

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

Spoke 3 Deliverable  
**S3.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



**Ministero  
dell'Università  
e della Ricerca**



**PNC**  
Piano nazionale per gli investimenti  
complementari al PNRR  
Ministero dell'Università e della Ricerca

**DARE** | Digital  
Lifelong  
Prevention

## S3.D1.1 List of process and outcome indicators

Deliverable information	
Spoke number and title	Spoke 3 - Digitally Enabling Secondary and Tertiary Prevention
WP number and title	WP1 - Evidence, Outcome Indicators, and Stakeholders Engagement
Related task(s)	T1.1, T1.2, T1.3
Lead beneficiary	UNIPD
Contributing beneficiaries	
Dissemination level	Public, fully open
Due date	July 10, 2023
Actual date of delivery	July 7, 2023
Author(s)	Dario Gregori (UNIPD), Luca Vedovelli (UNIPD), Honoria Ocagli (UNIPD)
Contributors	
Quality Assurance	Stefania Boccia (UCSC), Salvatore Vitabile (UNIPA)

## Document history

Version	Date	Author(s) / Reviewer(s) (Beneficiary)	Description
0.1	20/06/2023	Luca Vedovelli (UNIPD), Honoria Ocagli (UNIPD), Dario Gregori (UNIPD)	First draft
0.2	1/07/2023	Luca Vedovelli (UNIPD), Honoria Ocagli (UNIPD), Dario Gregori (UNIPD)	Revision
0.3	05/07/2023	Stefania Boccia (UCSC), Mario Cesare Nurchis (UCSC), Andrea Gentili (UCSC)	Revision
1.0	7/07/2023	Luca Vedovelli (UNIPD), Honoria Ocagli (UNIPD), Dario Gregori (UNIPD)	Final document

## Disclaimer

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

## Table of contents

Publishable Summary .....	5
List of abbreviations.....	<b>Errore. Il segnalibro non è definito.</b>
1. Introduction .....	6
1.1. Study Design Evaluation .....	6
1.2. Sample Size and Power Evaluation.....	6
1.3. Statistical Plan Evaluation.....	7
1.4. Outcomes Evaluation.....	7
1.5. Accurate reporting.....	10
1.6. Stakeholders' engagement.....	12
2. Pilots Evaluation – Methodology and Implementation.....	14
1.7. Activities - Evaluation procedures.....	15
2. Stakeholders Engagement .....	16
3. References .....	17
Appendix 1 – Pilot evaluation form .....	19
Do the analyses reflect study objectives? .....	20
Are the methods appropriate?.....	20
Are the considerations regarding missing values, unused, and spurious data acceptable?.....	20
Are the considerations regarding multiplicity acceptable?.....	20
Are the planned analyses appropriate? .....	20
Appendix 2 – TRIPOD Checklist.....	22

## Publishable Summary

The overarching goal of the DARE project is to equip the Italian Ministry of Health and the National Health Service (NHS) with efficient, effective, and innovative technologies. This transition is geared towards revolutionizing the healthcare landscape to better serve our communities' needs. As a part of this endeavor, Spoke 3 WP1 assumes a critical role in evaluating the quality and appropriateness of proposed pilots. This task involves a rigorous assessment of several key elements, including study design, sample size justification, statistical analysis plan, and outcomes evaluation. Our role aligns with the DARE project's objective of implementing innovative technologies and advancing prevention strategies, with the ultimate aim of improving health outcomes.

Our first duty lies in evaluating study design, which is fundamental to the success of any pilot study. We are tasked with ensuring that the design adequately addresses the research question, reduces bias and confounding, and maximizes validity (Porta, 2008). Furthermore, it is our responsibility to ensure the appropriateness of the sample size. An adequate sample size provides sufficient statistical power, controls for random error, and enhances the credibility and generalizability of the findings (Charan & Biswas, 2013).

In addition, we scrutinize the statistical analysis plan. A robust plan strengthens the validity of a study and ensures the proper interpretation of results (Gelman et al., 2013). Moreover, we will encourage pilots to apply statistical methodologies that can handle deviations from the protocols since we are aware that the research process is not always linear, which should be accounted for in the statistical plan. Finally, we evaluate outcomes, focusing on their appropriateness, validity, and sensitivity to changes in the factors under investigation (Haber et al., 2021).

