Scientific Research Project Proposal – TEMPLATE – pls WRITE DOWN IN English

Supervisor name (same as in SRP checklist): Janet MacNeil-Vroomen

Supervisor email address (same as in SRP checklist): j.l.macneil-vroomen@amsterdamumc.nl

KIK tutor: Martijn Kersloot, m.g.kersloot@amsterdamumc.nl

Title of SRP project: Improving data quality and reusability of the Amsterdam Ageing Cohort

Location of SRP project: Amsterdam UMC, location **AMC and VUmc**- Department of Internal Medicine, Geriatric Medicine Section

Introduction

Longitudinal data, collected from the same individuals over time, are increasingly important in health research because they can show how health, functioning and care needs change [1]. This matters especially for ageing in place, where such data have been used to identify why some older adults with comparable medical needs remain living independently longer than others [1]. The Amsterdam Ageing Cohort (AAC) is an ongoing longitudinal cohort of patients visiting the outpatient geriatric clinic of Amsterdam UMC, location VUmc. It is built from Comprehensive Geriatric Assessment (CGA) data collected during routine care [2]. The cohort currently contains over 2500 patients and, with a fast-growing older population, is an increasingly valuable resource for research.

Whilst much data is collected in healthcare, including through longitudinal cohort studies, it is not necessarily easily for clinicians and researchers to access and use this data. Dutch data standards authority Nictiz defines usable data as those data accessible through information systems that are user-friendly, efficient and aligned with work processes [3]. Whilst usability of data specifically addresses what is required to make data easy to process, it is also considered increasingly important that data collected for care and research are FAIR: findable, accessible, interoperable and reusable [4]. Whilst well-designed user-facing software and user interfaces are required to make data usable, additional steps must be taken to make data FAIR.

Description of the Problem and SRP Project

Like many cohorts built from routine care data, the AAC faces data quality issues. For example, the status of participants is not updated regularly, so deceased patients may still be contacted to participate in follow-ups. In addition, the cohort lacks outcomes such as hospital admissions or causes of death. Linking the cohort to other datasets, such as hospital EHR data or national registries, could enrich the data, but only if the data are interoperable. The dataset is stored in a Microsoft Access database on the internal network, which makes linkage, sharing, and automated quality checks against external sources (e.g., EMR data) difficult. This negatively impacts the usability of data for researchers and health care professionals. Lastly, the dataset and its metadata are not described or findable outside the research team, and there is no procedure to request access, so potential users do not know it exists and cannot use it to its full potential.

The project consists of three phases: (1) identifying barriers and facilitators for data quality and reusability, (2) implementing improvements based on these findings, and (3) evaluating the improvements from the perspective of data quality, usability of the data for end-users, and the extent to which the cohort data are FAIR.

Research objective and research questions

The objective is to identify the barriers and facilitators to the data quality and reusability of the AAC, to implement targeted improvements, and to evaluate whether these improvements make the cohort data easier to use for clinicians and researchers in the geriatrics department who are not data scientists.

The main research question is: How can the data quality, interoperability and accessibility of the Amsterdam Ageing Cohort be improved so that clinicians without a data science background can reliably use it?

1. What is required to sustainably improve the quality of the data in the Amsterdam Ageing Cohort?
2. What is required to make the data and the management of the dataset usable for clinicians and clinical researchers?
3. What is required to make the cohort dataset more-FAIR?

Expected results

The project will deliver:

- A scientific manuscript in the form of a master thesis, one or more chapters of which may be the basis of a scientific publication
- A prioritised overview of barriers and facilitators
- A set of implemented improvements (e.g., automated quality checks and updates, a machine-readable data dictionary, and a published cohort description)
- An evaluation of the effect of implemented improvements.

References

1. Amankour F, Krabbe-Alkemade Y, Koolman X, Broese van Groenou M, Portrait F. Non-medical barriers and facilitators of ageing in place: a cohort study in the Netherlands. Eur J Ageing. 2025 May 30;22(1):23. doi: 10.1007/s10433-025-00859-9. PMID: 40447834; PMCID: PMC12125428.

2. Rhodius-Meester HFM, van de Schraaf SAJ, Peters MJL, Kleipool EEF, Trappenburg MC, Muller M. Mortality Risk and Its Association with Geriatric Domain Deficits in Older Outpatients: The Amsterdam Ageing Cohort. Gerontology 2021;67:194–201.

3. Nictiz. Bruikbare data and usability. Available from: https://www.nictiz.nl/wat-we-doen/zorginformatiestelsel/bruikbare-data-usability/, accessed 8-10-26.

4. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

GenAI Authenticity Statement

I accept my responsibility to follow the SRP Rules and Guidelines document fully, including its Generative AI (GenAI) policy (in Section 16.3). I also certify that I am fully responsible for all the content in this document and that I only used GenAI tools (if any) for the purposes I disclose below.

I did not use any GenAI tool in any way as part of my SRP proposal.