# Research, Statistical Methods and Scientific Writing in Sleep Medicine

An integrated approach to research and scientific writing, supported by statistical methodology, with a focus on sleep medicine and related fields.

2nd Edition of the Postgraduate Course in Sleep Medicine
Faculty of Medicine – UCP

30 May 2025

**Ricardo Miguel Vieira de São João**
**ORCID:** https://orcid.org/0000-0003-3137-0891
ricardo.sjoao@esg.ipsantarem.pt/rsj@net.sapo.pt

GADES DATA ANALYSIS SOLUTIONS

CEAUL Centro de Estatística e Aplicações Universidade de Lisboa

IPS POLITÉCNICO DE SANTARÉM

## Topics Covered

**1.Epidemiology**: etymology, definition, historical milestones, and objectives

**2.Types of Epidemiological Studies** and **Systematic Reviews with Meta-Analysis** (PICO framework)

**3.Scientific Writing in Health**: focus on Sleep Medicine and related fields

**4.Key Statistical Concepts**: what to consider when critically reading scientific articles


**Ricardo São João**

# Epidemiology

## Etymology

The word 'epidemiology' is derived from the Greek words: epi meaning 'upon', demos meaning 'people', and logos meaning 'study'.

## Definition

"The study of the distribution and determinants of health-related states or events in specific populations, and their application in the prevention and control of health problems."

Last JM (2001). A dictionary of epidemiology, 4th ed. Oxford, Oxford University Press.



**Ricardo São João**

# Objectives of Epidemiology

### Determine the Frequency

Statistical analysis of the distribution of disease(s) in population(s).

### Identify Causes

Identification of risk factors that lead to health problems.

### Formulate Strategies

Development of effective methods for prevention and control.

### Evaluate Interventions

Monitoring the effectiveness of implemented public health measures.



**Ricardo São João**

# Historical Milestones
# Scurvy among English Sailors

1

## Problem

18th century: scurvy devastated crews of the British Royal Navy. Approximately 1,400 of the 1,900 crew members (≈74%) died during Admiral George Anson's circumnavigation voyage

2

## Investigation

James Lind (1716–1794), the "Hippocrates of naval medicine"conducted the first controlled clinical trial in 1747.

3

## Solution

Lind demonstrated that citrus fruits effectively prevented and cured the disease.

4

## Impact

The simple dietary change saved thousands of lives.

Image taken from https://www.medicina.ulisboa.pt



**Ricardo São João**

# Cholera Epidemic in London

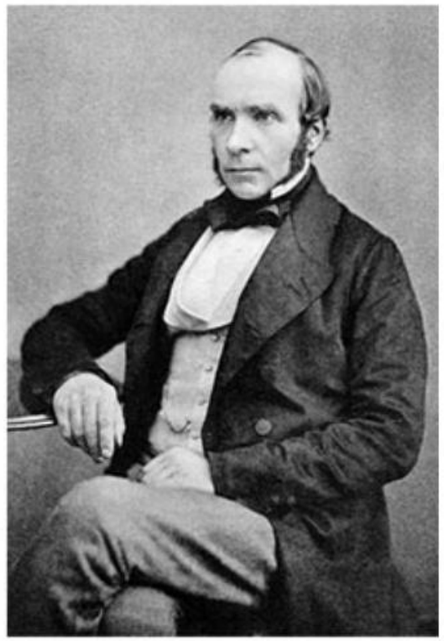



Image taken from https://eldiariodesalud.com

**outbreak**

In 1854, London was devastated by a cholera epidemic.

**Investigation**

John Snow (1813–1858) mapped cases around the Broad Street water pump.

**Discovery**

He identified water contamination as the main source of the disease.

**Action**

He removed the pump handle, interrupting the transmission of the disease.


**Ricardo São João**

# Smallpox Vaccine

## Development

Edward Jenner (1749–1823) developed the first effective smallpox vaccine in 1796.

## Implementation

In 1967, the WHO launched the eradication program through a strategy of surveillance and mass vaccination

## Eradication

In 1980, the WHO officially declared smallpox eradicated—the first human disease eliminated by medical science.



**Ricardo São João**

# Fight Against AIDS

### Identification

Discovery of HIV as the causative agent of AIDS in 1983

### Surveillance

Epidemiological systems monitored the spread and risk behaviours.

### Quality of Life

Significant increase in life expectancy of those infected.

### Treatment

Development of highly effective antiretroviral therapies.



**Ricardo São João**

# And more “recently”... the COVID-19 vaccine.

Accelerated Development

Scientists developed vaccines in record time during 2020.

Immediate Impact

Significant reduction in hospitalization and mortality rates.

Epidemiological monitoring identified variants and guided strategies.

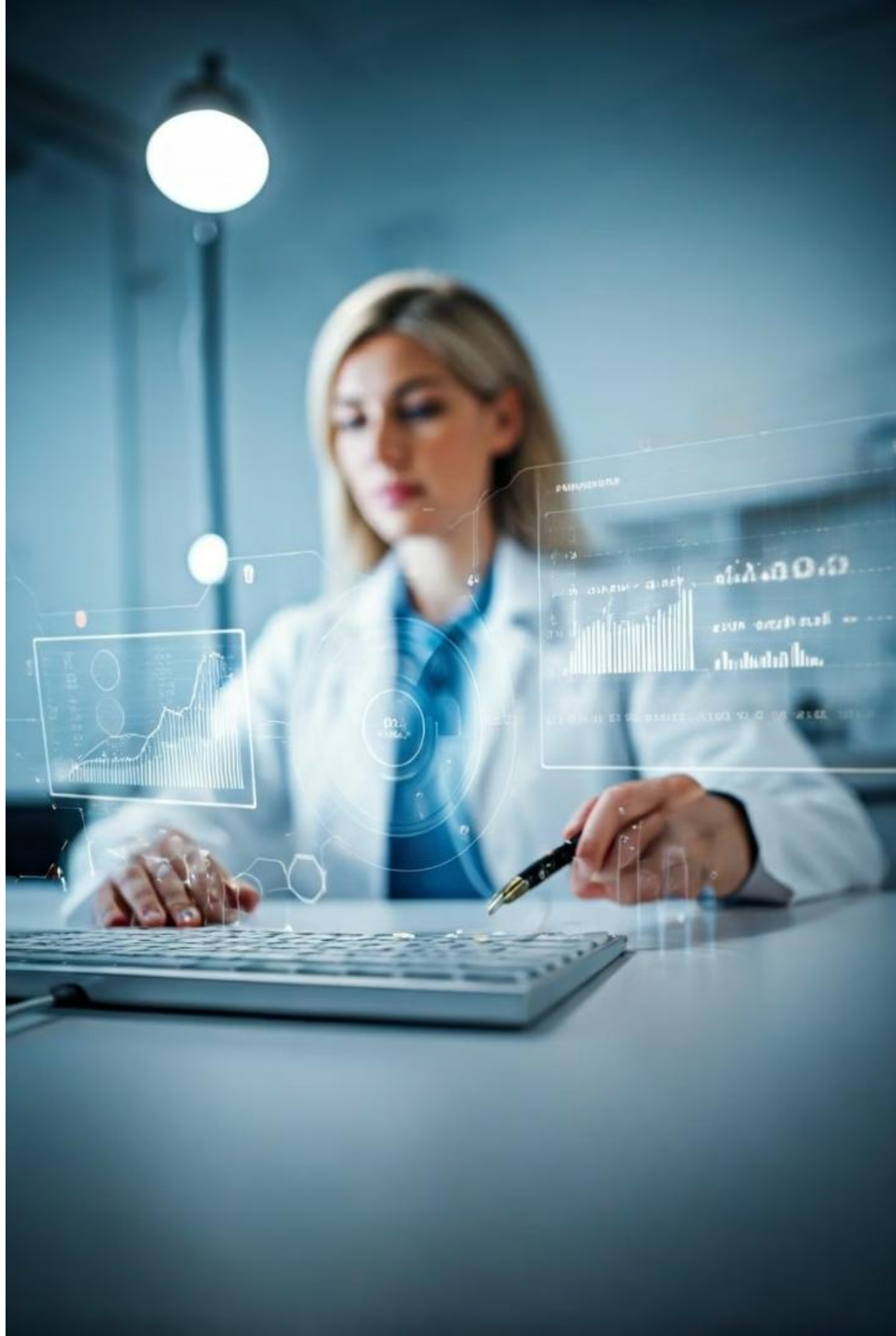
Continuous Surveillance



Ricardo São João

# Types of Epidemiological Studies and Systematic Review with Meta-Analysis (PICO)

Fundamental studies for the practice of evidence-based medicine.