At WP1, we collaborate closely with various DARE consortium partners to develop a comprehensive list of indicators. These will be instrumental in assessing the utility and implementability of innovative technologies within the framework of pilot studies. We use two frameworks for this purpose: the GRADE Evidence to Decision (EtD) framework for partners implementing clinical studies (Alonso-Coello et al., 2016) and the Predictive Clinical Model framework for partners creating predictive clinical models using artificial intelligence, machine learning, and deep learning.

In sum, our mission at WP1 of Spoke 3 is to ensure the rigorous, scientific, and ethical conduct of pilot studies, all with the aim of advancing the objectives of the broader DARE project.

## 1. Introduction

### 1.1. Study Design Evaluation

A significant duty of WP1 in Spoke 3 of the DARE project is the rigorous evaluation of the study designs proposed in pilot studies. This involves ensuring that the chosen designs are appropriate for answering the research questions and meeting the study's objectives. The design of a study sets the blueprint for data collection and analysis, directly influencing the robustness and validity of findings (Leon, Davis, & Kraemer, 2011).

Our examination of study design includes an assessment of the chosen design's strengths and weaknesses and an appraisal of its alignment with the research question. This involves determining whether the study design is suitable for addressing the relationship between variables of interest, whether it is feasible within the constraints of available resources, and whether it reduces bias and confounding while maximizing internal and external validity (Sedgwick, 2014).

For example, a randomized controlled trial might be the most appropriate study design when evaluating the effectiveness of an intervention, while a cohort study might be more suitable when investigating the incidence and natural history of a condition. Furthermore, we examine the ethical considerations implicit in each study design, such as the balance of risks and benefits to participants (World Medical Association, 2013).

By examining the study design, WP1 ensures that the research undertaken as part of the DARE project is not only scientifically robust but also ethical and feasible, thereby ensuring that the evidence generated from each pilot is credible and meaningful.

### 1.2. Sample Size and Power Evaluation

As part of WP1's role in Spoke 3, we are responsible for the meticulous evaluation of sample size within the proposed pilots. This step is of paramount importance in maintaining the integrity of the research process and ensuring the results drawn from each pilot are both reliable and generalizable.

When assessing the proposed sample size, we utilize well-established statistical methods and principles to ascertain whether the size is appropriate for the pilot's specific context, the research question at hand, and the statistical techniques to be used (Charan & Biswas, 2013).

Adequacy in sample size not only impacts the feasibility of the study but also significantly determines the statistical power - the probability that the study will detect an effect of a certain size if it exists (Button et al., 2013). Insufficient sample sizes may lead to type II errors, thereby falsely concluding there is no effect when there is one. Conversely, an unnecessarily large sample size could lead to wasted resources and potential ethical issues. Therefore, we assess sample sizes against these principles, ensuring the efficient use of resources, ethical research practice, and robust conclusions (Moore et al., 2013).

### 1.3. Statistical Plan Evaluation

WP1 is also committed to the detailed assessment of the proposed statistical analysis plans. This scrutiny is crucial in assuring the credibility and validity of results drawn from the pilots. A robust statistical plan is the bedrock of reliable data interpretation and, thus, is an invaluable tool in health research (Gelman et al., 2013).

We evaluate whether the proposed statistical methods align with the objectives of the pilot, are appropriate for the study design, and are sensitive enough to detect anticipated effects. This process involves verifying the suitability of methods used to address missing data, account for confounding factors, and handle potential outliers. Furthermore, we consider the appropriateness of methods for controlling type I and type II errors and the validity of any interim analysis or stopping rules.

In essence, we aim to ensure that the proposed statistical plan is not only scientifically rigorous but also transparent, reproducible, and comprehensible to all stakeholders involved (The American Statistician, 2019).

### 1.4. Outcomes Evaluation

Lastly, WP1 is tasked with outcomes evaluation. This involves an in-depth assessment of the relevance, validity, and sensitivity of the outcomes chosen for each pilot study (Haber et al., 2021).

Outcomes should directly correspond to the research questions and objectives and be of paramount significance to the intended beneficiaries of the research.

We consider whether the chosen outcomes adequately reflect the effectiveness of the intervention and whether they are measurable, both qualitatively and quantitatively. In addition, we assess whether outcomes are patient-centered, ensuring that they reflect what truly matters to patients and their health.

