

**DARE**  
**DIGITAL LIFELONG PREVENTION**  
**CODE NO. PNC0000002**

Spoke 1 Deliverable  
**S1.D1.1 List of process and  
outcome indicators**

This research is co-funded by the Ministry of University and Research  
within the Complementary National Plan PNC-I.1  
“Research initiatives for innovative technologies  
and pathways in the health and welfare sector”  
D.D. 931 of 06/06/2022, PNC0000002 DARE - Digital Lifelong Prevention

## S1.D1.1 List of process and outcome indicators

Deliverable information	
Spoke number and title	Spoke 1 - Enabling Factors and Technologies for Digital Prevention
WP number and title	WP1 - Ecosystem Building
Related task(s)	T1.1
Lead beneficiary	UCSC
Contributing beneficiaries	GIMBE, UNIBO, UNIBA
Dissemination level	Public, fully open
Due date	15 June 2023
Actual date of delivery	31 May 2023
Author(s)	Stefania Boccia (UCSC), Andrea Gentili (UCSC), Alessandra Maio (UCSC), Mario Cesare Nurchis (UCSC), Luigi Russo (UCSC)
Contributors	Giorgio Bedogni (UNIBO), Antonino Cartabellotta (GIMBE), Tiziano Innocenti (GIMBE)
Quality Assurance	Dario Gregori (UNIPD), Salvatore Vitabile (UNIPA)
Document version	1.0

## Document history

Version	Date	Author(s) /Reviewer(s) (Beneficiary)	Description
0.1	8 May 2023	Stefania Boccia (UCSC), Andrea Gentili (UCSC), Alessandra Maio (UCSC), Mario Cesare Nurchis (UCSC), Luigi Russo (UCSC)	First draft
0.2	29 May 2023	Leader WP1 SP2 (Salvatore Vitabile - UNIPA), WP1 Leader SP3 (Dario Gregori - UNIPD)	Revision
0.3	31 May 2023	Partners WP1 SP1	Final document
1.0	31 May 2023	HUB final check	Final version submitted

## Disclaimer

This publication reflects only the authors' views, and the Funding Agency is not liable for any use that may be made of the information contained therein.

## Table of contents

Document history .....	3
Disclaimer .....	3
Table of contents.....	4
Publishable Summary.....	5
List of abbreviations.....	7
1. Introduction.....	8
1.1. Objectives of the deliverable .....	8
1.2. What is an indicator?.....	8
1.3. Why are indicators essential in DARE?.....	8
1.4. Which indicators? Clinical studies vs. predictive model studies.....	9
2. Clinical studies: the GRADE Evidence to Decision Framework .....	9
2.1. List of indicators according to the two main GRADE EtD domains.....	10
2.1.1. Detailed explanation of each domain.....	11
2.1.2. Examples.....	13
3. Predictive models: the importance of defining indicators for proper performance assessment	13
3.1. List of indicators for predictive models' performance assessment .....	14
3.1.1. Detailed explanation of each domain.....	14
3.1.2. Examples.....	19
4. References .....	20
Appendix A.....	22
Appendix B.....	40

## Publishable Summary

The DARE project aims to provide the Italian Ministry of Health and NHS with innovative technologies that are effective, efficient, and ready to be used, enabling the country to accelerate and revolutionize the health paradigm encompassing prevention through to treatment..

Indicators are defined as specific, observable, and measurable changes showing progress toward achieving a specific output or outcome in a logic model or work plan. The purpose of this deliverable is to provide all DARE consortium partners with a list of possible indicators that will be used to assess the utility and implementability of innovative technologies during the development of a pilot.

The DARE project consortium involves many entities and promotes the implementation of a wide variety of studies that focus on prevention, and the maturity level of the proposed pilots could be very different. Two frameworks have been selected for the identification of indicators: the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) Evidence to Decision (EtD) framework, to be used by any partner that will implement a clinical study, and the Predictive Clinical Model framework, to be followed by any proponent that will focus on creating a predictive clinical model using artificial intelligence, machine learning, and deep learning.

The GRADE EtD framework is a tool used to inform and guide health system and public health decision-making. It presents a set of criteria and indicators that can be used to evaluate and assess the potential benefits, harms, feasibility, acceptability, and impact of an intervention or option. Each domain of the GRADE EtD is filled with a list of possible indicators that should be gathered in the course of the development of the pilots.

The list of indicators includes the seriousness of the problem, priority of the problem, clinical utility of the intervention, impact of the intervention, benefits and harms of the intervention, balance between the desirable and undesirable effects, equity and human rights, financial and economic considerations, values and preferences, acceptability, and feasibility.

The Predictive Clinical Model is a framework for developing and evaluating predictive models in clinical research. An indicator for predictive models is a metric or set of metrics used to evaluate the performance of a model in predicting specific outcomes. Such indicators are essential for ensuring that the model is effective in achieving the defined tasks and objectives. Indeed, without indicators, it is difficult to determine whether the model is working properly or whether improvements are needed.

The selection of appropriate indicators will depend on the type of predictive model and the application domain, and they should provide meaningful insights into the model's performance.

The list includes the seriousness of the problem, the priority of the problem, impact analysis, discrimination of the model, calibration of the model, equity and human rights, values and preferences, and feasibility.



## List of abbreviations

ACT: Acceptance and Commitment Therapy

AUC-ROC: Area Under the Receiver Operating Characteristic Curve

CBA: Cost-benefit Analysis

CDC: Center for Disease Control and Prevention

CEA: Cost-effectiveness Analysis

CHD: Coronary Heart Disease

CI: Confidence Interval

CUA: Cost-utility Analysis

CVD: Cardiovascular Disease

DALYs: Disability-adjusted life year

DCA: Decision Curve Analysis

DCIS: Ductal Carcinoma In Situ

EtD: Evidence to Decision

GRADE: Grading of Recommendations Assessment, Development and Evaluation

HIV: Human Immunodeficiency Virus

IVR: Interactive Voice Response

MMS: Multimedia Message Service

NHS: National Health Service

PROCAM: Prospective Cardiovascular Münster

QALYs: Quality-Adjusted Life Years

QoL: Quality of Life

ROC: Receiver-operating Characteristics

SF-12: 12-Item Short Form Health Survey

SF-36: 36-Item Short Form Health Survey

SMS: Short Message Service

SRH: Sexual and Reproductive Health and Research

STI: Sexually Transmitted Infections

WHO: World Health Organization

## 1. Introduction

### 1.1. Objectives of the deliverable

The main objective of the deliverable is to compile a list of process and outcome indicators to measure the utility and implementability of digital preventive interventions.

### 1.2. What is an indicator?

The Center for Disease Control and Prevention (CDC) states that an indicator is a “marker of accomplishment/progress. It is a specific, observable, and measurable accomplishment or change showing the progress toward achieving a specific output or outcome in your logic model or work plan [1].” Indicators help determine what data should be collected during the process. They both help the evaluation process and indicate in which direction it should go and what parameters to consider [2].

### 1.3. Why are indicators essential in DARE?

To properly assess the utility and implementability of innovative technology during the development of a pilot, it is vital to consider a number of actors and to take into account all the potential health-related and non-correlated outcomes that are relevant to citizens, patients, and all other stakeholders (e.g., clinicians and policymakers). Indeed, our ultimate goal is to provide the Italian Ministry of Health and NHS with innovative technologies that are at the forefront but, above all, effective, efficient, and ready to be used, enabling our country to accelerate and revolutionise the health paradigm, spanning from prevention to treatment.

Indicators are used to systematize as much as possible the uncertainty that innovative technologies bring, thus ensuring that all relevant aspects are taken into account when the utility and implementability of innovative technologies have to be assessed from a public health perspective. By harmonizing each of the required steps during the development and

implementation of the pilot, we will reach a comprehensive assessment ready to give precise answers and clear directions. The present document is intended as a guideline for researchers to ease the collection, compilation, and dissemination of key indicators and outcomes along the process of pilot development. The scope is to collect the relevant utility and implementability indicators through data collection and evidence synthesis across the pilot journey. An evidence synthesis (i.e., systematic review) should be conducted for each utility and implementability indicator (please refer to the “Supplementary Materials Review Guide” as a short guidance on review conduction) in order to make the formulation of final recommendations possible.

#### **1.4. Which indicators? Clinical studies vs. predictive model studies**

DARE is a consortium that involves many entities (Universities, private companies, non-profit Organizations) and that promotes the implementation of a wide variety of studies.

Those studies will regard the prevention topic (primary, secondary, and tertiary), but the maturity level of the proposed pilots could be very different.

We decided to assist the pilot’s proponent in the identification of indicators, providing two different frameworks to follow:

- The GRADE EtD framework: this is intended to be used by any partner that will implement a clinical study;
- The Predictive Clinical Model framework: this is intended to be followed by any proponent that will focus their pilot to create a predictive clinical model, using artificial intelligence, machine learning, and deep learning.

## **2. Clinical studies: the GRADE Evidence to Decision Framework**

Given the great diversity of the pilots in DARE, it is indeed unfeasible to identify unique indicators for each pilot, as the technologies operate at various levels of prevention and address a wide variety of pathologies. In order to identify the indicators reported below, we have inspired our work to the GRADE EtD criteria for health systems and public health [3]. This is a largely used instrument for both public health recommendations and policy-making

decisions, that will be used for the production of the recommendations in the context of DARE, where possible.