**Ricardo São João**

# Classification of Epidemiological Studies

## Observational Studies

The researcher observes **without intervening.**

## Experimental Studies

They include **active intervention** by the researcher. Higher methodological rigor.

## Descriptive Studies

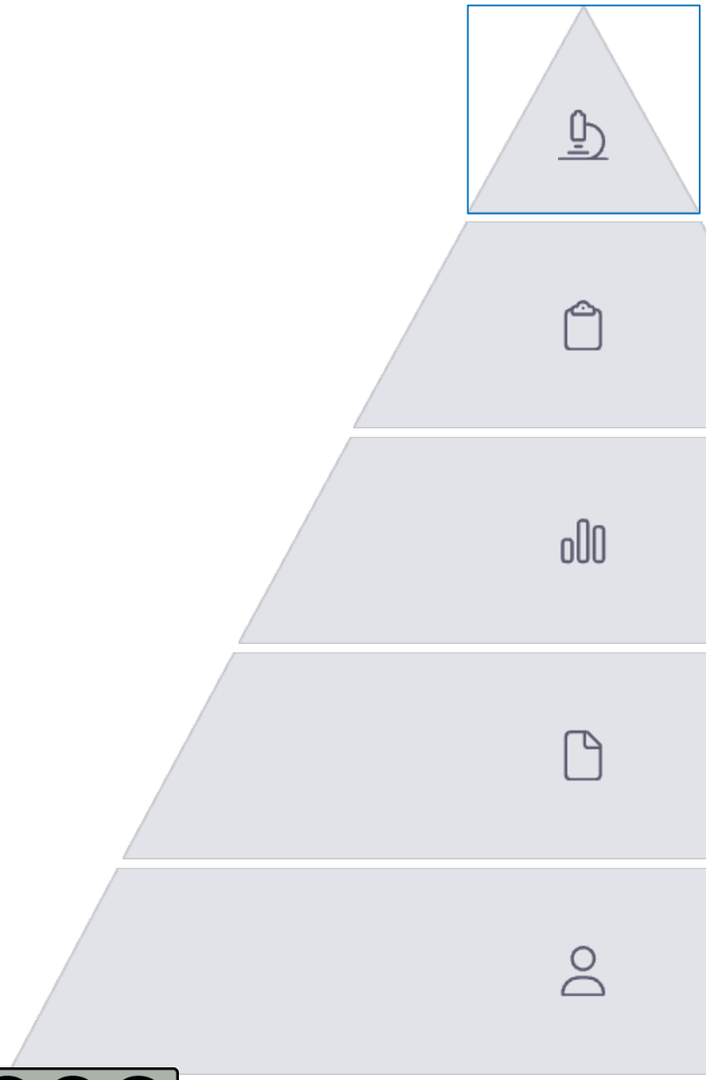
They characterise the frequency and distribution of events. **They generate initial hypotheses**.

## Analytical Studies

They **test hypotheses** about causal relationships. They provide the basis for health policies.


**Ricardo São João**

# Observational Studies - Descriptive and Analytical



**Case Report**

Detailed description of a single or rare case

**Case Series**

Collection of similar cases

**Cross-Sectional Studies**

Simultaneous measurement of exposure and disease

**Case-Control**

Retrospective, compares diseased and non-diseased individuals

**Cohort**

Prospective, follows exposed and non-exposed groups over time



**Ricardo São João**

# Experimental Studies

## Randomized Clinical Trials (RCT)

Participants randomly assigned to groups.
Gold standard for testing interventions..

## Field Trials

Conducted with healthy individuals at risk. Focus on primary prevention.

## Community Trials

Applied to entire populations. Assess the impact of public policies.


Ricardo São João

# Systematic Review: Concept and Methodology

## Formulation of the Question

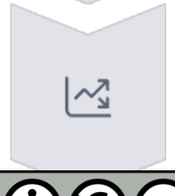
Clear definition of the problem using the PICO framework.

## Systematic Search

Comprehensive search across multiple databases.

## Study Selection

Application of inclusion/exclusion criteria.



## Qualitative Synthesis

Critical analysis and narrative summary of the results..

**Ricardo São João**

# Meta-Analysis: Quantitative Aggregation of Results

**Data Extraction**

Systematic collection of statistical data from included studies. Focus on comparable effect measures.

.

**Statistical Analysis**

Application of fixed or random effects models. Assessment of heterogeneity among studies.

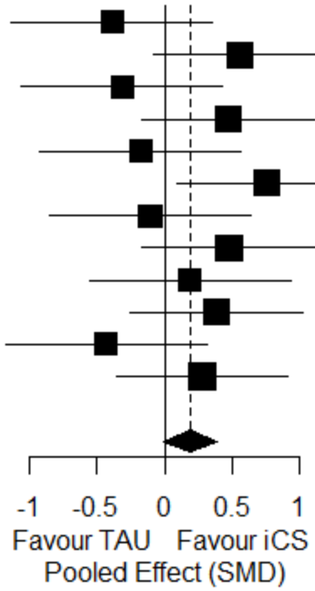
**Overall Interpretation**

Synthesis of the effect estimate. Assessment of precision and statistical significance..


Ricardo São João

| Study | iCS group Total | Mean | SD | TAU group Total | Mean | SD | SMD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| MMSE - Alzheimer's disease severity: Moderately severe | 14 | -0.86 | 3.3700 | 14 | 0.50 | 3.3900 | -0.39 | [-1.14; 0.36] | 7.0% |
| MMSE - Alzheimer's disease severity: Moderate | 16 | 1.81 | 3.7600 | 24 | -0.54 | 4.3200 | 0.56 | [-0.08; 1.21] | 9.5% |
| ADAS-COG - Alzheimer's disease severity: Moderately severe | 14 | -5.00 | 6.4100 | 14 | -3.00 | 5.9000 | -0.32 | [-1.06; 0.43] | 7.1% |
| ADAS-COG - Alzheimer's disease severity: Moderate | 16 | 0.06 | 8.6200 | 24 | -3.75 | 7.4900 | 0.47 | [-0.17; 1.11] | 9.6% |
| MAT - Alzheimer's disease severity: Moderately severe | 14 | -0.79 | 7.6400 | 14 | 0.29 | 3.0200 | -0.18 | [-0.92; 0.56] | 7.2% |
| MAT - Alzheimer's disease severity: Moderate | 16 | 4.50 | 6.4700 | 24 | -0.54 | 6.6100 | 0.75 | [ 0.10; 1.41] | 9.2% |
| FCSRT IR - Alzheimer's disease severity: Moderately severe | 14 | -3.79 | 8.5900 | 14 | -2.93 | 6.7400 | -0.11 | [-0.85; 0.63] | 7.2% |
| FCSRT IR - Alzheimer's disease severity: Moderate | 16 | 4.56 | 8.6500 | 24 | -0.12 | 10.1800 | 0.48 | [-0.16; 1.12] | 9.6% |
| FCSRT DR - Alzheimer's disease severity: Moderately severe | 14 | -0.64 | 2.0200 | 14 | -1.07 | 2.3700 | 0.19 | [-0.55; 0.93] | 7.2% |
| FCSRT DR - Alzheimer's disease severity: Moderate | 16 | 1.25 | 2.7400 | 24 | -0.08 | 3.8000 | 0.38 | [-0.26; 1.02] | 9.7% |
| FAB - Alzheimer's disease severity: Moderately severe | 14 | -1.14 | 2.8500 | 14 | -0.14 | 1.4100 | -0.43 | [-1.18; 0.32] | 7.0% |
| FAB - Alzheimer's disease severity: Moderate | 16 | 0.00 | 2.0000 | 24 | -0.50 | 1.5600 | 0.28 | [-0.36; 0.92] | 9.8% |
| **Common effect model** | **180** | | | **228** | | | **0.19** | **[-0.01; 0.39]** | **100.0%** |

