

Is Data Collection Harmonized Across Italian Clinical Cohorts on Alzheimer's and Other Dementias? A Systematic Review

Migliazzo Claudia^(1,2), Lodico Manuela^(1,2), Allegra Patrizio⁽²⁾, Maniscalco Laura⁽²⁾, Tarantino Domenico^(1,3), Piccoli Tommaso^(1,3), Vanacore Nicola⁽⁴⁾, Salemi Giuseppe^(1,3), Matranga Domenica⁽²⁾

(1) Department of Diagnostic, Interventional and Stroke Radiology, UOC Neurology, AOUP 'P. Giaccone', Palermo

(2) Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties" - PROMISE - University of Palermo, 90127 Palermo, Italy

(3) Department of Biomedicine, Neuroscience and Advanced Diagnostics (BIND), University of Palermo, Palermo, Italy

(4) National Center for Disease Prevention and Health Promotion, Italian National Institute of Health, Rome, Italy

CORRESPONDING AUTHOR: Migliazzo Claudia, claudia.migliazzo@gmail.com.

INTRODUCTION

Dementia, particularly Alzheimer's disease, is a progressive neurodegenerative condition that severely affects patients' quality of life and imposes a substantial burden on healthcare and social systems. Recent estimates indicate that over 57 million people worldwide are living with dementia, and this number is expected to triple by 2050. In Italy, current estimates exceed 1.2 million cases. The increasing incidence highlights the urgent need to enhance prevention strategies, especially considering that approximately 40% of cases can be attributed to modifiable risk factors, as noted by the Lancet Commission in 2020 report (1)

Despite scientific recognition of the importance of prevention, Italy's involvement in international dementia research networks (e.g., GAAIN, DPUK, ADDI) remains marginal. A major barrier is the fragmentation and heterogeneity of data collected across different Italian observational studies. The lack of standardized methodologies for data collection, classification, and risk factor analysis hampers comparability between studies and limits their integration into meta-analyses or cohort inclusion in existing networks (2).

OBJECTIVES

As part of the PREV-ITA-DEM project (PN-RR-MAD-2022-12375822), funded by the Italian Ministry of Health, we conducted a systematic review to identify, map, and evaluate observational studies carried out in Italy between 2019 and 2024 focusing on dementia, Alzheimer's disease, and modifiable risk factors.

Specifically, the review aimed to: assess the methodologies and the tools used; develop data harmonization rules,

based on internationally recognized standards (3) to support integration of the Italian cohorts into global research networks; propose validated protocols and tools to improve the data quality and comparability, suggesting a methodology that can be transferred also to non-modifiable factors such as biomarkers and genetic data.

METHODS

A systematic review of observational studies of Italian clinical cohorts was conducted using MEDLINE, Embase, and Scopus, covering the period from January 2019 to December 2024. Studies were included if published in English and focused on at least one of the eight modifiable risk factors for dementia (BMI, hypertension, diabetes mellitus, dietary habits, alcohol consumption, depressive symptoms, physical inactivity, and smoking).

Data were extracted on measurement methods, units used, assessment tools (e.g., questionnaires, clinical tests), diagnostic criteria, and classification categories.

Definitions and international criteria were used to assess compatibility and the DataSHaPER (Data Schema and Harmonization Platform for Epidemiological Research) methodology was applied to evaluate the degree of variables' harmonization (3). Variables were classified as completely harmonizable when they strictly adhered to standard definitions and formats, allowing direct comparison without any loss of information; partially harmonizable when some discrepancies or incomplete data could result in information loss; and impossible to harmonize when insufficient or incompatible data prevented meaningful alignment.

This rigorous approach facilitates the identification of gaps and inconsistencies in data collection methods.

RESULTS

Of the 365 articles initially identified, 18 Italian observational studies met the inclusion criteria. Study designs included cross-sectional, longitudinal, and case-control approaches, with sample sizes ranging from fewer than 100 to more than 5.000 participants.

The analysis revealed significant heterogeneity in data collection protocols. Obesity, assessed via BMI, was the best documented factor: 38% of the studies showed complete harmonization with international standards, while 33% showed partial harmonization. Dietary habits, assessed through FFQ questionnaires, also showed reasonable harmonization (30% complete).

Smoking (28% complete harmonization), physical inactivity (using IPAQ), and alcohol consumption were addressed with variable tools and categorizations, while depression (measured with GDS or CES-D) and diabetes (through fasting blood glucose) showed only 22% complete harmonization. Variability in tools and their application emerged as a major obstacle to data comparability. Moreover, in many cases, key methodological details such as diagnostic thresholds or data collection frequency were missing.

DISCUSSION

The results of the review highlight methodological fragmentation that undermines the ability to integrate Italian data into large international consortia. The systematic adoption of recognized tools and validated protocols for the assessment of risk factors will enable the generation of more robust evidence, enhance the statistical power of analyses, and support the inclusion of Italian cohorts in international consortia. This represents a critical step toward making a meaningful contribution to the fight against dementia, with direct benefits for public health and the sustainability of the national healthcare system (2).

REFERENCES

1. Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., ... & Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The lancet*, 396(10248), 413-446.
2. Toga, A. W., Phatak, M., Pappas, I., Thompson, S., McHugh, C. P., Clement, M. H., ... & Gallacher, J. (2023). The pursuit of approaches to federate data to accelerate Alzheimer's disease and related dementia research: GAAIN, DPUK, and ADDI. *Frontiers in Neuroinformatics*, 17, 1175689.
3. Fortier, I., Burton, P. R., Robson, P. J., Ferretti, V., Little, J., L'heureux, F., ... & Hudson, T. J. (2010). Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *International journal of epidemiology*, 39(5), 1383-1393.