Further documentation on the GRADE EtD will follow. Still, in brief, this framework presents a list of criteria that are useful to inform about the opportunities and ways to implement an option (such as the effects and economic impact of an option, the priority of the problem, the impact of the option on equity and its acceptability and feasibility) and to making public health decisions.

Each domain of GRADE EtD is filled with a list of possible indicators to be gathered in the course of the development of the pilots as well as in the production of the systematic reviews, which will ultimately provide the required evidence in accordance with the GRADE EtD methodology [3]. This approach is feasible for pilots that operate in the clinical or public health domain, while it is not applicable for pilots related to predictive models based on safety big data. For such pilots, please refer to the next section about predictive models.

To select the best indicators, we also defer to the individual pilots to identify and suggest the most suitable ones, according to the technology in the exam and the medical field of application, always referring to the core domains [4].

The following is a list of indicators that are usually considered in order to fill the domains of GRADE EtD. Please note that all indicators should be addressed (at least with one key outcome), and an evidence synthesis should be conducted for each indicator to contextualize the results of the pilot within the existing knowledge.

## 2.1. List of indicators according to the two main GRADE EtD domains

General and utility domains:

- Seriousness of the problem;
- Priority of the problem;
- Clinical utility of the intervention;
- Impact of the intervention;
- Benefits and harms of the intervention;

- Balance between the desirable and undesirable effects.

Implementability domains:

- Equity and human rights;
- Financial and economic considerations;
- Values and preferences;
- Acceptability;
- Feasibility.

### **2.1.1. Detailed explanation of each domain**

**Seriousness of the problem:** the more serious a problem is, the more likely it is that an intervention that addresses the problem should be a priority. It refers to the disease or condition that we are analyzing in terms of Disability-Adjusted Life Years (DALYs) or mortality.

**Priority of the problem:** considering whether the problem being addressed in a health system and public health decision is a priority; in addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency, and consequences of the problem, and whether it is a recognised priority by people or stakeholders.

For example, from a population perspective, helpful interventions for serious conditions that are fatal or disabling are likely to be a higher priority and to be recommended than the ones for transient conditions or those that cause only minor and reversible distress. However, it is important that, while addressing the priority, we consider the seriousness, number of people, and frequency of the disease in a more complex analysis than just evaluating how serious it is in terms of mortality or disability caused by the problem.

**Clinical utility of the intervention:** The value of innovative technology in diagnosing/ruling out the disease, suggesting treatment or prevention strategies, and evaluating associated risks and benefits. For example, an innovative technology that provides early diagnosis of a disease but does not change its clinical history will have low utility.

**Impact of the intervention:** The absolute magnitude of the effect (such as the proportion of people who would benefit) and the importance of the outcome (such as how much it is valued by the people affected).

**Benefits and harms of the intervention:** How substantial are the desirable and undesirable anticipated effects? The more substantial the desirable effects, the more likely it is that an intervention should be recommended.

**Balance between the desirable and undesirable effects:** Does the balance between the desirable and undesirable effects favour the intervention or the comparison? For example, an intervention that reduces the pain caused by a disease but increases the risk of cancer is clearly balanced toward the undesirable effects.

**Equity and human rights:** The impacts on equity of the health system and public health options are important because these decisions are always taken from a population perspective. We can address potential impacts on equity by considering the possible differential effects of options on disadvantaged populations, including economic status, employment or occupation, education, place of residence, sex, and ethnicity.

**Financial and economic considerations:** a fundamental aspect while evaluating an innovative technology is to address its cost and the resources it requires. If resource use is considered critical for a recommendation, the greater the cost, the less likely an intervention will be recommended.

**Values and preferences:** specific values or preferences of the patients around the sharing of sensitive personal data or the use of technologies that might even only induce the suspicion of a risk of privacy violation are possible in a subset of patients.

**Acceptability:** the acceptability of an option may depend on evidence presented for some of the preceding criteria, such as the distribution and timing of harms, benefits and costs, and how much different stakeholders value the harms and benefits.

**Feasibility:** the ability of the intervention to be accomplished is a key factor to be addressed. Barriers to implementing an intervention can modify the strength of a recommendation; clinicians might find it unhelpful to receive strong recommendations if the interventions are

not implementable in their settings. However, policymakers could evaluate to make a strong recommendation and address the limitations of the proposed intervention.

### 2.1.2. Examples

In Tables 1-3 (Appendix A), we report three different examples of the use of each of the aforementioned criteria in the context of three recommendations produced for different digital technologies [6, 7, 8] in order to demonstrate the usefulness of the proposed indicators to formulate recommendations. For each criterion, a detailed explanation of the related indicator is reported, along with a practical example. Such examples should be considered to better understand how the process of pilot implementation and related data collection in each relevant criterion, is constructed.

If possible, according to the technical expertise, each pilot should conduct an assessment of the certainty of the evidence following the GRADE EtD methodology. However, in addition to the data collected from each pilot, a rapid review [9] or systematic review should be conducted to inform each criterion (please refer to the “Supplementary Materials Review Guide” as a short guidance on review conduction). This will enable the delegated Work Package to make evidence-based recommendations, using the GRADE EtD methodology if deemed appropriate, in a second phase of the DARE project.

## 3. Predictive models: the importance of defining indicators for proper performance assessment

An indicator for predictive models is a metric or set of metrics used to evaluate the performance of a model in predicting specific outcomes. As stated by Kuhn and Johnson, it is of paramount importance to define the proper indicators for assessing the performance of predictive models [10]. Such indicators are essential for ensuring that the model is effective in achieving the defined tasks and objectives. Indeed, without indicators, it is difficult to determine whether the model is working properly or whether improvements are needed.

The selection of appropriate indicators will depend on the type of predictive model and the application domain, and they should provide meaningful insights into the model's performance.

### 3.1. List of indicators for predictive models' performance assessment

Based on the type of problem, different metrics need to be utilized for evaluating the model.

General and predictive performance domains:

- Seriousness of the problem;
- Priority of the problem;
- Impact analysis;
- Discrimination of the model;
- Calibration of the model.

Implementability domains:

- Equity and human rights;
- Values and preferences;
- Feasibility.

#### 3.1.1. Detailed explanation of each domain

**Seriousness of the problem:** the more serious a problem is, the more likely it is that an intervention that addresses it should be a priority. It refers to the disease or condition that we are analyzing in terms of disability-adjusted life years (DALYs) or mortality.

**Priority of the problem:** considering whether the problem being addressed in a health system and public health decision is a priority; in addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency, and consequences of the problem, and whether it is a recognised priority by people or stakeholders.

For example, from a population perspective, helpful interventions for serious conditions that are fatal or disabling are likely to be a higher priority and to be recommended than the ones for transient conditions or those that cause only minor and reversible distress. However, it is important that, while addressing the priority, we consider the seriousness, number of people,

and frequency of the disease in a more complex analysis than just evaluating how serious, in terms of mortality or disability caused, the problem is.

**Impact analysis (Clinical utility):** The analysis of the impact is required to evaluate the clinical usefulness of a predictive model beyond its performance metric. While performance is an important aspect of a predictive model, clinical utility analysis takes into consideration the practical implications of using the model in a clinical setting to inform clinical decision-making. It considers the impact of the model on patient outcomes, the cost-effectiveness and economic sustainability of using the model, and the feasibility of implementing the model in clinical practice [11].

Common metrics for assessing clinical utility include Decision Curve Analysis (DCA), net benefit, and economic evaluations:

- Net benefit: it is a measure of the clinical utility of a predictive model that considers the harms and benefits of different decision thresholds. It is similar to DCA but does not rely on graphical representation;
- Decision curve analysis (DCA): it is a graphical method that assesses the clinical utility of a predictive model. DCA evaluates the net benefit of using a model or test compared to assuming that all individuals have the same risk or outcome. It considers the clinical consequences of true positive and false positive predictions, as well as the threshold probability or risk at which the model or test is used;
- Economic evaluations: Economic evaluations are a framework for assessing the economic impact of healthcare interventions, technologies, or policies. Economic evaluations aim to provide decision-makers with information on the costs and benefits of different interventions to make informed decisions about the allocation of resources.

There are different types of economic evaluations, but they typically involve comparing the costs and outcomes of different interventions using a standardized framework. The most commonly used frameworks for economic evaluations are cost-



effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA).

In a cost-effectiveness analysis, the costs of different interventions are compared to their effectiveness in achieving a specific health outcome, such as a reduction in disease incidence or mortality. The outcome of interest is typically measured in physical units, such as years of life saved or cases prevented. In a cost-utility analysis, the costs of different interventions are compared to their effectiveness in improving health-related quality of life. The outcomes are typically measured in quality-adjusted life years (QALYs), which take into account both the length and quality of life.

In a cost-benefit analysis, the costs of different interventions are compared to their monetary benefits, which are typically measured in terms of the monetary value of the outcomes achieved, such as increased productivity or avoided healthcare costs [12, 13, 14].