Heterogeneity: $I^2$ = 23.1%, $\tau^2$ = 0.0346, $p$ = 0.2163




*Abbreviations:*

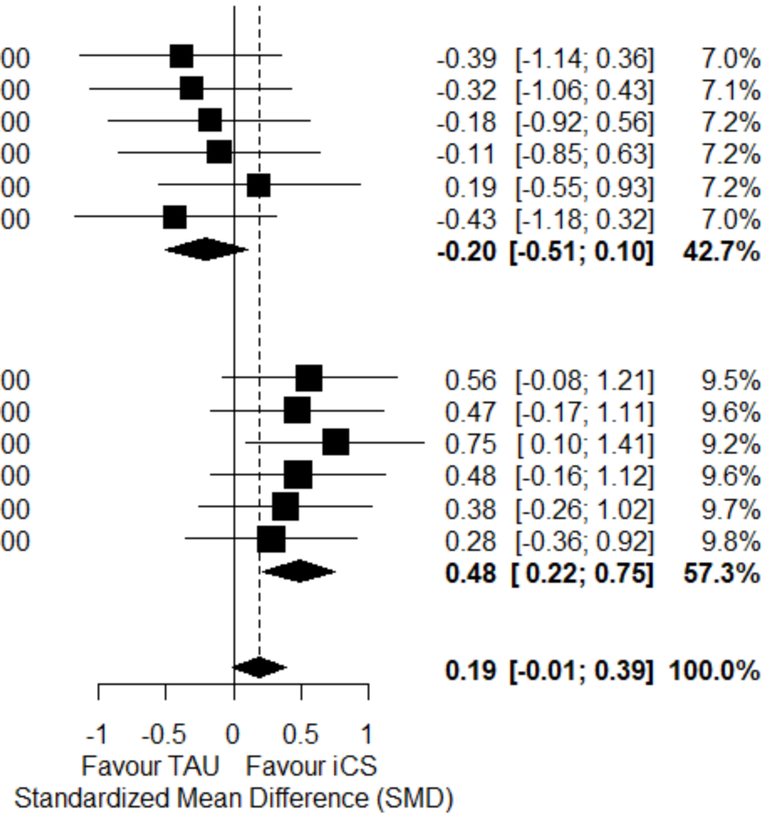
ADAS-COG= Alzheimer's Disease Assessment Scale-Cognitive;ADS= Alzheimer's disease severity; FAB= Frontal Assessment Battery;FCSRT IR= Free and Cued Selective Reminding Test Immediate Recall; FCSRT DR= Free and Cued Selective Reminding Test Delayed Recall; iCS= Individual Cognitive Stimulation; MMSE= Mini-Mental State Examination; MAT= Memory Alteration Test; TAU = Treatment As Usual.

*Note:* In ADAS-Cog higher scores reflecting lower performance.



Ricardo São João

| Study | iCS group Total | Mean | SD | TAU group Total | Mean | SD | SMD | 95%-CI | Weight |
|---|---|---|---|---|---|---|---|---|---|
| **Subscale = Alzheimer's disease severity: Moderately severe** | | | | | | | | | |
| MMSE - Alzheimer's disease severity: Moderately severe | 14 | -0.86 | 3.3700 | 14 | 0.50 | 3.3900 | -0.39 | [-1.14; 0.36] | 7.0% |
| ADAS-COG - Alzheimer's disease severity: Moderately severe | 14 | -5.00 | 6.4100 | 14 | -3.00 | 5.9000 | -0.32 | [-1.06; 0.43] | 7.1% |
| MAT - Alzheimer's disease severity: Moderately severe | 14 | -0.79 | 7.6400 | 14 | 0.29 | 3.0200 | -0.18 | [-0.92; 0.56] | 7.2% |
| FCSRT IR - Alzheimer's disease severity: Moderately severe | 14 | -3.79 | 8.5900 | 14 | -2.93 | 6.7400 | -0.11 | [-0.85; 0.63] | 7.2% |
| FCSRT DR - Alzheimer's disease severity: Moderately severe | 14 | -0.64 | 2.0200 | 14 | -1.07 | 2.3700 | 0.19 | [-0.55; 0.93] | 7.2% |
| FAB - Alzheimer's disease severity: Moderately severe | 14 | -1.14 | 2.8500 | 14 | -0.14 | 1.4100 | -0.43 | [-1.18; 0.32] | 7.0% |
| **Common effect model** | **84** | | | **84** | | | **-0.20** | **[-0.51; 0.10]** | **42.7%** |
| Heterogeneity: $I^2$ = 0%, $\tau^2$ = 0, $p$ = 0.8737 | | | | | | | | | |
| **Subscale = Alzheimer's disease severity: Moderate** | | | | | | | | | |
| MMSE - Alzheimer's disease severity: Moderate | 16 | 1.81 | 3.7600 | 24 | -0.54 | 4.3200 | 0.56 | [-0.08; 1.21] | 9.5% |
| ADAS-COG - Alzheimer's disease severity: Moderate | 16 | 0.06 | 8.6200 | 24 | -3.75 | 7.4900 | 0.47 | [-0.17; 1.11] | 9.6% |
| MAT - Alzheimer's disease severity: Moderate | 16 | 4.50 | 6.4700 | 24 | -0.54 | 6.6100 | 0.75 | [ 0.10; 1.41] | 9.2% |
| FCSRT IR - Alzheimer's disease severity: Moderate | 16 | 4.56 | 8.6500 | 24 | -0.12 | 10.1800 | 0.48 | [-0.16; 1.12] | 9.6% |
| FCSRT DR - Alzheimer's disease severity: Moderate | 16 | 1.25 | 2.7400 | 24 | -0.08 | 3.8000 | 0.38 | [-0.26; 1.02] | 9.7% |
| FAB - Alzheimer's disease severity: Moderate | 16 | 0.00 | 2.0000 | 24 | -0.50 | 1.5600 | 0.28 | [-0.36; 0.92] | 9.8% |
| **Common effect model** | **96** | | | **144** | | | **0.48** | **[ 0.22; 0.75]** | **57.3%** |
| Heterogeneity: $I^2$ = 0%, $\tau^2$ = 0, $p$ = 0.9449 | | | | | | | | | |
| **Common effect model** | **180** | | | **228** | | | **0.19** | **[-0.01; 0.39]** | **100.0%** |
| Heterogeneity: $I^2$ = 23.1%, $\tau^2$ = 0.0346, $p$ = 0.2163 | | | | | | | | | |
| Test for subgroup differences: $\chi^2_1$ = 11.29, df = 1 ($p$ = 0.0008) | | | | | | | | | |




Note: in ADAS-Cog, higher scores reflect lower performance.

| P | I | C | O |
|---|---|---|---|
| Population/ Problem | Intervention/ Exposure | Comparison | Outcome |
| Population characteristics or health issue of interest | Drug, surgery, policy, community program, etc. | No intervention, common practice, control group | Health outcomes of interest |

In medicine, PICO is an acronym that represents a **strategy to formulate research questions** and subsequently seek evidence to answer those questions.


**Ricardo São João**

# PICO Elements

**Patients/Population/Problem**

Defines the patient group or population of interest

**Intervention**

Indicates the treatment, procedure, or exposure of interest.

**Comparison**

Defines the condition to compare with the intervention (e.g., a placebo or standard treatment).

**Outcome**

Specifies the outcomes or endpoints to be evaluated (mortality, quality of life, etc.).


**Ricardo São João**

# Example

Does the administration of sertraline hydrochloride, compared to a placebo, reduce depression scores in patients diagnosed with moderate to severe major depressive disorder?” Source:: https://www.blog.psicometriaonline.com.br/estudos-de-revisao-sistematica-o-uso-da-estrategia-pico/

Note: Sertraline hydrochloride is a drug primarily used to treat depression (antidepressant), anxiety disorders, Obsessive-Compulsive Disorder (OCD), and post-traumatic stress disorder. It is available in prescription-only tablets.