Further, we examine the validity of the methods for measuring outcomes. This includes checking the use of validated measurement tools and scales and assessing plans for handling missing outcome data. Through this rigorous process, we aim to safeguard the relevance and utility of the results generated by each pilot, ultimately fostering research that impacts positively on health and wellbeing.

In the context of evaluating outcomes within the DARE project, WP1 has chosen to leverage the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement and checklist. The TRIPOD Statement was developed as an international initiative to improve the reliability and value of published health research literature by promoting transparent and accurate reporting of research studies (Collins et al., 2015).

When evaluating the outcomes of predictive models, the TRIPOD checklist offers a detailed guide to ensure comprehensive reporting of the methodologies employed. The checklist covers 22 key items that are critical for transparency in the development and validation of clinical prediction models. These include aspects such as data sources, participant details, handling of missing data, model-building strategies, presentation of model performance, results, and interpretation (Moons et al., 2015).

The robust application of the TRIPOD statement and checklist serves as a blueprint for WP1's evaluation of outcomes. It provides us with a standardized approach to ensuring that the results presented in the pilot studies are based on rigorous methodologies and are comprehensively reported. By doing so, we aim to enhance the replicability and utility of the findings, ultimately contributing to the validity of the outcomes of the DARE project.

Outcome Evaluation can be further understood by examining the discrimination of a predictive model. Discrimination is the model's capacity to distinguish between positive and negative outcomes.

These metrics analyze the model's competence in ranking observations based on their predicted probabilities. This is frequently employed in binary classification problems where the objective is to determine which observations belong to a specific category. In healthcare, for example, discrimination involves predicting the presence or absence of a particular condition or disease using various features or predictors. A common classification problem in medicine could be the utilization of medical records and diagnostic tests to predict whether a patient has conditions like diabetes, cancer, or heart disease. The model's discrimination ability is typically assessed using metrics such as the C-statistic and the D-statistic, although there are more advanced metrics suitable for various contexts (Fawcett T, 2006; Powers DEW, 2011; Steyeberg EW, 2010)).

**C-statistic:** Also known as the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), when logistic regression is the underlying modeling strategy, the C-statistic measures the model's capacity to distinguish between positive and negative outcomes. It plots sensitivity (true positive rate) against (1 - specificity) or the false positive rate for different decision thresholds of the model. It evaluates the trade-off between sensitivity and specificity. Hence, the 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 model that predicts randomly has an AUC of 0.5.

**D-statistic:** Also known as Somers' D, the D-statistic is a rank correlation coefficient that measures the strength of the association between the predicted probabilities and the observed outcomes.

In synergy with Spoke 1, we will stress the importance of calibration in clinical prediction models. Model calibration plays a pivotal role in outcome evaluation, and it denotes the level of correspondence between the predicted probabilities of a model and the actual outcomes within the examined data. In healthcare research, it is paramount to ensure model calibration because it ascertains that the predicted probabilities genuinely mirror the real likelihood of an outcome in the studied population. This implies that a well-calibrated model will, on average, accurately predict the possibility of an event transpiring within a patient group. For instance, if a model projects a high

risk of a particular disease for 70% of patients, then, in reality, 70% of the patients in the actual population should indeed be at high risk for that disease.

To assess calibration in predictive models, the Van Calster hierarchy (Van Calster B et al., 2016) was selected as one of the most appropriate frameworks. The Van Calster triad systematically evaluates calibration assessment levels, which consist of varying calibration degrees:

**Average calibration (calibration-in-the-large):** This aspect pertains to a scenario wherein a predictive model demonstrates good overall calibration, with negligible discrepancies between the predicted probabilities and the actual risk probabilities. In essence, the model accurately predicts the risk probabilities on average, albeit with potential variability in predictions for individual patients. Average calibration suggests a satisfactory overall performance of the model, although it may benefit from fine-tuning to enhance its performance for specific patient subgroups or outcomes.

**Weak calibration (calibration intercept and slope):** This degree of calibration is relevant when a predictive model shows some calibration, but the predicted probabilities are consistently too extreme or too conservative. In other words, the model may tend to either overestimate or underestimate the actual risk probabilities. Weak calibration, being the second level of calibration assessment in the Van Calster hierarchy, implies that the model requires improvement to reach better concurrence between the predicted probabilities and the actual risk probabilities.