Overall, clinical utility analysis provides a comprehensive framework for evaluating the clinical usefulness of a predictive model and can help guide decision-making in the adoption and implementation of the model in clinical practice.

Discrimination of the model: The discrimination of a predictive model is a measure of its ability to distinguish between positive and negative outcomes.

These metrics evaluate how well the model ranks observations based on their predicted probabilities, and they are often used in binary classification problems where the goal is to identify which observations belong to a particular category.

Staying in the field of healthcare, discrimination problems involve predicting the presence or absence of a specific condition or disease based on various features or predictors. For instance, a classification problem in medicine is the use of medical records and diagnostic tests to predict whether a patient has diabetes, cancer, or heart disease. Common metrics for assessing the model's discrimination ability include the C-statistic and the D-statistic. Other more sophisticated metrics are available and are more appropriate for several contexts.

- C-statistic: the C-statistic, also known as the area under the receiver operating characteristic curve (AUC-ROC) when logistic regression is the underlying modeling strategy, is a measure of the model's ability to distinguish between positive and negative outcomes. It plots sensitivity (true positive rate) against (1 - specificity), i.e., false positive rate, for different decision thresholds of the model. It is used to evaluate the trade-off between sensitivity and specificity. Therefore, AUC is a summary measure of the ROC curve that indicates the overall performance of the model. A perfect model has an AUC of 1, while a random model has an AUC of 0.5 [15, 16];
- D-statistic: the D-statistic, also known as the Somers' D, is a rank correlation coefficient that measures the strength of the association between the predicted probabilities and the observed outcomes [17].

Calibration: a models' calibration is the degree to which its predicted probability matches the actual outcomes in the data.

In healthcare, calibration is important because it ensures that the predicted probabilities of a model accurately reflect the true probability of an outcome in the population being studied. In other words, a well-calibrated model will predict the correct probability of an event occurring in a group of patients, on average. For instance, if a model predicts that 70% of patients are at high risk for a certain disease, then 70% of the patients in the actual population should actually be at high risk for that disease.

The Van Calster hierarchy was chosen as one of the most suitable frameworks for assessing calibration in predictive models. The Van Calster hierarchy (triad) provides a systematic approach to assessing the level of calibration assessment, consisting of different degrees of calibration [18, 19]:

- Average calibration (calibration-in-the-large): It refers to a situation where a predictive model shows good overall calibration, with no significant deviations between the predicted probabilities and the true risk probabilities. In other words, the model predicts the risk probabilities accurately on average, but there may still be some variability in the predictions for individual patients. Average calibration suggests that

the model is performing well overall but may still benefit from some fine-tuning to improve its performance for specific subgroups of patients or outcomes;

- Weak calibration (calibration intercept and slope): it refers to a situation where a predictive model shows some calibration, but the predicted probabilities are systematically too extreme or too conservative. In other words, the model may tend to overestimate or underestimate the true risk probabilities. Weak calibration is the second level of calibration assessment in the Van Calster hierarchy and suggests that the model needs to be improved to achieve better agreement between the predicted probabilities and the true risk probabilities;
- Moderate calibration: it refers to a situation where a predictive model shows reasonable calibration, but there are still some systematic deviations between the predicted probabilities and the true risk probabilities. In other words, the model may have a tendency to overestimate or underestimate the true risk probabilities to some extent. Moderate calibration is the intermediate level of calibration assessment in the Van Calster hierarchy and suggests that the model can be further improved to achieve better agreement between the predicted probabilities and the true risk probabilities.

The Van Calster hierarchy emphasizes the importance of assessing model calibration at multiple levels to fully evaluate the performance of a predictive model.

**Equity and human rights:** The impacts on equity of the health system and public health options are important because these decisions are always taken from a population perspective. We can address potential impacts on equity by considering the possible differential effects of options on disadvantaged populations, including economic status, employment or occupation, education, place of residence, sex, and ethnicity.

**Values and preferences:** specific values or preferences of the patients around the sharing of sensitive personal data or the use of technologies that might even only induce the suspicion of a risk of privacy violation are possible in a subset of patients.

**Feasibility:** the ability of the intervention to be accomplished is a key factor to address. Barriers to implementing an intervention can modify the strength of a recommendation;

clinicians might find it unhelpful to receive strong recommendations if the interventions are not implementable in their settings; However, policymakers could evaluate to raise a strong recommendation and address the limitations of the proposed intervention.

### 3.1.2. Examples

In Table 4 (Appendix B), we report an example of the use of each of the aforementioned criteria in the context of a predictive clinical model.

This example should be considered in order to better understand how the process of pilot implementation and related data collection in each relevant criterion is constructed. Eventually, for each criterion, a detailed explanation of the related indicator is reported, along with the practical example.

## 4. References

- [1] <https://www.cdc.gov/std/Program/pupestd/Developing%20Evaluation%20Indicators.pdf>
- [2] <https://mypeer.org.au/monitoring-evaluation/indicators-for-evaluation/>
- [3] Moberg J, Oxman AD, Rosenbaum S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst.* 2018;16(1):45. Published 2018 May 29. doi:10.1186/s12961-018-0320-2
- [4] Rejon-Parrilla, J. C., Espin, J., & Epstein, D. (2022). How innovation can be defined, evaluated and rewarded in health technology assessment. *Health economics review*, 12(1), 1. <https://doi.org/10.1186/s13561-021-00342-y>
- [5] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol.* 2011 Apr;64(4):395-400. doi: 10.1016/j.jclinepi.2010.0
- [6] [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=87234&VersionID=1136179](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=87234&VersionID=1136179)
- [7] [https://www.crd.york.ac.uk/prospero/display\\_record.php?RecordID=87234&VersionID=1136179](https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=87234&VersionID=1136179)
- [8] <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD012928/full>
- [9] <https://methods.cochrane.org/rapidreviews/cochrane-rr-methods>
- [10] Applied Predictive Modeling. <https://doi.org/10.1007/978-1-4614-6849-3>
- [11] Drummond MF, Sculpher MJ, Claxton K, et al. *Methods for the Economic Evaluation of Health Care Programmes.* Oxford: Oxford University Press; 2015.
- [12] Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 1977;296(13):716-721.
- [13] Gold MR, Siegel JE, Russell LB, et al. *Cost-Effectiveness in Health and Medicine.* Oxford: Oxford University Press; 1996.
- [14] Neumann PJ, Sanders GD, Russell LB, et al. Cost-effectiveness in health and medicine. In: Culyer AJ, Newhouse JP, eds. *Handbook of Health Economics.* Vol. 1. Amsterdam: Elsevier; 2000: 463-536.
- [15] Fawcett, Tom (2006). "An Introduction to ROC Analysis". *Pattern Recognition Letters.* 27 (8): 861-874.  HYPERLINK

"[https://en.wikipedia.org/wiki/Doi\\_\(identifier\)](https://en.wikipedia.org/wiki/Doi_(identifier))"doi::10.1016/j.patrec.2005.10.010.10.1016/j.patrec.2005.10.010.

[16] Powers, David M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". *Journal of Machine Learning Technologies*. 2 (1): 37-63.. *Journal of Machine Learning Technologies*. 2 (1): 37-63.

[17] Somers RH. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 1962; 27(6): 799-811.

[18] Steyerberg, E. W., et al. (2010). "Internal validation of predictive models: efficiency of some procedures for logistic regression analysis." *Journal of clinical epidemiology* 54(8): 774-781.

[19] Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M. J., & Steyerberg, E. W. (2016). A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of clinical epidemiology*, 74, 167-176. <https://doi.org/10.1016/j.jclinepi.2015.12.005>

[20] Gerd Assmann, and Paul Cullen. Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-Year Follow-Up of the Prospective Cardiovascular Münster (PROCAM) Study Helmut Schulte. <https://doi.org/10.1161/hc0302.102575>

[21] Bedogni, G.; Tamini, S.; Caroli, D.; Cicolini, S.; Domenicali, M.; Sartorio, A. Development and Internal Validation of Fatty Liver Prediction Models in Obese Children and Adolescents. *J. Clin. Med.* 2021, 10, 1470. <https://doi.org/10.3390/jcm10071470>



## Appendix A

**Table 1.** Example of application of the indicators, according to domains and criteria, in the context of primary health care (s) that populated a GRADE EtD recommendations