**Ricardo São João**

| Abbreviation | Description | Example |
|---|---|---|
| P | Population/Patients | Patients diagnosed with moderate to severe major depressive disorder |
| I | Intervention/Exposure | Administration of sertraline hydrochloride |
| C | Comparison | Administration of placebo pill |
| O | Outcome | Depression scores |


Ricardo São João

## How it works

1. Define the clinical question
2. PICO Strategy
3. Search for evidence

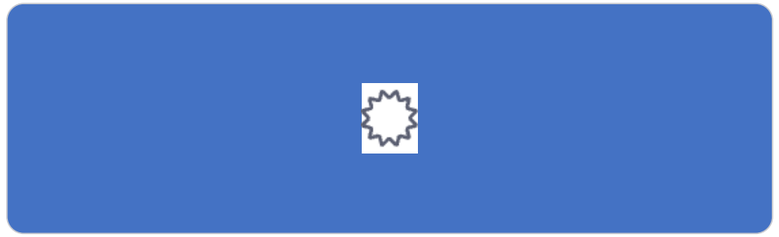
## Benefits

1. Precision in research questions
2. Better use of resources
3. Improved decision-making

# Synthesis

**Meta-Analysis with PICO**

Highest level of scientific evidence



**RCT**

Gold standard for interventions

**Observational Studies**

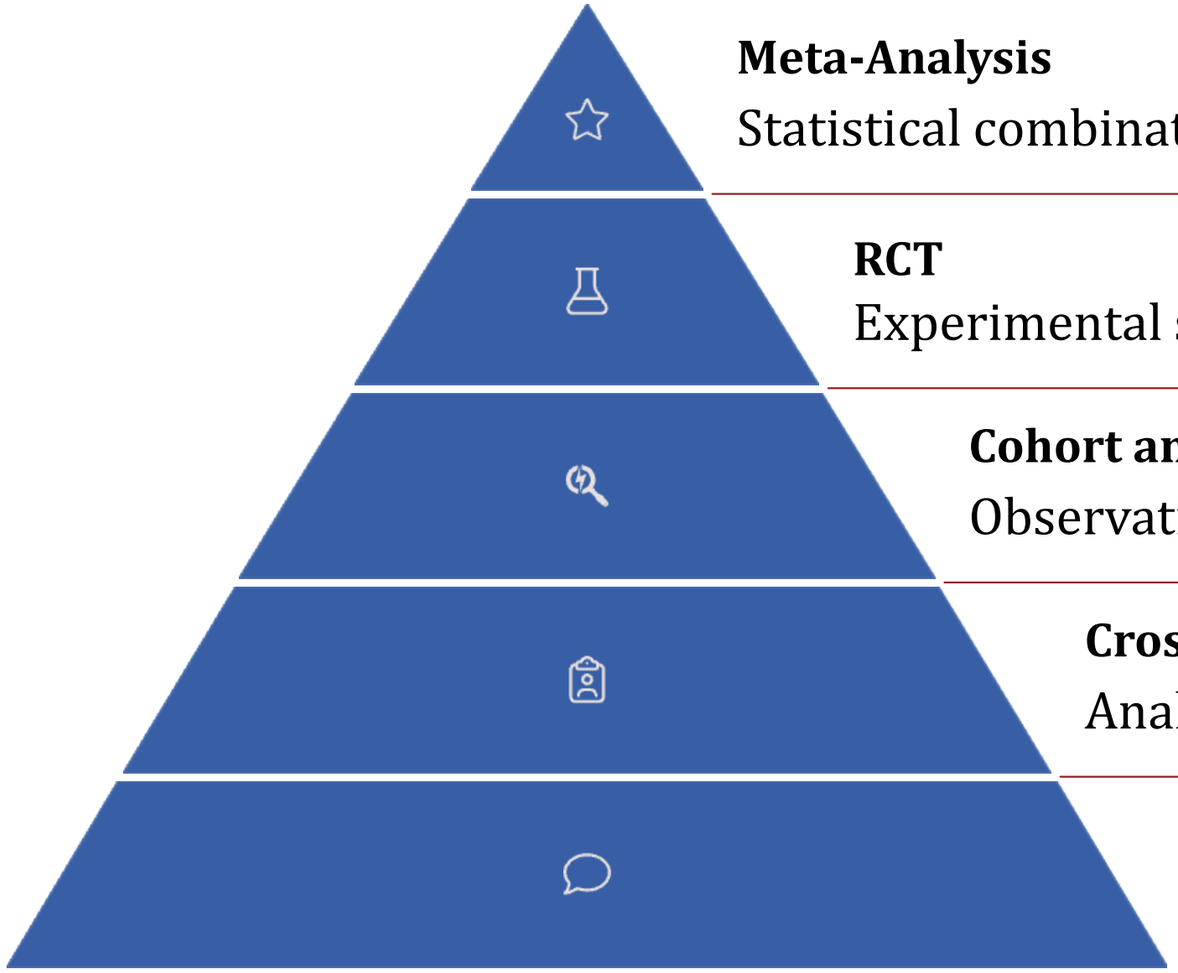
Fundamental for descriptive epidemiology

The methodological choice should align with the scientific question. Each study type has specific strengths.



**Ricardo São João**

# Hierarchy of methodological robustness



**Meta-Analysis**
Statistical combination of results from multiple studies.

**RCT**
Experimental studies with participant randomisation.

**Cohort and Case-Control Studies**
Observation of groups over time or retrospectively.

**Cross-Sectional Studies and Case Series**
Analysis of data at a specific point in time.

**Expert Opinions**
Based on individual clinical experience

The higher up the pyramid, the lower the risk of bias and the greater the reliability of the results.



**Ricardo São João**

# **Scientific Writing in Health**: focus on Sleep Medicine and related fields

Scientific writing is a fundamental skill for all who wish to contribute to the advancement of knowledge.

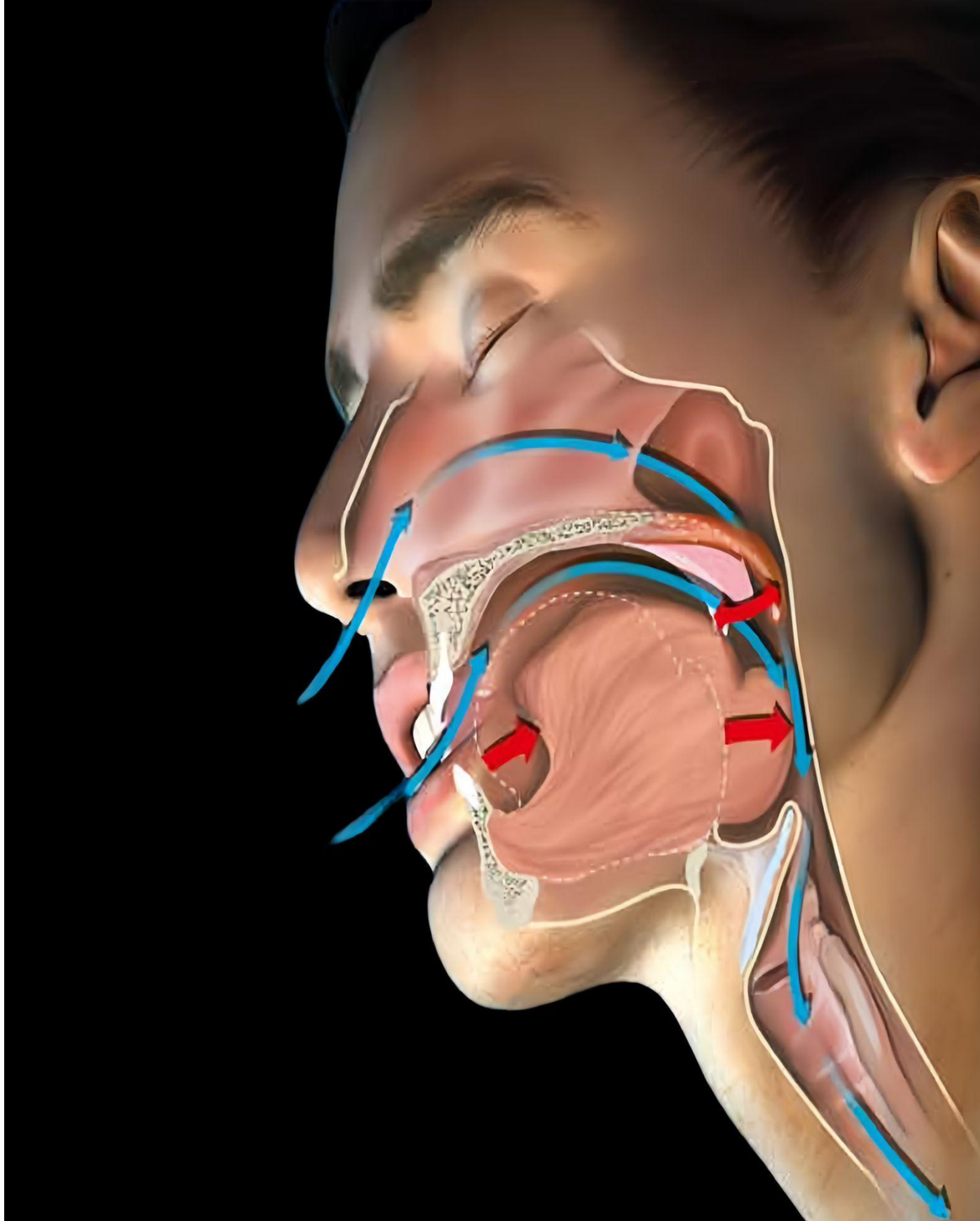


**Ricardo São João**

1. What are the essential components of a scientific article?

2. How to communicate results?

3. What mistakes should be avoided in scientific writing?



ada de https://omaxstudio.com.br/ronco-apneia/

# Motivation for Scientific Writing

## Effective Communication

The communication of scientific discoveries with rigor and clarity enables others to understand and apply the results in clinical practice, ensuring that knowledge is effectively shared