**Moderate calibration:** This degree applies to a situation where a predictive model exhibits reasonable calibration, yet there are systematic deviations between the predicted probabilities and the actual risk probabilities. This means that the model might somewhat overestimate or underestimate the actual risk probabilities. Moderate calibration, as the intermediary level of calibration assessment in the Van Calster hierarchy, proposes that the model can be further refined to achieve improved alignment between the predicted probabilities and the actual risk probabilities.

## 1.5. Accurate reporting

Reporting is a crucial aspect of outcome evaluation. Transparency and thoroughness in reporting not only validate the integrity of a study but also facilitate its replication, further investigation, and the practical application of its findings.

To uphold these principles, it is vital to adopt clear and clean data management practices. This includes accurate and comprehensive record-keeping, adherence to standardized data format and nomenclature, and meticulous data validation and verification procedures. These practices ensure the robustness and reliability of the data, making it a reliable foundation for analysis and interpretation.

Advocating for open data and open code is another crucial element in transparent reporting. Open data principles promote the public availability and usability of data, while open code allows for transparency in the methodology of analysis. This open approach enables other researchers to verify findings, replicate studies, and build upon existing work, thereby advancing scientific knowledge. Furthermore, for an analysis or experiment to be fully reproducible, the original data and the complete method of analysis must be available. This includes all the code and algorithms used in data cleaning, manipulation, analysis, and visualization. By providing full details about the research design, data collection, and analysis procedures, other researchers can understand, evaluate, and replicate the study, fostering a scientific environment built on trust, transparency, and collaborative progress.

Ensuring comprehensive, transparent reporting through open data and open code principles is a responsibility that requires meticulous attention and organized procedures. The following key steps will guide our strategy to maintain and enforce these principles:

- 1. Data Management:** Effective data management begins with a well-structured data management plan, which outlines how data will be collected, processed, stored, and shared. This should include standardizing formats and nomenclature and consistently applying validation and verification measures to maintain the accuracy and reliability of the data. Regular audits will be conducted to ensure compliance with these standards.
- 2. Adherence to Open Data Principles:** Data generated during the project should be made publicly available in a manner consistent with open data principles. This includes using accessible, non-proprietary formats, providing comprehensive metadata for understanding and re-use, and storing the data in a secure, reputable repository that ensures long-term availability. To facilitate these efforts, we will establish procedures for anonymizing or de-identifying sensitive data and obtaining necessary permissions for data sharing.

3. **Open Code Practices:** In alignment with the principles of open science, the code used for data cleaning, manipulation, analysis, and visualization should be publicly shared. This includes detailed documentation that explains the purpose and operation of each section of the code, thereby making our analytical methods transparent and understandable. Platforms like GitHub, Open Science Framework, Zenodo, or similar for storing and sharing the code will enable others to inspect, use, and build upon it.
4. **Transparent Reporting:** Finally, when reporting our results, projects should adhere to established guidelines such as the TRIPOD statement for transparent reporting of prediction model studies. This involves clearly describing all steps of the research process, from study design and data collection to data analysis and interpretation, ensuring others can understand, evaluate, and replicate our study.

By incorporating these steps into our work practices, we can uphold our commitment to transparent, reproducible research and contribute to a culture of openness and integrity in science.

## 1.6. Stakeholders' engagement

Inherent in the conception and execution of the DARE project is our diverse stakeholders' active and substantial engagement. These encompass researchers, policymakers, funding agencies, health authorities, and health professionals, each bringing unique insights and perspectives that enrich the project's breadth and depth.

Our approach to stakeholder engagement is rooted in a reciprocal and ongoing dialogue that seeks to align the project's objectives and strategies with our stakeholders' varied needs and expectations. This active discourse not only facilitates transparency and inclusivity but also encourages knowledge exchange, fosters collaboration, and promotes a shared understanding and ownership of the project's mission and outcomes.

Researchers provide critical insights from the frontline of technological innovations and advancements in preventative health. At the same time, policymakers ensure that our project is aligned with the broader health policy landscape, helping to drive change at a system level. Funding agencies, through their financial contributions and oversight, ensure accountability and facilitate strategic planning. Health authorities offer necessary regulatory guidance and approval, ensuring

that our outputs are compliant, safe, and ethically sound. The invaluable perspectives of health professionals, those who will ultimately implement these innovations in their practice, ground our project in practical realities and patient needs, ensuring that our efforts are focused on generating actionable and meaningful improvements in health outcomes.