DOMAIN	DEFINITION	INDICATORS	“Digital, targeted client communication for a standard care in primary healthcare settings”	
			Outcomes	Standard care
SERIOUSNESS OF THE PROBLEM	The more serious a problem is, the more likely it is that an intervention that addresses the problem	DALYs, mortality, morbidity, hospitalization, complications	<u>Intervention:</u> Targeted communication to H information/education via SMS, MMS, IVR, or voice calls)	
			Mortality of HIV in 2021	650.000
			Incidence of HIV in 2021	1.5 million
			Prevalence of HIV in 2019	36.8 million



	should be a priority. It refers to the disease or condition that we are analyzing.		DALYs of HIV in 2019	47.6 million		
PRIORITY OF THE PROBLEM	In addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency and consequences of the problem, and whether it is a recognized priority by people or stakeholders.	Number of people affected, severity, urgency, consequences, unmet need, frequency of hospital admissions, length of the hospital stay	Number of people affected by HIV in 2021	38.4 million		
CLINICAL UTILITY OF THE INTERVENTION	The value of innovative technology in	Therapeutic benefit, adherence to the	HIV viral load suppression (< 400 copies per mL) Follow-up: up to 12 months	680 per 1.000	727 per 1.000 (605 to 884)	RR 1.07 (0.89 to 1.30)



	diagnosing/ruling out a disease.	therapy, clinic attendance, hospitalization and length	Adherence to anti-retroviral medication (objective and self-report) Follow-up: up to 12 months	489 per 1.000	538 per 1.000 (475 to 612)	RR 1.10 (0.97 to 1.25)
			Clinic attendance for STI/HIV testing (objective) Follow-up: 2 to 12 weeks	278 per 1.000	545 per 1.000 (287 to 1.000)	RR 1.96 (1.03 to 3.75)
			Health behavior - use of effective contraception method (self-report) Follow-up: 12 months	428 per 1.000	500 per 1.000 (393 to 633)	RR 1.17 (0.92 to 1.48)
			Health behavior - Condom use 50% of the time (self-report) Follow-up: up to 12 months	243 per 1.000	472 per 1.000 (243 to 919)	RR 1.94 (1.00 to 3.78)
IMPACT OF THE INTERVENTION	The absolute magnitude of the effect (such as the proportion	Health behavior, status and well-being, life	Well-being among people living with HIV and AIDS Follow-up: up to 6 months	This outcome can be measured by SF-12 or WHO QoL physical wellbeing subscale.		



	of people who would benefit) and the importance of the outcome (such as how much it is valued by the people affected).	expectancy, health-related quality of life				
BENEFIT AND HARMS OF THE INTERVENTION	How substantial are the desirable and undesirable anticipated effects?	Patients' experience, unintended consequences	Women's experience of physical violence, in the context of receiving targeted communication on contraception (self report) Follow-up: 4 months	65 per 1.000	109 per 1.000 (68 to 170)	OR 1.74 (1.04 to 2.92)
BALANCE BETWEEN THE DESIRABLE AND UNDESIRABLE EFFECTS	Does the balance between the desirable and undesirable effects favor the intervention or the comparison?	Safety	Other unintended consequences	Three studies reported on unintended consequences as a result of the intervention. One study explicitly reported no adverse events, while the other reported that one female in the intervention arm requested		/



				to withdraw because she felt it had compromised her undisclosed status. One study reported that at four months follow-up, no participants experienced involvement in a road traffic accident or domestic abuse as a result of the intervention or control.	
EQUITY AND HUMAN RIGHTS	It explores the equity of the interventions, from a population perspective.	Economic status, employment or occupation, education, place of residence, sex, and ethnicity	To fill this domain, it is necessary to consider the possible differential effects of intervention on disadvantaged populations.		
FINANCIAL AND ECONOMIC CONSIDERATIONS	Evaluating an innovative technology addressing its cost and the resources it requires.	Resource use, cost-effectiveness	Resource use	One study reported a cost of about \$2.41 for each additional person to be HIV tested, that is, the cost to get people to test over and above those who were likely to test without the intervention.	/

VALUES AND PREFERENCES	About the sharing of sensitive personal data or the use of technologies.	Patients' experience about the risk of privacy violation	To fill this domain, it is necessary to consider patients' thoughts about the risk of privacy violation.		
ACCEPTABILITY	It explores if the intervention is acceptable for the people and the stakeholders.	Patient usefulness, caregiver usefulness, administration, satisfaction of the users	Client acceptance of and satisfaction with the approach/ intervention	In general, all studies reported moderate to high levels of satisfaction and acceptability with the intervention. (Only intervention group assessed for this outcome thus non-comparable)	/
FEASIBILITY	The ability of the intervention of being accomplished.	Interoperability, infrastructure availability, bottlenecks, barriers	To fill this domain, it is necessary to analyze the capacity of the intervention to be able to implemented.		

**Table 2.** Example of application of the indicators, according to domains and criteria, in the context of secondary prevention with digital intervention (s) that populated a GRADE EtD recommendations.



CRITERIA	DEFINITION	INDICATORS	“Digital, targeted client communication for adult users of SRH services, compared to non-digital, targeted communication in primary healthcare settings” [7] <u>Intervention:</u> Targeted communication to breast cancer patients (reminders and/or information/education via SMS, MMS, IVR, instant messaging, app instant messaging or voice calls)			
			Outcomes	Standard care	Targeted client communication	Relative effect (95% CI)
SERIOUSNESS OF THE PROBLEM	The more serious a problem is, the more likely it is that an intervention that addresses the problem should be a priority. It refers to the disease or condition that we are analyzing.	DALYs, mortality, morbidity, hospitalization, complications	Mortality of breast cancer in 2023	43.700 women		
			Incidence of invasive breast cancer	297.790 women		
			Incidence of ductal carcinoma in situ (DCIS)	55.720 women		
			DALYs of breast cancer in 2019	20.6 million		



<p>PRIORITY OF THE PROBLEM</p>	<p>In addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency and consequences of the problem, and whether it is a recognized priority by people or stakeholders.</p>	<p>Number of people affected, severity, urgency, consequences, unmet need, frequency of hospital admissions, length of the hospital stay</p>	<p>Breast cancer survivors  3.8 million</p>			
<p>CLINICAL UTILITY OF THE INTERVENTION</p>	<p>The value of innovative technology in diagnosing/ruling out a disease.</p>	<p>Therapeutic benefit, adherence to the therapy, clinic attendance,</p>	<p>Clinic attendance for breast cancer screening (self-report)  Follow-up: 6 months</p>	<p>250 per 1.000</p>	<p>400 per 1.000 (235 to 685)</p>	<p>RR 1.60 (0.94 to 2.74)</p>



		hospitalization and length	Clinic attendance for cervical screening (objective report) Follow-up: 8 weeks	188 per 1.000	216 per 1.000 (152 to 306)	RR 1.15 (0.81 to 1.63)
IMPACT OF THE INTERVENTION	The absolute magnitude of the effect (such as the proportion of people who would benefit) and the importance of the outcome (such as how much it is valued by the people affected).	Health behavior, status and well-being, life expectancy, health-related quality of life	Health behavior, status and well-being	This outcome can be measured by SF-12 or WHO QoL physical wellbeing subscale or other scales.		
BENEFIT AND HARMS OF THE INTERVENTION	How substantial are the desirable and undesirable anticipated effects?	Patients' experience, unintended consequences	To fill this domain, is it necessary to considerer patients' experience about intervention and unintended consequences.			
BALANCE BETWEEN THE	Does the balance between the desirable	Safety	To fill this domain, is it necessary to take a balance between desirable and undesirable effects.			



DESIRABLE AND UNDESIRABLE EFFECTS	and undesirable effects favor the intervention or the comparison?				
EQUITY AND HUMAN RIGHTS	It explores the equity of the interventions, from a population perspective.	Economic status, employment or occupation, education, place of residence, sex, and ethnicity	To fill this domain, it is necessary to consider the possible differential effects of intervention on disadvantaged populations.		
FINANCIAL AND ECONOMIC CONSIDERATIO NS	Evaluating an innovative technology addressing its cost and the resources it requires.	Resource use, cost- effectiveness	Resource use	One study reported that the total cost of a screening program, using SMS reminders, was cheaper than phone calls or normal letters.	/
VALUES AND PREFERENCES	About the sharing of sensitive personal data	Patients' experience about the risk of privacy violation	To fill this domain, it is necessary to consider patients' thoughts about the risk of privacy violation.		



	or the use of technologies.				
ACCEPTABILITY	It explores if the intervention is acceptable for the people and the stakeholders.	Patient usefulness, caregiver usefulness, administration, satisfaction of the users	Client acceptance of and satisfaction with the approach/ intervention	Clients reported high levels of satisfaction with the intervention. (Only intervention group assessed for this outcome thus non-comparable)	/
			Clients' and providers' acceptability/satisfaction	No studies were identified that reported this outcome	/
FEASIBILITY	The ability of the intervention of being accomplished.	Interoperability, infrastructure availability, bottlenecks, barriers	To fill this domain, it is necessary to analyze the capacity of the intervention to be able to implemented.		

**Table 3.** Example of application of the indicators, according to domains and criteria, in the context of tertiary prevention with digital intervention (s) that populated a GRADE EtD recommendations.