## Evolution in Clinical Practices

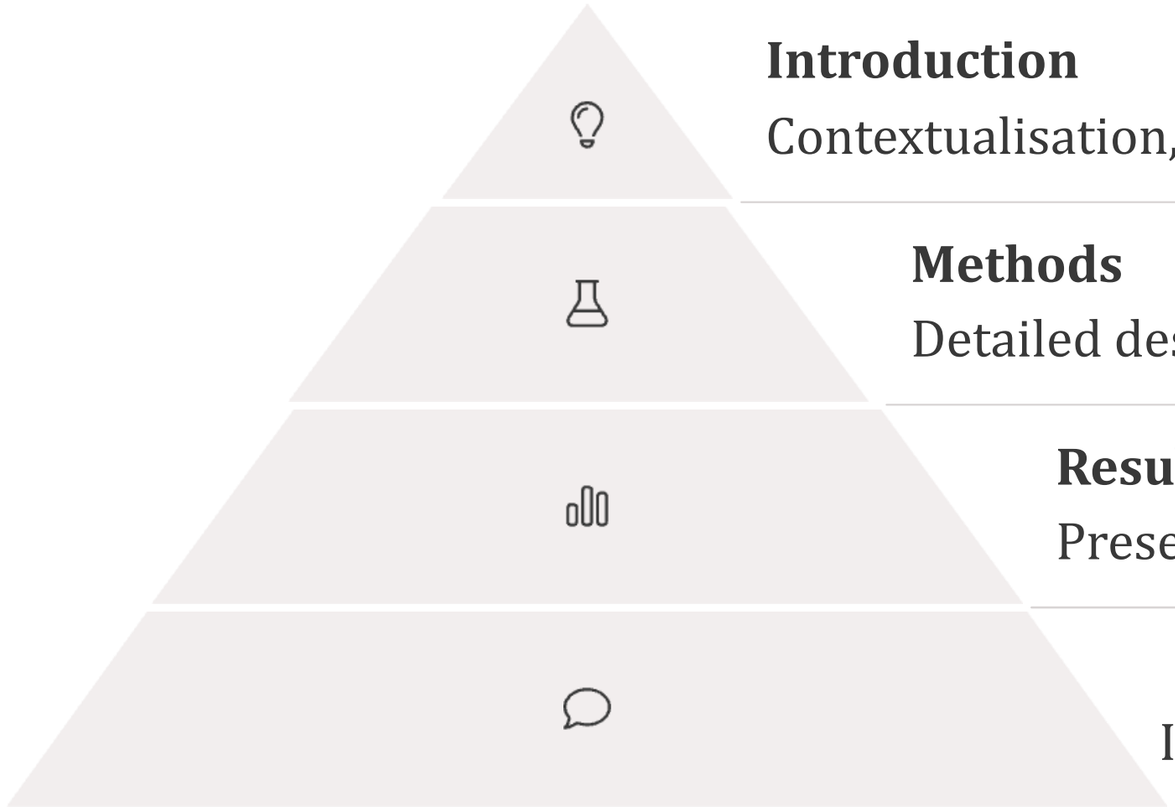
Well-written articles directly contribute to the improvement of patient care (Evidence-based medicine) by promoting continuous updating.

## Visibility and Reproducibility

A clearly described methodology allows other researchers to replicate studies, validating the results and building a solid and reliable body of scientific knowledge


**Ricardo São João**

# IMRAD Structure of the Scientific Article



**Introduction**

Contextualisation, identification of the gap, and definition of the objective

**Methods**

Detailed description of the methodology used

**Results**

Presentation of the data obtained **without interpretation**

**Discussion**

Interpretation, implications, and limitations.



**Ricardo São João**

# IMRAD Structure

- It is universally accepted in the health sciences
- The logical organisation facilitates reading and evaluation of the work by the scientific community, allowing quick location of relevant information.
- Helps authors organise their thinking clearly and objectively during the writing process.


Ricardo São João

# What makes a good Introduction?

## Presentation of the Problem

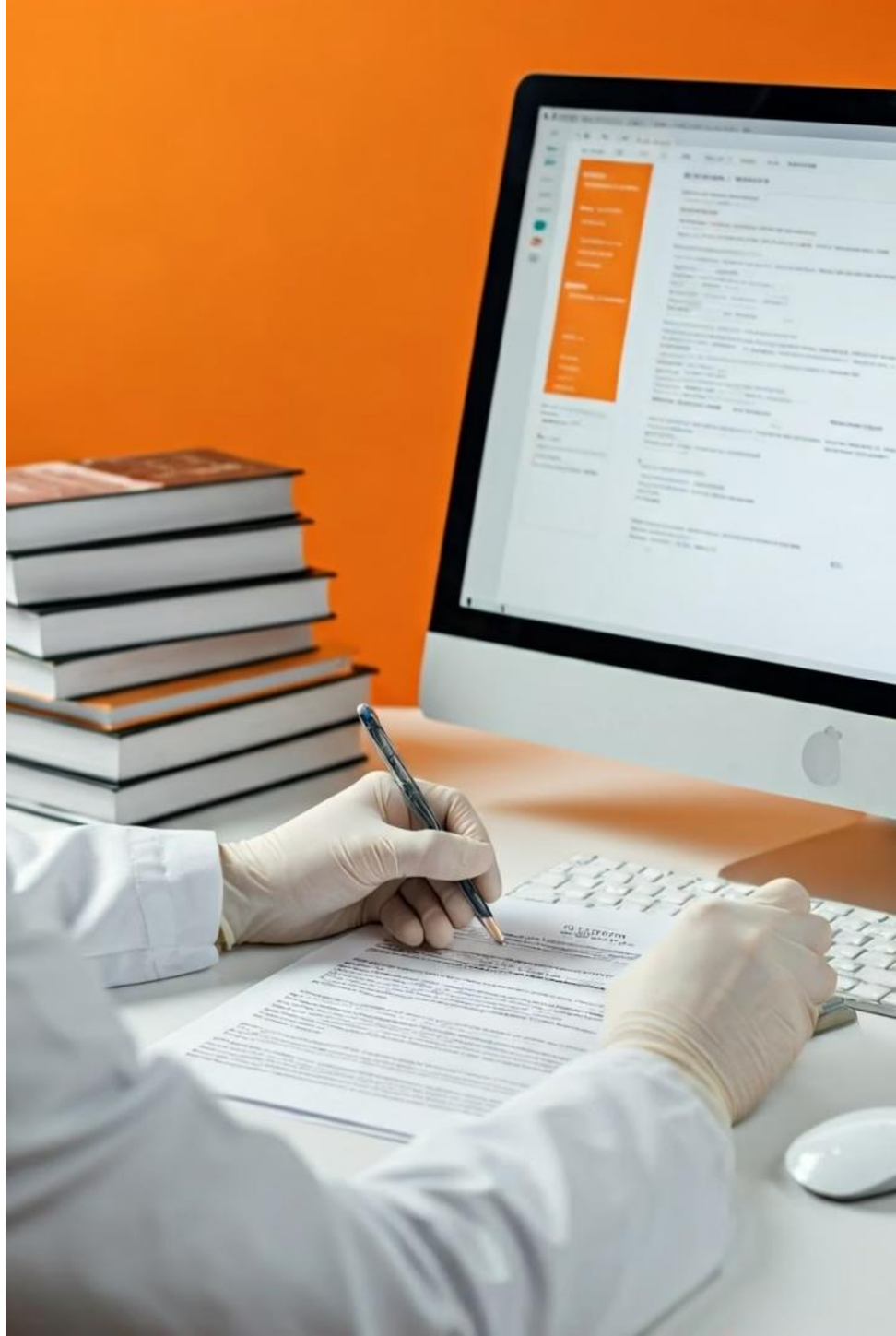
Clearly identify the research problem and justify its clinical relevance for Sleep Medicine and related fields.

## Literature Review

Summarise the most relevant and current studies on the topic, highlighting the knowledge gap your work intends to fill.