As WP1 in Spoke 3 of the DARE project, our role in stakeholder engagement is multifaceted and integral to the overall success of the project. We will be involved in:

- 1. Mapping and Identifying Stakeholders:** WP1 is responsible for identifying key stakeholders related to the secondary and tertiary prevention areas of the project. This involves mapping the stakeholder landscape, identifying individuals and groups who can influence or be influenced by the project, and understanding their needs, interests, and potential contributions.
- 2. Engagement Strategy Development:** We are tasked with devising effective strategies for stakeholder engagement, which includes setting clear objectives, outlining methods of communication, and defining the frequency and type of interaction. This strategy is aligned with the overall project goals and adjusted according to the evolving needs of the project.
- 3. Facilitating Inclusive Dialogue:** WP1 organizes regular meetings and workshops to involve stakeholders in various stages of the project, creating a platform for constructive dialogue and cooperation. These interactive sessions encourage stakeholders to share their knowledge, voice their concerns, and contribute to decision-making processes.
- 4. Communication and Transparency:** We ensure consistent and transparent communication with all stakeholders, keeping them updated on the project's progress, achievements, and challenges. We utilize various tools and platforms for communication to ensure that all stakeholders can access information and engage effectively.
- 5. Feedback and Consultation:** As WP1, we actively seek feedback from our stakeholders and incorporate their insights into the project. We establish mechanisms to capture stakeholder feedback and implement changes based on these inputs to ensure the project stays relevant and impactful.

6. **Training and Support:** WP1 also undertakes capacity-building activities to empower stakeholders to contribute effectively to the project. This includes training sessions on the GRADE EtD and Predictive Clinical Model frameworks and other relevant areas.
7. **Advocacy and Liaison:** We act as advocates for the stakeholders, ensuring that their voices are heard and their concerns are addressed in the broader project framework. We also facilitate effective liaison between stakeholders and the different work packages within the DARE project.

Through these roles, WP1 serves as a central pillar of stakeholder engagement in the DARE project, ensuring the meaningful involvement of diverse stakeholders in all aspects of the project's development and implementation.

## 2. Pilots Evaluation – Methodology and Implementation

The validity of the proposed pilots' statistical plan and design is a fundamental factor in the credibility and reliability of the research conducted. To this end, we have constructed an evaluation form incorporating six key aspects derived from the European Medicines Agency (EMA) guidelines for ethical committees ([euclinicaltrials.eu](http://euclinicaltrials.eu)). This ensures that the proposed research adheres to exemplary scientific practices, thereby enhancing the dependability and applicability of its outcomes in the healthcare sector.

The development of the form involved a collegiate process of adaptation that capitalized on the knowledge of our expert panel, comprising Prof. Dario Gregori, two post-doctoral researchers experienced in clinical trials and studies, and a resident in the third year of Medical Statistics. The evaluation form achieved consensus after two meetings and was then assessed externally by a non-participating Assistant Professor of medical statistics from the University of Ferrara.

The evaluation process was undertaken by a senior post-doctoral researcher and a resident in the third year of medical statistics, with a final assessment by Prof. Dario Gregori. Discrepancies, if any, were addressed through discussion among reviewers and recorded separately for transparency.

The items of evaluation are as follows:

**Study Plan and Design:** The underpinning plan and design of a study represent critical components as they delineate the methodological approach, thereby determining the reliability and utility of the results (Grimes & Schulz, 2002).

**Randomization and Blinding:** These techniques are crucial in mitigating bias in experimental research. Randomization ensures an equal probability of participant allocation to the groups under investigation, thereby controlling for confounding variables. Blinding prevents intervention knowledge from influencing participant behaviour or experiment results (Suresh, 2011; Hróbjartsson & Boutron, 2011).

**Sample Size, Trial Power, and Level of Significance Used:** These factors substantially influence the study's capacity to detect an effect if it exists. Inadequate sample size could undermine statistical power, possibly leading to a false negative result. Therefore, the justification and calculation for these factors must be appropriately substantiated (Biau et al., 2008).

**Planned Analysis:** The proposed analyses must align with the research questions and the data. This item considers handling missing, unused, or spurious data, and multiple comparisons, as these factors could impact results interpretation (ICH E9).

**Interim Analysis:** Interim analyses contribute to the ongoing data monitoring during the trial. They can facilitate early termination of a trial if the results become evident, conserving resources and potentially reducing unnecessary participant exposure (Fleming & DeMets, 1993).