CRITERIA	DEFINITION	INDICATORS	“Mobile- based technologies to support client to healthcare provider communication and management of care (unpublished review)” [8] <u>Intervention:</u> Telemedicine to patients with cardiovascular disease (CVD) (phone-based consultations, home-based monitoring followed by phone-based consultations, web-delivered physical rehabilitation, smartphone application for contacting provider)			
			Outcomes	Standard care	Targeted client communication	Relative effect (95% CI)
SERIOUSNESS OF THE PROBLEM	The more serious a problem is, the more likely it is that an intervention that addresses the problem should be a priority. It refers to the disease or	DALYs, mortality, morbidity, hospitalization, complications	Mortality of CVD in 2019	2.0 million		
			DALYs of CVD in 2019	40.8 million		



	condition that we are analyzing.					
PRIORITY OF THE PROBLEM	In addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency and consequences of the problem, and whether it is a recognized priority by people or stakeholders.	Number of people affected, severity, urgency, consequences, unmet need, frequency of hospital admissions, length of the hospital stay	Number of people affected of CVD in 2017	126 million		
CLINICAL UTILITY OF THE	The value of innovative technology in diagnosing/ruling out a disease.	Therapeutic benefit, adherence to the therapy, clinic attendance,	Number of hospital admission Follow-up: 2 to 12 months	328 per 1,000	322 per 1,000 (220 to 361)	RR 0.98 (0.67 to 1.10)



INTERVENTI ON		hospitalization and length	Length of hospital stay Follow-up: 8 to 12 months	Two studies recruiting clients with heart conditions show both negative and positive impacts. The first study shows an increase of 2 days (95% CI -0.18, 4.18) in length of hospital stay. The second study shows a decrease of 2.5 days (95% CI -4.64, - 0.36) in length of hospital stay		/
			Number of hospital and clinic visits (number of clients visiting) Follow-up: 1-12 months	90 per 1,000	34 per 1,000 (11 to 108)	RR 0.38 (0.12, 1.20)
IMPACT OF THE INTERVENTI ON	The absolute magnitude of the effect (such as the proportion of people who would benefit)	Health behavior, status and well- being, life expectancy, health- related quality of life	Mortality among individuals with heart conditions Follow-up: 6-12 months	48 per 1,000	27 per 1,000 (18 to 42)	RR 0.57 (0.38 to 0.87)



	and the importance of the outcome (such as how much it is valued by the people affected).		Health-related quality of life Follow-up: 1-6 months	This outcome can be measured by ACT, SF-36 or other scales.	
BENEFIT AND HARMS OF THE INTERVENTIONS	How substantial are the desirable and undesirable anticipated effects?	Patients' experience, unintended consequences	Unintended consequences - adverse clinical events	One study that recruited adults who sought help for smoking cessation reported more adverse events in the intervention group (RR 1.52, 95% CI 1.30 to 1.77). Two studies reported little or no difference between groups.	/
			Unintended consequences related to the intervention	One study recruiting women who had an induced abortion reported that there were little or no differences between groups for adverse effects of the intervention (specifically, car accidents caused by driving while using a mobile phone to access support).	/
BALANCE BETWEEN	Does the balance between the desirable	Safety	To fill this domain, is it necessary to take a balance between desirable and undesirable effects.		



THE DESIRABLE AND UNDESIRABLE EFFECTS	and undesirable effects favor the intervention or the comparison?				
EQUITY AND HUMAN RIGHTS	It explores the equity of the interventions, from a population perspective.	Economic status, employment or occupation, education, place of residence, sex, and ethnicity	To fill this domain, it is necessary to consider the possible differential effects of intervention on disadvantaged populations.		
FINANCIAL AND ECONOMIC CONSIDERATIONS	Evaluating an innovative technology addressing its cost and the resources it requires.	Resource use, cost- effectiveness	Healthcare costs Follow-up: 10-18 months	One study with adults with heart failure found that mean total all-causes costs per client during 18 months follow-up was USD 4678 less, (95% CI = -4758 to -4597) in the intervention group than the control group. Another study with children with complex congenital heart disease reported that delivering the intervention by videoconferencing	/



				was less costly than usual care. One study that measured costs to the health system and parents, evaluated an e-health portal for children with a skin condition and reported little or no difference between groups; one study measured costs to the health system of a PhysioDirect telephone intervention and reported little of no difference.	
VALUES AND PREFERENCES	About the sharing of sensitive personal data or the use of technologies.	Patients' experience about the risk of privacy violation	To fill this domain, it is necessary to consider patients' thoughts about the risk of privacy violation.		
ACCEPTABILITY	It explores if the intervention is acceptable for the people and the stakeholders.	Patient usefulness, caregiver usefulness, administration, satisfaction of the users	Client acceptability/satisfaction with the intervention	One study measured satisfaction with the intervention among individuals in the intervention group and found that: 42/49 were satisfied with service, 34/46 thought it gave added value and 22/48 preferred face to face. In another study the following was measured among individuals in the intervention group: Intervention was approachable:	/



				90% agreed/strongly agreed; Intervention improved diabetes knowledge: 90% agreed/strongly agreed; Would prefer to see a provider in person: 50% agreed/strongly agreed		
			Number of clients satisfied with care	292 per 1,000	535 per 1,000 (403 to 710)	RR 1.83 (1.38, 2.43)
FEASIBILITY	The ability of the intervention of being accomplished.	Interoperability, infrastructure availability, bottlenecks, barriers	To fill this domain, it is necessary to analyze the capacity of the intervention to be able to implemented.			

## Appendix B

**Table 4.** Example of application of the indicators, according to domains and criteria, in the context of a predictive clinical model pilot

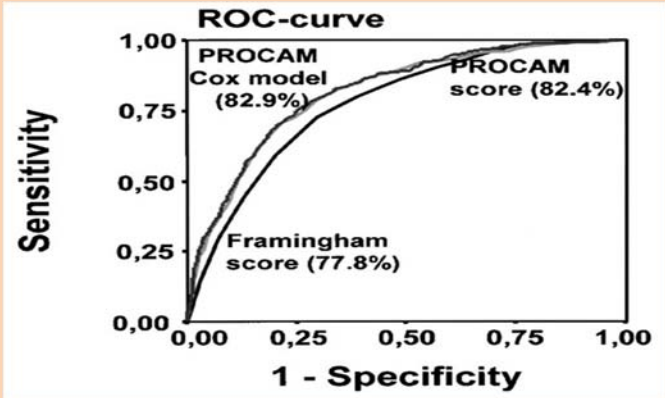
DOMAIN	DEFINITION	INDICATORS EXAMPLES	“Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-Year Follow-Up of the Prospective Cardiovascular Münster (PROCAM) Study” [20] [21]	
			Indicator	Exploitation
SERIOUSNESS OF THE PROBLEM	The more serious a problem is, the more likely it is that an intervention that	DALYs, mortality, morbidity, hospitalization, complications	Number of deaths for Ischemic Heart Disease in 2021	9.440.000 (8.820.000-9.960.000)
			<p>Intervention:</p> <ol style="list-style-type: none"> <li>1. Targeted communication to HIV-infected patients (reminders and/or information/education via SMS, MMS, IVR, instant messaging, app instant messaging or voice calls)</li> <li>2. Development and Internal Validation of Fatty Liver Prediction Models in Obese Children and Adolescents</li> </ol>	



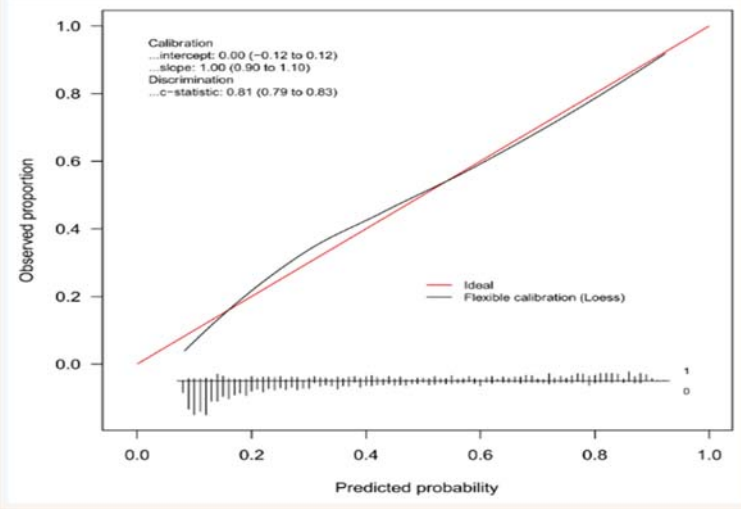
	addresses the problem should be a priority. It refers to the disease or condition that we are analyzing.		Number of DALYs for Ischemic Heart Disease in 2021	185.000.000 (175.000.000-196.000.000)
PRIORITY OF THE PROBLEM	In addition to seriousness, it is important to evaluate the number of people affected, the severity, urgency and consequences of the problem, and whether it is a recognized priority by people or stakeholders.	Number of people affected, severity, urgency, consequences, unmet need, frequency of hospital admissions, length of the hospital stay	Number of person affected by some CVD in 2021	523 million people had some form of CVD