## Definition of the Objective

Formulate clear and specific objectives, preferably accompanied by hypotheses that can be empirically tested..





Ricardo São João

# Methods: the heart of reproducibility

## Type of Study

- Clearly specify the study design and justify your choice in relation to the defined objectives.

- Describe the location and period of data collection.
- Mention obtained ethical approvals

## Population and Sample

- Detail the inclusion and exclusion criteria used, the sampling process, and the sample size calculation..
- Justify the sample size.

- Describe demographic characteristics.

## Procedures and Analysis

- Thoroughly describe the instruments, equipment, and techniques used, as well as the statistical methods applied in data analysis.
- Specify the software used.
- Indicate specific statistical tests.


**Ricardo São João**

# Results: clarity above all!

| What To Do | What To Avoid |
| --- | --- |
| Present data objectively | Interpret the results in this section |
| Use tables and figures for complex data | Repeat the same data in text and tables |
| Include measures of central tendency and dispersion | Present only p-values without context |
| Organise the results according to the order of the objectives | Include methodological details |
| Highlight the main results first | Omit negative or contradictory results |

**In sleep medicine and the like, it is important to clearly present baseline values and the changes observed after interventions, making it easier to understand the clinical impact.**


**Ricardo São João**

# Discussion

**Interpretation**

Explain the significance of the results in the light of current knowledge

**Context**

Compare the findings with previous studies, explaining agreements and discrepancies

**Implications**

Discuss the clinical and theoretical implications of the results

**Limitations**

Recognise ‘honestly’ the methodological limitations of the study

For studies in sleep medicine and related fields, it is especially important to discuss the practical implications for diagnosis and treatment, helping to translate research into clinical practice.



**Ricardo São João**

# Some bibliographic databases

## PubMed

It contains more than **38 million** biomedical and life sciences citations.

Offers access to high-quality peer-reviewed medical literature.

## Scopus

It has 20% greater coverage when analysing citations compared to Web of Science.

Includes scientific journals, books and conference proceedings.

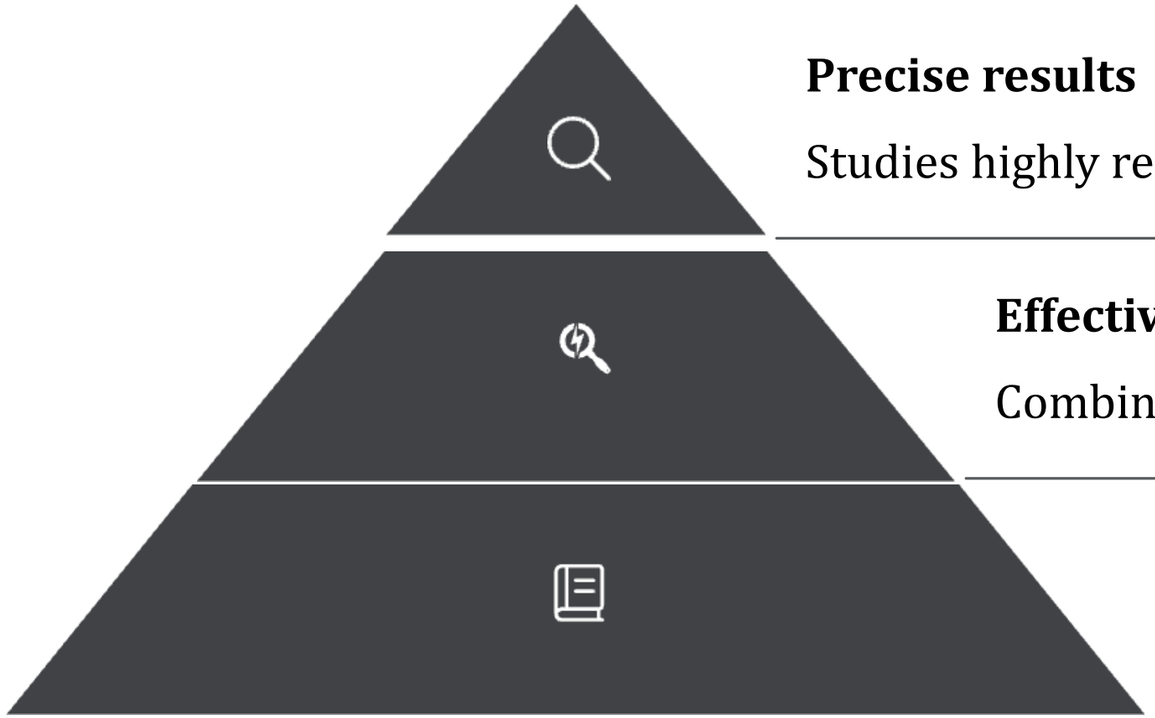
## Google Scholar

Multidisciplinary database with more than 400 million scientific documents.

Ideal for interdisciplinary research.


**Ricardo São João**

# Optimising Bibliographic Research



**Precise results**

Studies highly relevant to the research question

**Effective Strategies**

Combined use of specific databases

**Structured Keywords**

MeSH for PubMed and DeCS for Lusophone literature/Spanish/English

APA 7th ed. citation generators ensure bibliographic standardisation.

MeSH = Medical Subject Headings. This is a controlled vocabulary created and maintained by the US National Library of Medicine (NLM).

It is used mainly to index articles in the PubMed/MEDLINE database, organising them systematically by subject.

In analogue DeCS stands for Descritores em Ciências da Saúde (Descriptors in Health Sciences).



**Ricardo São João**

# Common mistakes in scientific publications

### Lack of focus on the goal

Define a precise and measurable research question. Align the methodology with specific objectives.

### Vague methods or poorly described

Detail procedures to allow reproducibility. Specify statistical analyses with clear justifications.

### Inadequate discussion

Interpret results within the limitations of the study. Contextualise findings in existing literature.



**Ricardo São João**

# Ethics and Scientific Integrity

### Ethical approval

Obtain approval from the ethics committee before starting any clinical study

### Responsible Publication

Avoid duplicate publication and declare all conflicts of interest

### Informed Consent

Document the free and informed consent of all participants.

### Plagiarism prevention

Make appropriate citations and recognise all the sources consulted



**Ricardo São João**

# Scientific Writing in Sleep Medicine

Alotaibi AD, Alosaimi FM, Alajlan AA, Bin Abdulrahman KA. The relationship between sleep quality, stress, and academic performance among medical students. J Fam Community Med 2020;27:23-8

https://journals.lww.com/jfcm/fulltext/2020/27010/the_relationship_between_sleep_quality,_stress,.4.aspx

The cross-sectional study by Alotaibi et al. (2020) on sleep quality and academic performance in university students exemplifies the structure of a scientific article in the field of sleep medicine. The use of validated sleep assessment instruments, such as the Pittsburgh Sleep Quality Index, was described in sufficient detail to ensure reproducibility.

# Scientific Writing in Sleep Medicine

64%

Students with poor sleep quality

Percentage of university students reporting sleep problems

42%

Reduced academic performance

Correlation between insufficient sleep and low ratings

3.2x

Higher risk of depression

Increased risk at students with chronic insomnia

The results were clearly presented and highlighted the relationship between disturbed sleep patterns and reduced academic performance, with implications for public health and educational policies.



**Ricardo São João**

São João, R., Cardoso, A., Domingues, T.D., Fradinho, M., Silva, V., Feliciano, A. (2022). A Retrospective Study on Obstructive Sleep Apnea. In: Bispo, R., Henriques-Rodrigues, L., Alpizar-Jara, R., de Carvalho, M. (eds) Recent Developments in Statistics and Data Science. SPE 2021. Springer Proceedings in Mathematics & Statistics, vol 398. Springer, Cham. https://doi.org/10.1007/978-3-031-12766-3_19

São João, R., Cardoso, A. , Domingues, T. D., Silva, V., Fradinho, M., Santos, L. & Feliciano, A.(Outubro, 2021). Observational and comparative study between automatic and manual analysis of sleep studies. XXV Congresso da Sociedade Portuguesa de Estatística, Evora http://hdl.handle.net/10400.15/3664

# Scientific Writing in Oral Health

A good example of scientific writing in oral health is the cohort study on poor oral hygiene and periodontitis in young adults

*Tonetti, M. S., et al. (2017). Staging and grading of periodontitis. J Clin Periodontol*
https://onlinelibrary.wiley.com/doi/epdf/10.1111/jcpe.12681

This work stood out for the clarity of its objective, its methodological robustness in assessing periodontal parameters and for using accessible language, even when dealing with complex technical concepts..

