the medicine, and listed all short-term and long-term side effects after taking medicine on a scale of 1 to 10.

Compared to the control group, participants in the intervention group were generally more satisfied with the treatment. The mean score on a scale of 1 to 10 for the effectiveness of the drug was higher in the intervention group than in the control group. Participants in the intervention group reported a more prolonged effect of the drug and, compared to participants in the control group, reported fewer side effects after treatment.

16. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

17. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

Scenario

This study aimed to determine how effectively Drug X alleviates the facial pain left behind after infection with the herpes zoster virus.

A total of 250 participants from the Republic of Croatia (130 women and 120 men of all age groups) participated in this study. All participants recovered from herpes zoster infection and had residual pain in the facial area innervated by the nerves affected by the virus.

The participants were randomly divided into two groups (125 respondents each), and neither the participants nor the researchers knew which participant belonged to which group. One group of participants (intervention group) took the painkiller Drug X twice a day after meals for one month, while the other group of subjects (control group) took ibuprofen at a dose of 600 mg twice a day for one month. Daily, the respondents marked the pain level immediately before the treatment and two hours after on a scale from 0 to 10. Also, possible side effects of the treatment were recorded.

The results showed that participants in the intervention group had a lower pain level two hours after treatment than subjects in the control group and had fewer side effects.

18. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

19. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

Scenario

In this study, the authors searched the medical research databases in search of all studies that compared the effect of Painkiller X with the effect of a placebo or another painkiller (ibuprofen, paracetamol, tramadol, etc.) in people who had their wisdom tooth surgically removed in the last 7 days.

A literature search resulted in 17 studies that included a total of 560 participants and that compared Drug X with a placebo or another pain reliever (ibuprofen 400 mg or paracetamol 500 mg).

Nine studies compared Drug X with ibuprofen 400 mg (275 participants), five studies compared it with paracetamol 500 mg (181 participants), and three compared it with placebo (104 participants).

An analysis of the results of all studies was done, and the results of nine studies comparing Drug X and ibuprofen 400 mg showed that there is moderate to high-quality evidence that Drug X is more effective than ibuprofen 400 mg in reducing pain after surgical extraction of wisdom teeth.

The results of studies comparing Drug X with paracetamol 500 mg showed that there is high-quality evidence that Drug X is more effective than paracetamol 500 mg in relieving pain after wisdom tooth extraction surgery.

Further, studies comparing Drug X with a placebo showed moderate to high-quality evidence that Drug X is more effective than placebo in relieving pain after wisdom tooth extraction surgery.

20. In your opinion, what level of evidence does this scenario provide for the effectiveness of Drug X?

(it provides no evidence)									(it provides all the evidence)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

21. How effective is Drug X?

(not at all effective)									(it is completely effective)
1	2	3	4	5	6	7	8	9	10
o	o	o	o	o	o	o	o	o	o

10.2 Appendix 2: Ten most frequently used words in the included PLSs, presented for each emotional category and valence

	Nr.	ANGER	ANTICIPATION	DISGUST	FEAR	JOY
NON-MEDICAL	1	limited	improve	larger	risk	improve
	2	bias	risk	adverse	adverse	found
	3	adverse	time	criminal	caution	majority
	4	caution	income	abuse	change	income
	5	violence	caution	finally	violence	child
	6	youth	public	poverty	youth	providing
	7	crime	child	excluded	problem	youth
	8	involvement	providing	violent	difficult	resources
	9	criminal	youth	weight	criminal	improvement
	10	abuse	prevention	morbidity	abuse	finally
MEDICAL	1	disease	risk	disease	risk	found
	2	adverse	improve	adverse	disease	improve
	3	versus	time	death	adverse	treat
	4	death	death	treat	pain	receiving
	5	limited	result	uncertain	death	birth
	6	treat	medical	cancer	prevent	good
	7	uncertain	treat	damage	surgery	improvement
	8	bias	patient	pregnancy	infection	safe
	9	cancer	receiving	bleeding	medical	confidence
	10	loss	birth	failure	treat	present

	Nr.	SADNESS	SURPRISE	TRUST	NEGATIVE	POSITIVE
NON-MEDICAL	**1**	intervention	larger	effective	intervention	included
	2	limited	youth	improve	small	main
	3	negative	assessment	found	risk	study
	4	adverse	finally	important	limited	effective
	5	income	present	provide	bias	intervention
	6	violence	good	policy	lack	improve
	7	problem	receiving	school	rigorous	increase
	8	lower	violent	show	negative	including
	9	abuse	money	larger	adverse	found
	10	case	weight	suggest	income	important
MEDICAL	**1**	disease	death	found	risk	found
	2	adverse	treat	effective	small	study
	3	pain	uncertain	important	disease	included
	4	death	receiving	improve	adverse	question
	5	surgery	good	medical	pain	effective
	6	limited	chance	treat	versus	important
	7	treat	randomly	provide	death	imporve
	8	intervention	present	hospital	limited	including
	9	cancer	receiving	bleeding	medical	confidence
	10	loss	birth	failure	treat	present