<p>IMPACT ANALYSIS (CLINICAL UTILITY)</p>	<p>It considers the impact of the model on patient outcomes, the cost-effectiveness and economic sustainability of using the model, and the feasibility of implementing the model in clinical practice.</p>	<p>Decision Curve Analysis, Net Benefit Analysis, economic evaluations (Markov models)</p>	<p>Example of a Decision Curve Analysis</p>	<p>The graph shows Net Benefit on the y-axis (0.00 to 0.10) and Threshold Probability on the x-axis (0% to 30%). Four lines are plotted: 'Treat All' (red), 'Treat None' (green), 'Family History' (cyan), and 'Prediction Model' (purple). The Prediction Model line is consistently above the other three, indicating superior performance across most threshold probabilities.</p>
<p>DISCRIMINATION ANALYSIS</p>	<p>It involves predicting the presence or absence of a specific condition or disease based on various features or risk factors.</p>	<p>C-statistics (ROC, AUC/ROC)</p>	<p>Example of AUC/ROC</p>	<p>“We compared the performance of our scoring system with that of the Cox model in calculating the relative risk of an acute coronary event using receiver-operating characteristics (ROC) curve analysis. Such ROC curves measure the discrimination of a prediction model, which is its ability to separate those who have had CHD events from those who do not. Although the area under the ROC curve obtained by means of the Cox function was 82.9%, the area under the ROC curve obtained with the PROCAM scoring scheme did not differ significantly, at 82.4% (P=0.251), indicating that the ability of the PROCAM scoring scheme to predict the</p>



				<p>relative risk of an acute coronary event was equivalent to that of the full PROCAM Cox algorithm.”</p> 
<p>CALIBRATION ANALYSIS</p>	<p>It is the degree to which the predicted probability of a model corresponds to the actual outcomes observed in the data. In other words, a well-calibrated model will predict the correct probability of an event</p>	<p>Weak calibration (calibration intercept and slope); moderate calibration; average calibration</p>	<p>Example of an Average Calibration</p>	<p>“The average expected rate of fatty liver (38%, 95%CI 36% to 40%) equaled the average observed rate (38%, 95%CI 36% to 41%), showing a satisfactory mean calibration. At logistic calibration, the average calibration slope was 1 and the average intercept was 0, showing a satisfactory weak calibration. Lastly, the examination of calibration plots showed an acceptable profile of moderate calibration” [21]</p>



	occurring in a group of patients, on average			<p><b>Figure 1. Internal calibration plot for the diagnosis of fatty liver from the BMI model.</b> Abbreviation: loess = locally estimated scatterplot smoother.</p> 
EQUITY AND HUMAN RIGHTS	It explores the equity of the interventions, from a population perspective.	Economic status, employment or occupation, education, place of residence, sex, and ethnicity	To fill this domain, it is necessary to consider the possible differential effects of intervention on disadvantaged populations.	
VALUES AND PREFERENCES	About the sharing of sensitive personal data		To fill this domain, it is necessary to consider patients' thoughts about the risk of privacy violation.	



	or the use of technologies.		
FEASIBILITY	The ability of the intervention of being accomplished.	Interoperability, infrastructure availability, bottlenecks, barriers	To fill this domain, it is necessary to analyze the capacity of the intervention to be able to implemented.



UNIVERSITÀ  
CATTOLICA  
del Sacro Cuore

# HOW TO PERFORM A LITERATURE REVIEW

Short guide for Systematic and Rapid Reviews

## Summary

<b>INTRODUCTION</b> .....	<b>2</b>
How to formulate a research question .....	2
<b>GUIDELINES FOR SYSTEMATIC REVIEWS</b> .....	<b>3</b>
What is a Systematic Review and what is its purpose? .....	3
How to perform a systematic review? .....	3
1. Formulate the search question .....	3
2. Refine the search by applying pre-determined inclusion and exclusion criteria .....	6
3. Extract the appropriate data and assess their quality and validity .....	6
4. Synthesize, interpret, and report data .....	7
<b>GUIDELINES FOR RAPID REVIEWS</b> .....	<b>8</b>
What is a Rapid Review and what is its purpose? .....	8
How to perform a rapid review? .....	8
1. Define/refine the topic question, identify biases, plan and execute search .....	8
2. Screen, select and critically appraise the results .....	9
3. Synthesize the evidence .....	9
4. Identifying applicability and transferability issues for further consideration during the decision-making process .....	9
5. Reporting .....	9
<b>Bibliography</b> .....	<b>11</b>
<b>Appendix 1. Most popular risk of bias evaluation tools</b> .....	<b>14</b>
<b>Appendix 2. PRISMA 2020 Checklist</b> .....	<b>15</b>

# INTRODUCTION

## How to formulate a research question

Formulating the correct research question is pivotal in the development of a pilot. In this content it is worth to mention the EPICOT+ framework of Brown and all (1), which is a tool to assess and resolve the uncertainties of the research questions of any health intervention.

The EPICOT+ framework, that will be deeply discussed in this document below, is based on core elements and optional elements. The most important aspect of this framework is the domain E, Evidence, that explore the current state of the art of the literature on the topic. Every research question, indeed, needs to be assessed with a preliminary systematic review of the literature. In fact, the investigators need to exclude that there is already robust and final evidence in favor of one of the proposed interventions in comparison.

As more recently pointed out by the EU funded project EVBRES (Evidence Based REsearch) working group (2), at the moment, a number of studies starts without responding to any actual need of research gaps, but rather on researchers' preferences and strategic considerations. While confirmatory studies are certainly important, redundant research is a practice that should be avoided when high quality and robust evidences are already available in the literature.

As such, it is fundamental to base the development of DARE pilots on the "Evidence-Based Research" (EBR) framework, defined by EVBRES as "the use of prior research in a systematic and transparent way to inform a new study so that the research is answering questions that matter in a valid, efficient and accessible manner".

In order to navigate the uncertainty of how to formulate a good research question and to provide guidance for the researcher to develop an interesting and complete one, the FINER framework (3) comes to aid. According to FINER, the research question should be Feasible, Interesting, Novel, Ethical and Relevant.

<b>Feasible</b>	Research questions should be answered under objective aspects like time, scope, resources, expertise, or funding.
<b>Interesting</b>	Regardless of your own personal motivation about a subject, it is important to check if your question corresponds to more practical and broader interests.
<b>Novel</b>	Answer to an existing gap in knowledge. Filling one of these gaps is important.
<b>Ethical</b>	In empirical research, ethics is an absolute MUST.
<b>Relevant</b>	Relevance can lead to real, visible changes in society.

# GUIDELINES FOR SYSTEMATIC REVIEWS

## What is a Systematic Review and what is its purpose?

A systematic review attempts to identify and synthesize the available empirical evidences that meet pre-specified eligibility criteria in order to answer a specific research question and to provide an overview of a particular topic.

The key characteristics are:

- a clearly stated set of objectives with pre-defined eligibility criteria;
- an explicit and systematic methodology;
- a systematic search that attempts to identify all studies related to the search question;
- an assessment of the validity and reliability of the findings;
- a systematic presentation and synthesis of the characteristics and findings of the included studies.

Meta-analysis is the use of statistical methods to summarize the results of independent studies.

## How to perform a systematic review?

1. Formulate the search question
2. Refine the search by applying pre-determined inclusion and exclusion criteria
3. Extract the appropriate data and assess their quality and validity
4. Synthesize, interpret, and report data

### 1 *Formulate the search question*

#### **Search strategy**

The first step in a systematic review is to determine its focus. It is therefore fundamental to clearly frame the question(s) the review seeks to answer. The research question should be clear and focused - not too vague, too specific or too broad. The detailed specification of the review question(s) requires consideration of several key components that could guide many aspects of the review process. The 'PICO' technique ([Figure 1](#)) may help researchers with this process.

Figure 1. The PICO model

P	I	C	O
<b>Patient, Population</b>	<b>Intervention (or Exposure)</b>	<b>Comparison (or Control)</b> if appropriate	<b>Outcome</b>
Most important characteristics of patient (e.g. age, disease/condition, gender)	Main intervention (e.g. drug treatment, diagnostic/screening test)	Main alternative (e.g. placebo, standard therapy, no treatment, gold standard)	What you are trying to accomplish, measure, improve, affect (e.g. reduced mortality or morbidity, improved memory)

A variant of PICO is **PICOT+**, with the same first 4 core elements, but completed with other core and optional elements

### Core elements

- **P: Population** (What is the population of interest?). We must consider various characteristics such as Diagnosis, disease stage, comorbidity, risk factor, sex, age, ethnic group, specific inclusion or exclusion criteria, clinical setting.
- **I: Intervention** (What are the interventions of interest?). We must consider type, frequency, dose, duration, prognostic factor.
- **C: Comparison** (What are the comparisons of interest?). For example: Placebo, routine care, alternative treatment or management.
- **O: Outcome** (What are the outcomes of interest?). Which clinical or patient related outcomes will the researcher need to measure, improve, influence, or accomplish? Which methods of measurement should be used?
- **T: Time stamp** (Date of recommendation). Date of the last literature search or recommendation in the field in exam.

### Optional elements

- **d: Disease burden** or relevance. A research recommendation will rarely have an absolute value in itself. Its relative priority will be influenced by the burden of ill health, which is itself dependent on factors such as local prevalence, disease severity, relevant risk factors, and the priorities of the organization considering commissioning the research.
- **t: Time aspect** of core elements of EPICOT. Relevant to each of the core elements in varying ways—for example, duration of treatment, length of follow-up.
- **s: Appropriate study type** according to local need. What is the most appropriate study design to address the proposed question?

So, as we can see, we have various domains that must be filled. The optional elements are not necessary to make a full assessment, but they

should be filled.