As demonstrated by Tonetti et al. (2017), the use of well-defined diagnostic criteria and the systematic presentation of results are essential for studies in the field of periodontology, allowing for international comparisons and the development of clinical guidelines.

# Scientific Writing in Oral Health

**Characterising the problem**

Epidemiology of periodontitis in young adults

**Specific methods**

Standardised clinical and microbiological assessmenta

**Specific methods**

Impact of interventions on periodontal health

# Scientific Writing in TMD

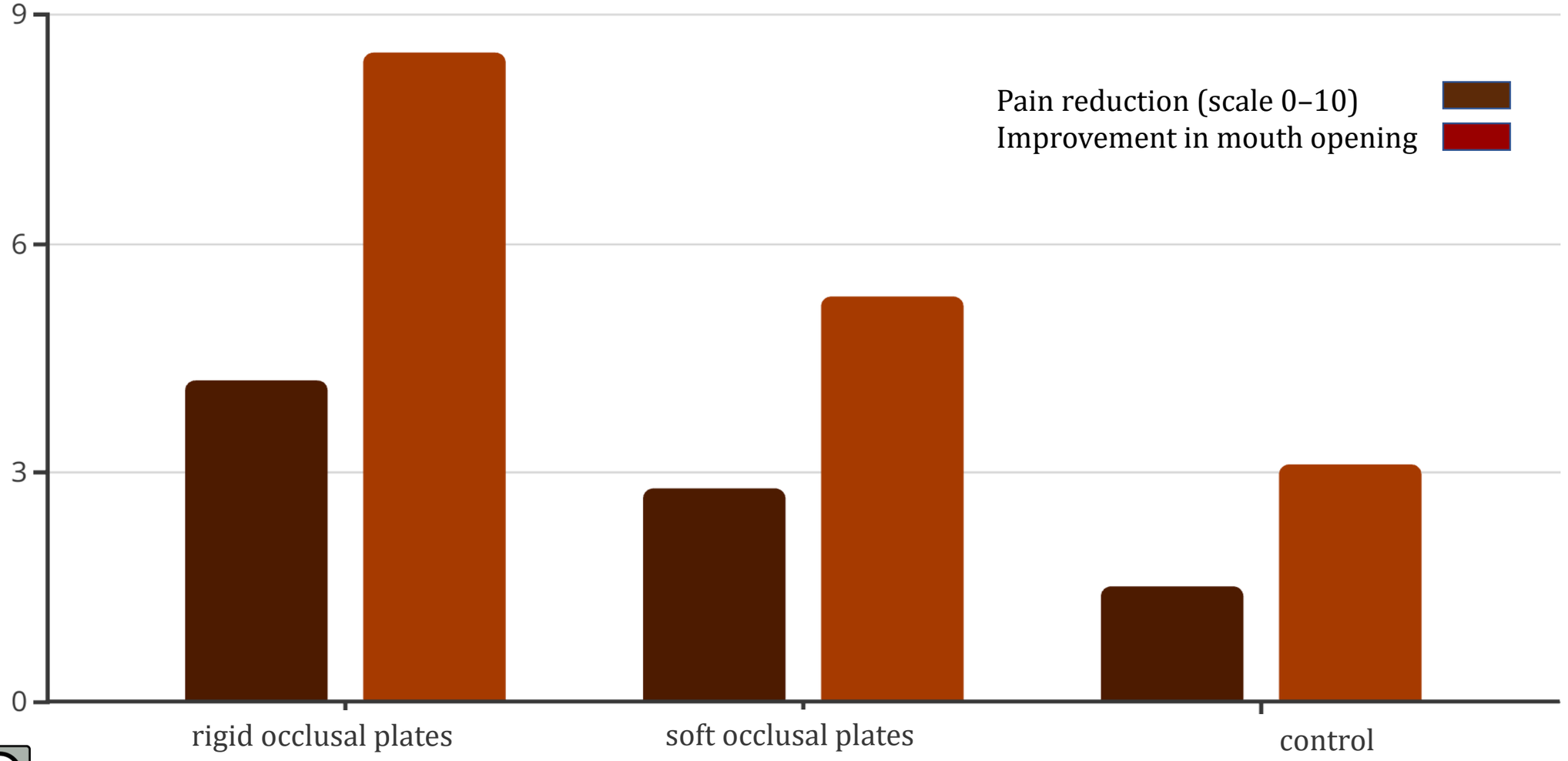
*Alajbeg, I. Z., et al. (2014). Occlusal splint therapy in TMD. J Oral Rehabil.*

https://www.sciencedirect.com/science/article/pii/S1013905222001754?via%3Dihub

The randomised clinical trial by Alajbeg et al. (2014) comparing rigid and soft occlusal plates in TMD patients is exemplary for its robust methodological structure. The authors clearly defined the diagnostic criteria used, the protocols for making the plates and the evaluation parameters.


**Ricardo São João**

The results section presented comparative data on pain reduction, improvement in mouth opening, and electromyographic activity across the different treatment groups. The discussion correlated the findings with the underlying physiological mechanisms, providing a solid foundation for the direct clinical application of this knowledge.






**Ricardo São João**

# Some important statistical concepts when reading articles

## Hypothesis testing

Objective: to determine whether a given hypothesis about the population is plausible.

$H_0$: ……  vs  $H_1$: ……

**Null hypothesis**

**reflects the situation in which there is no change**

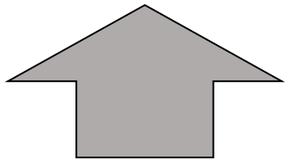
**Alternative hypothesis**

**hypothesis of equality**
**hypothesis of no effect**
**hypothesis of no difference**


Ricardo São João

# Hypothesis testing

**Objective:** to determine whether there is sufficient evidence to reject the null hypothesis



information available from the sample

**But when we make a decision, we are subject to error!**

**What errors are associated with decision-making?**


Ricardo São João

# Mechanism of Errors: analogy with the trial of a defendant in court

| Judge \ Defendant | innocent (Ho TRUE) | guilty (Ho FALSE) |
|---|---|---|
| ABSOLVE (Não Rejeita Ho) | Ok (1-α) | **Type II error= β** |
| CONDENA (Rejeita Ho) | **Type I error= α** | Ok (1-β) |

$P(\text{Reject } H_0 | H_0 \text{ TRUE}) = \alpha$ (level of significance).

Set by the researcher. Usually set at 1%, 5%, or 10%.

$P(\text{Not reject } H_0 | H_0 \text{ FALSE}) = \beta$;

$1 - \beta$=statistical power


Ricardo São João

## Hypothesis testing

**How to decide? Several ways. We will opt for the p-value.**

Reject $H_0$ if **valor-*p*** $\leq \alpha$

Interpretation:

| | | |
|---|---|---|
| $p > 0.1$ | ⟶ | Little or no evidence of a difference or relationship |
| $0.05<p<0.1$ | ⟶ | weak evidence of a difference or relationship |
| $0.01<p<0.05$ | ⟶ | evidence of a difference or relationship |
| $p<0.01$ | ⟶ | strong evidence of a difference or relationship |
| $p<0.001$ | ⟶ | very strong evidence of a difference or relationship |


Ricardo São João

# Effect measures in contingency tables

Relative Risk (RR) is a measure used in prospective studies where individuals are classified as exposed or unexposed to a certain factor, and the outcome (disease) is analysed.


Ricardo São João

# Risk estimation in a cohort study

**Incidence of a disease = absolute risk**

| | Disease progression /Death | No disease progression /Survival | Total |
|---|---|---|---|
| Exposed (E+) | a | b | a+b |
| Not Exposed (E-) | c | d | c+d |
| | a+c | b+d | N |

**Risk among the Exposed**
**RE+=a/(a+b)**

**Risk among the Unexposed**
**RE-=c/(c+d)**

**RR=RE+/RE-**

$$= \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}$$

**Relative Risk (RR) is a ratio of probabilities**


Ricardo São João

## Interpretation of RR for a disease

- **RR = 1**: Risk among the exposed = Risk among the unexposed **(no association)**;

- **RR > 1**: Risk among the exposed > Risk among the unexposed (**positive association; possibly causal?**);

- **RR < 1**: Risk among the exposed < Risk among the unexposed (**negative association; possible protective effect?)**.