11 RESUME

Nensi Bralić, DMD

Department of Research in Biomedicine and Health

University of Split School of Medicine

Education:

Doctor of Dental Medicine, University of Split School of Medicine, 2020

Published scientific papers:

Ursić L, **Bralić N**, Žuljević MF, Puljak L, Buljan I. Exploring the understanding of reproducibility among stakeholders within academia and their expectations for a web-based education tool: A qualitative study. Account Res. 2024 May 5:1-30. doi: 10.1080/08989621.2024.2345723.

Bralić, N., Mijatović, A., Marušić, A., & Buljan, I. (2024). Conclusiveness, readability and textual characteristics of plain language summaries from medical and non-medical organizations: a cross-sectional study. Scientific reports, 14(1), 6016. https://doi.org/10.1038/s41598-024-56727-6

Bralić, N., & Buljan, I. (2023). The association between research design and the perceived treatment effectiveness: a cross-sectional study. Frontiers in medicine, 10, 1220999. https://doi.org/10.3389/fmed.2023.1220999

Ursić, L., Žuljević, M. F., Vuković, M., **Bralić, N.**, Roje, R., Matas, J., Mijatović, A., Sapunar, D., & Marušić, A. (2023). Assessing the quality and completeness of reporting in health systems guidance for pandemics using the AGREE-HS tool. Journal of global health, 13, 06050. https://doi.org/10.7189/jogh.13.06050

Kantorová, L., Friessová, T., Slezáková, S., Langaufová, A., Kantor, J., Munn, Z., Barker, T. H., Katikireddi, S. V., Mustafa, R. A., Žuljević, M. F., Lukežić, M., Klugarová, J., Riad, A., Vrbová, T., Pokorná, A., Búřilová, P., Búřil, J., Kirkovski, A., **Ćaćić, N.**, Delač, L., … Klugar, M. (2022). Addressing Challenges When Applying GRADE to Public Health Guidelines: A Scoping Review Protocol and Pilot Analysis. International journal of environmental research and public health, 19(2), 992. https://doi.org/10.3390/ijerph19020992

Šimundić Munitić, M., Šutej, I., **Ćaćić, N.**, Tadin, A., Balić, M., Bago, I., & Poklepović Peričić, T. (2021). Knowledge and attitudes of Croatian Dentists Regarding Antibiotic Prescription in Endodontics: A Cross-sectional Questionnaire-based Study. Acta stomatologica Croatica, 55(4), 346–358. https://doi.org/10.15644/asc55/4/2

Borges do Nascimento, I. J., von Groote, T. C., O'Mathúna, D. P., Abdulazeem, H. M., Henderson, C., Jayarajah, U., Weerasekara, I., Poklepovic Pericic, T., Klapproth, H. E. G., Puljak, L., **Cacic, N.**, Zakarija-Grkovic, I., Guimarães, S. M. M., Atallah, A. N., Bragazzi, N. L., Marcolino, M. S., Marusic, A., Jeroncic, A., & International Task Force Network of Coronavirus Disease 2019 (InterNetCOVID-19) (2020). Clinical, laboratory and radiological characteristics and outcomes of novel coronavirus (SARS-CoV-2) infection in humans: A systematic review and series of meta-analyses. PloS one, 15(9), e0239235. https://doi.org/10.1371/journal.pone.0239235

Borges do Nascimento, I. J., **Cacic, N.**, Abdulazeem, H. M., von Groote, T. C., Jayarajah, U., Weerasekara, I., Esfahani, M. A., Civile, V. T., Marusic, A., Jeroncic, A., Carvas Junior, N., Pericic, T. P., Zakarija-Grkovic, I., Meirelles Guimarães, S. M., Luigi Bragazzi, N., Bjorklund, M., Sofi-Mahmudi, A., Altujjar, M., Tian, M., Arcani, D. M. C., … Marcolino, M. S. (2020). Novel Coronavirus Infection (COVID-19) in Humans: A Scoping Review and Meta-Analysis. Journal of clinical medicine, 9(4), 941. https://doi.org/10.3390/jcm9040941

Conference Presentations:

2023. Cochrane Colloquium, London: “Level of scientific evidence needed to make an informed health decision among researchers, healthcare workers and consumers” – oral presentation