Equal emphasis in addressing and equal precision in defining are not necessary for each PICO component. For example, a review might concentrate on competing interventions for a particular stage of breast cancer, with stage and severity of the disease being defined very precisely, or alternately focus on a particular drug for any stage of breast cancer, with the treatment formulation being defined very precisely.

### ***Protocol Development***

The protocol sets out the context in which the review is being conducted. It represents an opportunity to develop ideas that are foundational for the review or the definition of the eligibility criteria such as the study participants and the choice of comparators and outcomes.

### ***Databases selection***

Select the databases from which retrieving the records. The databases can be generic (*PubMed, Scopus, Web of Science, Embase, The Cochrane Library, etc.*) or related to a specific field [e.g., *CANCERLIT (oncology)*].

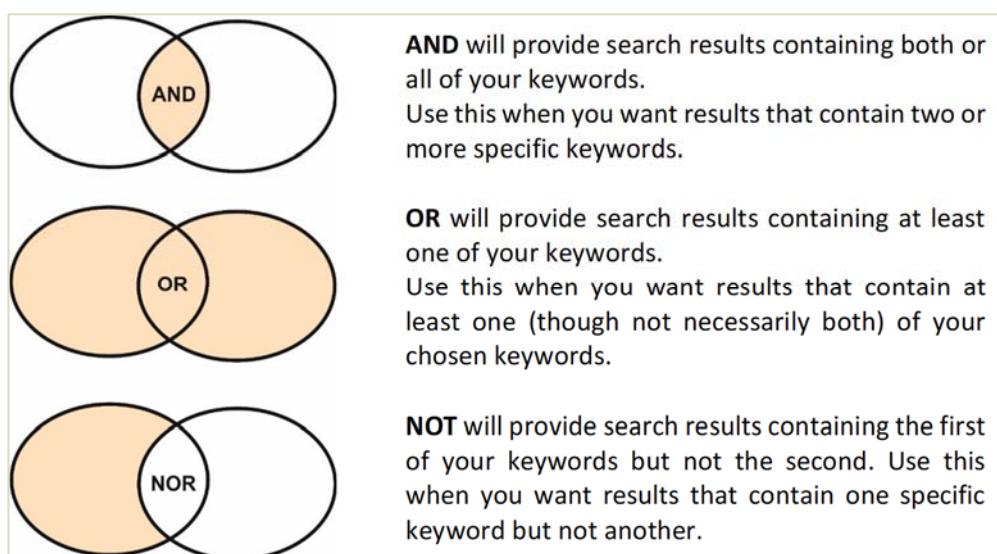
### ***Search filters***

Search filters are search strategies that are designed to retrieve specific types of records, such as those of a particular methodological design. For example, when searching for randomized trials in humans, a validated filter should be used to identify studies with the appropriate design.

### ***Keywords***

The *keywords*, which represent the major concepts of the search topic, are terms that, combined with each other through Boolean operators, are used in the databases to answer the search question. Boolean operators (Figure 2) are: “OR”, “AND”, “NOT”. They let expand or narrow search parameters when using a database or search engine. Indeed, using these operators can greatly reduce or expand the number of records returned. Boolean operators are useful in saving time by focusing searches for more 'on-target' results (that are more appropriate to search needs), eliminating unsuitable or inappropriate ones. Each search engine or database collection uses Boolean operators in a slightly different way or may require the operator be typed in capitals or have special punctuation.

Figure 2. Boolean operators conjunctions



## 2. Refine the search by applying pre-determined inclusion and exclusion criteria

Researchers should pre-specify the criteria for including and excluding studies. Inclusion criteria are the factors that would make a study eligible to be included, exclusion criteria those that would make a study ineligible. Criteria that should be considered include type of studies, type of participants, type of intervention, type of outcome measures.

Authors need to consider a priori what study designs are likely to provide reliable data to reach the aims of their review. The PICO elements present in the search question usually translate directly into eligibility criteria for the review, though this is not always a straightforward process and requires a thoughtful approach.

At the initial screening stage, just the title and abstract of the candidate studies have to be read and a decision to include or exclude each study has to be made.

At the next step, the full text of each study selected at the first screening stage has to be read and critically appraised to determine whether to include it in the review or not.

## 3. Extract the appropriate data and assess their quality and validity

Once all studies to be included in the review have been selected, all relevant data must be extracted. The list of data to be extracted should be agreed a priori during the study design phase. A specific data extraction form should be designed for the review so that the same data is extracted from each study and missing data is clearly evident. Furthermore, the data extraction has to be accurate and reproducible and should be performed by at least two independent readers.

As for quality assessment, more than 35 different quality assessment

instruments (most popular reported in the [Appendix1](#)) have been published in the literature, and most are designed for randomized clinical trials. The Jadad and the Consort scales are two examples of quality assessment tools for randomized clinical trials.

#### 4. Synthesize, interpret, and report data

Once the data have been extracted and their quality assessed, the results of individual studies may be pooled in a summary table ([Table 1](#)). If there are the required conditions to perform a meta-analysis, the heterogeneity among the gathered data has to be assessed, preferentially using both the Chi-squared test and the I-squared statistics. This will allow authors to choose either fixed- or random-effects models. Mostly, the data will be heterogeneous, therefore the random-effects model will be used much more commonly rather than the fixed-effects model. Unlike a fixed effects model, which assumes that the differences between the results of studies are due solely to chance, a random-effects model assumes that the effects being estimated in the different studies are not identical.

**Table 1. Example of summary results Table**

N°	AUTHORS (Author et al) + DATE	CONTEXT	STUDY DESIGN	STUDY POPULATION	STUDY AIMS	METHODS	RESULTS	CONCLUSIONS	ISSUES
1	XXXX et al. (2020)								

Once completed a thorough analysis of the existing literature on the subject, the authors are able to report their results making recommendations and suggestions for future investigation, that are based on the best available evidence. To report results, they can use the PRISMA STATEMENT checklist ([Appendix 2](#)). It consists of 27-items that aims to guide authors in improving the reporting of systematic reviews. While PRISMA's focus is systematic reviews of randomized trials, the tool can also be used as a blueprint for reporting systematic reviews of other types of studies, particularly those evaluating the effectiveness of health care intervention.

# GUIDELINES FOR RAPID REVIEWS

## What is a Rapid Review and what is its purpose?

A Rapid Review (or rapid evidence assessment) is a variation of a Systematic Review (SR) that balances time constraints with considerations for bias. “A rapid review is a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting specific methods to produce evidence for stakeholders in a resource-efficient manner.”

Evidence suggests that Rapid Reviews (RRs) are often used by policymakers as an efficient tool for their daily decision-making and are considered part of the knowledge synthesis family.

## How to perform a rapid review?

1. Define/refine the topic question, identify biases, plan and execute search
2. Screen, select and critically appraise the results
3. Synthesize the evidence ([Table 1](#))
4. Identifying applicability and transferability issues for further consideration during the decision-making process
5. Reporting

### ***1. Define/refine the topic question, identify biases, plan and execute search***

As previously described for SRs, the frameworks FINER and PICO can be used to find the key points and to refine the research question. Questions should be framed to be neutral rather than focused on a particular direction for the outcome. Framing the question in a neutral way helps identifying all of the research evidence relevant to the chosen topic. On the opposite, doing that with a particular direction for the outcome (e.g. the intervention has a positive effect) may bias the search for evidence.

In order to set the research question:

- Involve key stakeholders (e.g., review users such as consumers, health professionals, policymakers, decision-makers) to set and refine the review question, the eligibility criteria, and the outcomes of interest.
- Develop a protocol that includes review questions, PICO and inclusion and exclusion criteria.

If the findings of a RR can influence practice and policies, then the outcome selection needs to be relevant to clinicians, policymakers and patients. The

selection of databases will depend on the topic under review and access to them. It is recommended to limit grey literature and supplemental searching and place emphasis on higher quality study designs. If justified, study registries can be searched, and reference lists of other reviews or included studies after screening of the abstracts and full texts can be screened.

## **2. Screen, select and critically appraise the results**

Search results are screened based on the inclusion/exclusion criteria (using a table or worksheet to keep track of the screening and review process).

Critical appraisal is the process of assessing the study quality in order to determine if findings are trustworthy and meaningful. Critical appraisal helps to answer the question: *“Were the methods used good enough that I can be confident in the findings and apply these findings to public health practice?”*

If critical appraisal was not already conducted by a reliable source, it is necessary to critically appraise every document you include in the Rapid Review using an appropriate tool. These tools guide researcher through a series of questions to assess the quality of the methods used by the authors to conduct the research. ([Appendix 1](#))

## **3. Synthesize the evidence**

In this step, results from all of the high quality and relevant evidence gathered is aggregated. The overall goal is to come to a conclusion about what is known about the practice question in the literature. This step involves data extraction (including title, author(s), date of publication, type of study, settings and population studied, interventions implemented, outcomes measured and results). Also in this case, using a summary table to collect the evidences is recommended ([Table 1](#)).