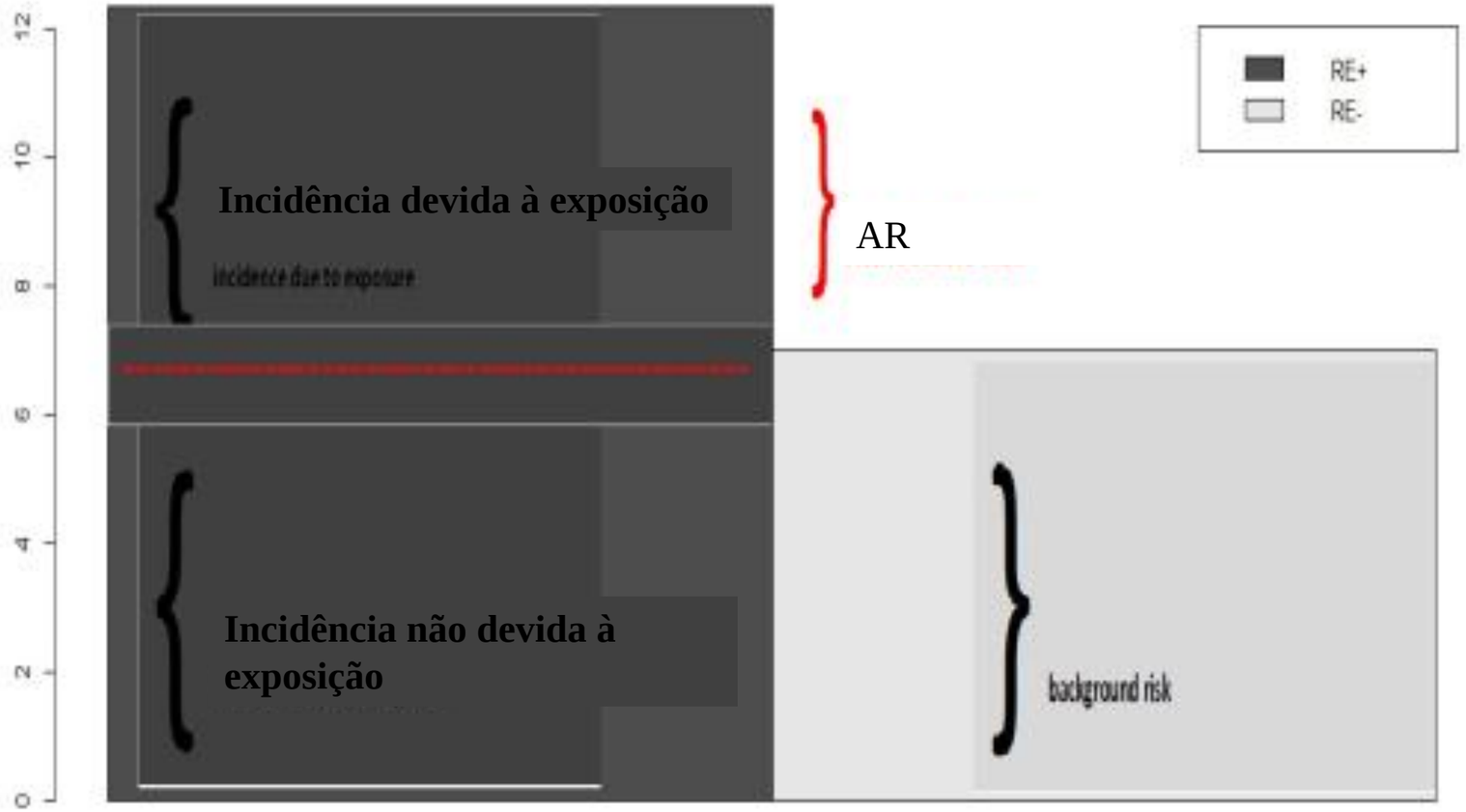

Ricardo São João

Attributable Risk (AR) is a measure of the excess risk that is attributed to the exposure.

- AR= RE+ - RE-
- AR= risk difference
- Proportion of Attributable Risk among the exposed = $\frac{AR}{RE^{+}}$


Ricardo São João

**AR** $= R_{E+} - R_{E-}$




risk

# *Let's get to work!*

Daly et al. (1983) conducted a study with 368 male patients who were smokers and had suffered a heart attack. These patients were divided into two groups:

(i) a group of patients who, after the heart attack, continued smoking;
(ii) a group of patients who, after the heart attack, stopped smoking.

The patients were followed for two years to analyse the relationship between smoking cessation and mortality. The following table summarises the results:

| | **death** | **life** | **Total** |
|---|---|---|---|
| **Exposed (E+)** | **19** | **135** | **154** |
| **Non Exposed (E-)** | **15** | **199** | **214** |
| | **34** | **334** | **368** |

Determine all the risks


**Ricardo São João**

# Risk calculation and interpretation

| | death | life | Total |
|---|---|---|---|
| Exposed (E+) | 19 | 135 | 154 |
| Non Exposed (E-) | 15 | 199 | 214 |
| | 34 | 334 | 368 |

RE+=a/(a+b)

RE-=c/(c+d)

RE+=19/(19+135)=12,34 (per 100 individuals)

Risk of death for a patient who continues to smoke

RE-=15/(15+199)=7 (per 100 individuals)

Risk of death for a patient who quits smoking

$$RR=RE+/RE- = \frac{\frac{19}{(19+135)}}{\frac{15}{(15+199)}} = 1,76$$

A patient who continues to smoke has a 1.76 times higher risk of death compared to a patient who has quit smoking.



Ricardo São João

# Risk calculation and interpretation

| | death | life | Total |
|---|---|---|---|
| **Exposed (E+)** | **19** | **135** | **154** |
| **Non Exposed (E-)** | **15** | **199** | **214** |
| | **34** | **334** | **368** |

**RA=RE+ - RE- =12,34-7=5,34 (per 100 individuals)**

Total Risk=Incidence Rate= $\frac{34}{368} = 9{,}24$ **(per 100 individuals)**

**Increased risk of death for patients who continue to smoke**

**Proportion of AR among the exposed** $= \frac{AR}{RE^+} = \frac{5{,}34}{12{,}34} = 43\%$

**43% of the risk of death is due to not quitting smoking.**


**Ricardo São João**

# Odds

**Odds= Number of occurrences / Number of non-occurrences**

Example: Out of 100 births, 60 are boys and 40 are girls. The odds of being a boy is 60/40 = 1.5..

| | Disease progression /case | No disease progression /control | Total |
|---|---|---|---|
| **Exposed (E+)** | **a** | **b** | **a+b** |
| **Not Exposed (E-)** | **c** | **d** | **c+d** |
| | **a+c** | **b+d** | **N** |

The odds that an exposed person develops the disease = a/b

The odds that an unexposed person develops the disease = c/d


**Ricardo São João**

Odds Ratio (OR) is a measure of association used in retrospective studies, where the outcome is already known and the aim is to analyse the exposure (a possible cause of the outcome).

| | **Disease progression /case** | **No disease progression /control** | **Total** |
|---|---|---|---|
| **Exposed (E+)** | **a** | **b** | **a+b** |
| **Not Exposed (E-)** | **c** | **d** | **c+d** |
| | **a+c** | **b+d** | **N** |

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{cb}$$


**Ricardo São João**

## OR interpretation

- **OR = 1:** Exposure is not associated with the disease;
- **OR > 1:** Exposure is positively associated with the disease;
- **OR < 1:** Exposure is negatively associated with the disease


Ricardo São João

$$OR \ \frac{19 \times 199}{15 \times 135} = 1,87; RR = 1,76$$

OR and RR values are sometimes similar! Why is that?

**If the incidence/risk is low**, such that $a+b \approx b$ and $c+d \approx d$,
then $RR \approx OR$.


Ricardo São João

# Effect size

It is a standardized measure that shows how strong the effect of an intervention, association, or difference is, **regardless of sample size**.

**While the p-value indicates whether there is a difference (statistical significance), the effect size indicates how large that difference is (practical or clinical significance).**


Ricardo São João

## 1) Cohen's d

Used when comparing two means (e.g., control group vs. experimental group)

$$d = \frac{M_1 - M_2}{SD_{\text{pooled}}}$$ where: $M_1, M_2$: Average groups; $$SD_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

## 2) Pearson correlation coefficient (parametric) / Spearman (non-parametric) - R

## 3) R-squared ($R^2$)

## 4) Eta-squared ($\eta^2$)

## 5) Cramér's V


Ricardo São João

*Thank you for your* PPA

throughout this session!

* Presence, Participation,
and Attention!


**Ricardo São João**