**Assessor's Overall Conclusion on the Statistical Part:** This section enables the assessor to render an overarching judgement on the statistical robustness of the proposed study, pinpointing any concerns or areas necessitating further clarification.

## 1.7. Activities - Evaluation procedures

In WP1, we have worked towards establishing rigorous evaluation procedures for the DARE project's pilots. The inception of these procedures took the form of a collegiate adaptation process, leveraging the expertise of our panel comprising Prof. Dario Gregori, two post-doctoral researchers with experience in clinical trials and studies, and a third-year resident in Medical Statistics. The consensus on the evaluation form was reached after two meetings and an external evaluation by an

Assistant Professor of Medical Statistics from the University of Ferrara, who was not involved in the form's creation.

Following the DARE project's kick-off plenary on January 26, 2023, and Spoke 3's initiation on February 10, 2023, a meeting was organized by WP1 leader Prof. Dario Gregori to settle on the tools for evaluating the methodological and statistical components of the pilots. The selected tool was adopted from the European Medicines Agency (EMA) standards and tailored to meet WP1's specific needs. The final version, an electronic form, was created by a team of three senior and one junior experts, along with an external reviewer. Each pilot was evaluated using this form, the completed versions of which were recorded and sent to the respective Spoke leaders.

To further support these efforts, several WP meetings and events were organized on April 20, 2023, for brainstorming on evaluation tools; May 5, 2023, for evaluation of March and April's pilots; June 4, 2023, for evaluation of June's pilots; and June 20, 2023, for evaluation of pilots submitted in late June.

## 2. Stakeholders Engagement

In WP1, we have made considerable strides towards effective stakeholder engagement. Initially, we established a collaboration with Spoke One to delineate the profile of our stakeholders and strategize on effective interaction methods. Key points of synergy were identified to streamline the engagement process and foster productive collaboration. We highlighted areas of shared interest and benefit, which serve as focal points around which our engagement activities revolve. Furthermore, we are currently in the process of developing robust involvement procedures. These procedures aim to standardize how we involve stakeholders in various project stages, ensuring consistency, inclusivity, and effectiveness in our stakeholder engagement endeavors. Through these steps, we are continually working to strengthen our rapport with stakeholders and promote their active and meaningful participation in the DARE project.

### 3. References

- [1] Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016.
- [2] Porta M, editor. *A Dictionary of Epidemiology*. 5th ed. Oxford University Press; 2008.
- [3] Charan J, Biswas T. How to calculate sample size for different study designs in medical research? *Indian J Psychol Med*. 2013;35(2):121–126.
- [4] Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis*. 3rd ed. CRC Press; 2013.
- [5] Haber N, Smith ER, Moscoe E, et al. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PLoS One*. 2018;13(5):e0196346.
- [6] Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365–376.
- [7] Moore CG, Carter RE, Nietert PJ, Stewart PW. Recommendations for planning pilot studies in clinical and translational research. *Clin Transl Sci*. 2011;4(5):332–337.
- [8] Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The Role and Interpretation of Pilot Studies in Clinical Research. *Journal of Psychiatric Research*, 45(5), 626–629.
- [9] Sedgwick, P. (2014). What is a non-randomised controlled trial? *BMJ*, 348, g4111.
- [10] World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20), 2191–2194.
- [11] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
- [12] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet*. 2002 Jan 5;359(9300):57–61. doi: 10.1016/S0140-6736(02)07283-5.
- [13] Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011 Jan;4(1):8–11. doi: 10.4103/0974-1208.82352.

- [14] Hróbjartsson A, Boutron I. Blinding in randomized clinical trials: imposed impartiality. *Clin Pharmacol Ther.* 2011 Nov;90(5):732-6. doi: 10.1038/clpt.2011.207.
- [15] Biau DJ, Kernéis S, Porcher R. Statistics in brief: The importance of sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res.* 2008 Sep;466(9):2282-8. doi: 10.1007/s11999-008-0346-9.
- [16] Statistical principles for clinical trials (ICH E9): <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials-scientific-guideline#current-version-section>
- [17] Fleming TR, DeMets DL. Monitoring of clinical trials: issues and recommendations. *Control Clin Trials.* 1993 Apr;14(2):183-97. doi: 10.1016/0197-2456(93)90002-u.
- [18] Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1–W73.
- [19] Fawcett, Tom (2006). "An Introduction to ROC Analysis". *Pattern Recognition Letters.* 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010.
- [20] 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.
- [21] Somers RH. A new asymmetric measure of association for ordinal variables. *American Sociological Review* 1962; 27(6): 799–811.
- [22] 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.
- [23] 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>