## **4. Identifying applicability and transferability issues for further consideration during the decision-making process**

The Applicability and Transferability tool (accessible at the following link: <http://www.nccmt.ca/publications/9/view-eng.html>) is designed to capture important issues which are related to the proposed policy or program. Team members should discuss which factors included in the Applicability and Transferability tool it would be important to consider prior to making policy or practice decisions. The identified factors could then be included in supplementary documentation accompanying the rapid review, to be discussed by managers and senior managers, and any other relevant stakeholders.

In order to focus on equity issues related to the applicability and transferability of evidence, the Knowledge Translation Toolkit can be useful (<http://stmichaelshospitalresearch.ca/research-programs/urban-health-solutions/resources-and-reports/knowledge-translation-toolkit/>).

## **5. Reporting**

The steps of the process and their results and implications are documented in a final report (evidence brief). Synthesizing the evidence requires careful consideration of the commonalities and differences across the studies, along with weighting of results by their methodological quality.

## Bibliography

1. Brown P, Brunnhuber K, Chalkidou K, et al. How to formulate research recommendations. *BMJ*. 2006;333(7572):804-806. doi:10.1136/bmj.38987.492014.94
2. White paper for European Commission "The need for an Evidence-Based Research approach in health science". From the COST Action "EVBRES"
3. <https://scientific-publishing.webshop.elsevier.com/research-process/finer-research-framework/>
4. Bonczar M, Ostrowski P, D'Antoni AV, Tubbs RS, Iwanaga J, Ghosh SK, Klejbor I, Kuniewicz M, Walocha J, Moryś J, Koziej M. How to write an Umbrella Review? A step-by-step tutorial with tips and tricks. *Folia Morphol (Warsz)*. 2022 Dec 27. doi: 10.5603/FM.a2022.0104. Epub ahead of print. PMID: 36573368.
5. Henry BM, Tomaszewski KA, Walocha JA. Methods of Evidence-Based Anatomy: a guide to conducting systematic reviews and meta-analysis of anatomical studies. *Annals of Anatomy - Anatomischer Anzeiger*. 2016; 205: 16–21, doi: 10.1016/j.aanat.2015.12.002
6. Higgins JPT, Green S (eds). *Cochrane Handbook for Systematic Reviews of Interventions*
7. Higgins JPT, Thomas J, Chandler J, et al. (eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley 2019, doi: 10.1002/9781119536604
8. Italian National Health Institute. Strumenti di riferimento per la valutazione della qualità (reporting e conduct) e del rischio di bias degli studi. Available online at [https://snlg.iss.it/wp-content/uploads/2019/01/Strumenti\\_val\\_qualita%CC%80\\_studi.pdf](https://snlg.iss.it/wp-content/uploads/2019/01/Strumenti_val_qualita%CC%80_studi.pdf) (last access: 08.02.2023).
9. Lasserson TJ, Thomas J, Higgins JPT. Chapter 1: Starting a review. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).
10. Margaliot, Zvi M.D., M.S.; Chung, Kevin C. M.D., M.S.. Systematic Reviews: A Primer for Plastic Surgery Research. *Plastic and Reconstructive Surgery* 120(7):p 1834-1841, December 2007. | DOI: 10.1097/01.prs.0000295984.24890.2f
11. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).
12. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when

- choosing between a systematic or scoping review approach. *BMC Med Res Methodol*. 2018 Nov 19;18(1):143. doi: 10.1186/s12874-018-0611-x. PMID: 30453902; PMCID: PMC6245623.
13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71. PMID: 33782057; PMCID: PMC8005924.
  14. Page, M.J., McKenzie, J.E., Bossuyt, P.M. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* **10**, 89 (2021). <https://doi.org/10.1186/s13643-021-01626-4>
  15. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc*. 2015 Sep;13(3):141-6. doi: 10.1097/XEB.000000000000050. PMID: 26134548.
  16. PRISMA Transparent Reporting of Systematic Reviews and Meta-Analyses. Available online at: [www.prisma-statement.org](http://www.prisma-statement.org) (last access: 08.02.2023).
  17. PRISMA Transparent Reporting of Systematic Reviews and Meta-Analyses. Available online at: <https://www.prisma-statement.org/Extensions/ScopingReviews> (last access: 08.02.2023).
  18. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995; 123(3):A12-3. PMID: 7582737.
  19. Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).
  20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garrity C, Lewin S, Godfrey CM, Macdonald MT, Langlois EV, Soares-Weiser K, Moriarty J, Clifford T, Tunçalp Ö, Straus SE. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018 Oct 2;169(7):467-473. doi: 10.7326/M18-0850. Epub 2018 Sep 4. PMID: 30178033.
  21. Garrity C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J Clin Epidemiol*. 2021 Feb;130:13-22. doi: 10.1016/j.jclinepi.2020.10.007. Epub 2020 Oct 15. PMID: 33068715; PMCID: PMC7557165.
  22. Brouwers MC, Kerkvliet K, Spithoff K; AGREE Next Steps Consortium. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 2016 Mar 8;352:i1152. doi: 10.1136/bmj.i1152. Erratum

- in: BMJ. 2016 Sep 06;354:i4852. PMID: 26957104; PMCID: PMC5118873.  
(<https://www.agreetrust.org/resource-centre/agree-reporting-checklist/>)
23. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.  
(<http://onlinelibrary.wiley.com/doi/10.1002/9780470988343.app1/pdf>)
  24. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.  
(<http://onlinelibrary.wiley.com/doi/10.1002/9780470988343.app1/pdf>)
  25. <https://www.ncbi.nlm.nih.gov/books/NBK115843/bin/appe-fm2.pdf>
  26. Des Jarlais DC, Lyles C, Crepaz N; TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004 Mar;94(3):361-6. doi: 10.2105/ajph.94.3.361. PMID: 14998794; PMCID: PMC1448256. (<http://www.cdc.gov/trendstatement/>)
  27. <https://www.strobe-statement.org/>
  28. <https://www.equator-network.org/wp-content/uploads/2015/03/STARD-2015-checklist.pdf>
  29. <https://methods.cochrane.org/sites/methods.cochrane.org.prognosis/files/public/uploads/QUIPS%20tool.pdf>
  30. [https://cdn.elsevier.com/promis\\_misc/ISSM\\_COREQ\\_Checklist.pdf](https://cdn.elsevier.com/promis_misc/ISSM_COREQ_Checklist.pdf)

## Appendix 1. Most popular risk of bias evaluation tools

Study design	Tool	Web site
<b>Guidelines</b>	AGREE II	
		<a href="https://www.agreetrust.org/resource-centre/agree-ii/agree-ii-instructions/">https://www.agreetrust.org/resource-centre/agree-ii/agree-ii-instructions/</a>
<b>Systematic reviews with or without meta-analysis</b>	AMSTAR-2	<a href="https://amstar.ca/docs/AMSTAR-2.pdf">https://amstar.ca/docs/AMSTAR-2.pdf</a>
	ROBIS	<a href="http://www.bristol.ac.uk/population-health-sciences/projects/robis/">http://www.bristol.ac.uk/population-health-sciences/projects/robis/</a>
<b>Observational (etiology)</b>	ROBINS-E	<a href="https://www.riskofbias.info/welcome/robins-e-tool">https://www.riskofbias.info/welcome/robins-e-tool</a>
<b>Observational (prognosis)</b>	QUIPS Tool	<a href="https://www.acpjournals.org/doi/10.7326/0003-4819-158-4-201302190-00009">https://www.acpjournals.org/doi/10.7326/0003-4819-158-4-201302190-00009</a>
<b>Non randomised studies of interventions</b>	ROBINS-I	<a href="http://www.bmj.com/content/bmj/355/bmj.i4919.full.pdf">http://www.bmj.com/content/bmj/355/bmj.i4919.full.pdf</a>
<b>Diagnostic</b>	QUADAS-2	<a href="https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/">https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/</a>
<b>Clinical cases</b>	CARE	<a href="https://www.care-statement.org/checklist">https://www.care-statement.org/checklist</a>
<b>Qualitative studies</b>	NICE	<a href="https://www.nice.org.uk/process/pmg10/chapter/appendix-g-methodology-checklist-qualitative-studies">https://www.nice.org.uk/process/pmg10/chapter/appendix-g-methodology-checklist-qualitative-studies</a>
<b>RCT</b>	Cochrane Risk Of Bias 2 Tool	<a href="https://www.riskofbias.info/welcome/rob-2-0-tool">https://www.riskofbias.info/welcome/rob-2-0-tool</a>

Modified from: Italian National Health Institute. [https://snlg.iss.it/wp-content/uploads/2019/01/Strumenti\\_val\\_qualita%CC%80\\_studi.pdf](https://snlg.iss.it/wp-content/uploads/2019/01/Strumenti_val_qualita%CC%80_studi.pdf) (last access: 08.02.2023).



## Appendix 2. PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
<b>TITLE</b>			
Title	1	Identify the report as a systematic review.	
<b>ABSTRACT</b>			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
<b>METHODS</b>			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	

Section and Topic	Item #	Checklist item	Location where item is reported
<b>RESULTS</b>			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
<b>DISCUSSION</b>			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	
	23b	Discuss any limitations of the evidence included in the review.	
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	
<b>OTHER INFORMATION</b>			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71; For more information, visit: <http://www.prisma-statement.org/>