## Appendix 1 – Pilot evaluation form

### STATISTICAL/METHODOLOGICAL ASSESSMENT

#### Study plan and design

<b>Type of design:</b>	
<b>Controlled/non controlled?</b>	Controlled <input type="checkbox"/> Non controlled <input type="checkbox"/>
<b>Randomised?</b>	Yes <input type="checkbox"/> No <input type="checkbox"/>
<b>Blinding (Masking)?</b>	Open-label <input type="checkbox"/> Blinded evaluator <input type="checkbox"/> Single-blind <input type="checkbox"/> Double-blind <input type="checkbox"/>
<b>Brief description of the study plan and design:</b> <a href="#">Note</a>	
<i>E.g. Multicenter, open-label extension of two ongoing studies to assess the safety of Eplontersen (ION-682884) in patients with Transthyretin-Mediated Amyloid Cardiomyopathy.</i>	
Is the proposed study design acceptable? Yes <input type="checkbox"/> No <input type="checkbox"/>	
<b>Workspace:</b>	
<b>Assessor's comment:</b>	
<i>E.g. The planned design is appropriate.</i>	

#### Randomisation and blinding

<b>Brief description of the randomisation and blinding procedures:</b> <a href="#">Note a</a>
<b>Workspace:</b>
<b>Assessor's comment:</b>
<i>E.g. Not applicable.</i>

#### Sample size, trial power and level of significance used

Number of subjects planned to be enrolled:
--

E.g. 102

The sample size calculation/justification is acceptable. Yes  No

Trial power and level of significance is acceptable. Yes  No  NA

**Brief** description of sample size, trial power and level of significance: [Note](#)

**Workspace:**

**Assessor's comment:**

*E.g. No sample size calculations were performed as this is an extension study to two previous studies. Up to 1400 subjects from both studies will be enrolled.*

### Planned analysis

**Brief** description of the planned analyses: [Note](#)

Do the analyses reflect study objectives?  
Yes  No  Other, comment

Are the methods appropriate?  
Yes  No  Other, comment

Are the considerations regarding missing values, unused, and spurious data acceptable?  
Yes  No  Other, comment

Are the considerations regarding multiplicity acceptable?  
Yes  No  Other, comment

Are the planned analyses appropriate?  
Yes  No

**Workspace:**

**Assessor's comment:**

*E.g. No formal hypothesis testing will be done. The study is mainly descriptive. Multiplicity is not applicable to this open label extension study.*

**Interim analysis**

Interim analysis (IA) is proposed for this trial? Yes  No

Brief description of the interim analysis(es): [Note](#)

**Workspace:**

**Assessor's comment:**

*E.g. N/A*

**Assessor's Overall Conclusion on the Statistical Part**

The statistical aspects of the application are acceptable

Supplementary information needs to be provided (refer to the list of requests for additional information)

**Workspace:**

**Overall comment/ conclusion on the statistical assessment:**

*E.g. There are no further comments in relation to the statistical methodology.*

**REQUESTS FOR ADDITIONAL INFORMATION: STATISTICAL:**

**6.6.1.A PROPOSED LIST OF REQUESTS FOR ADDITIONAL INFORMATION BY RMS**

**Workspace (List of proposed RFI):**

*E.g. Provide more details on cross-validation approaches*

## Appendix 2 – TRIPOD Checklist

Section/Topic	Item	Checklist Item	Page
<b>Title and abstract</b>			
Title	1	D;V Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	
Abstract	2	D;V Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	
<b>Introduction</b>			
Background and objectives	3a	D;V Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	
	3b	D;V Specify the objectives, including whether the study describes the development or validation of the model or both.	
<b>Methods</b>			
Source of data	4a	D;V Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	

	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	
	5b	D;V	Describe eligibility criteria for participants.	
	5c	D;V	Give details of treatments received, if relevant.	
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	
Sample size	8	D;V	Explain how the study size was arrived at.	
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	

	10c	V	For validation, describe how the predictions were calculated.	
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	
<b>Results</b>				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	

	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	
	15b	D	Explain how to use the prediction model.	
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	
<b>Discussion</b>				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	
<b>Other information</b>				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	

Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	
---------	----	-----	---	--

\* Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.