CODE NO. PNC0000002

Spoke 2 Deliverable

SP2.D5.2
Risk analyses and modeling

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

Spoke 2 - D5.2. Risk Analyses and Modeling

Deliverable information	
Spoke number and title	SP2 - Community-based Digital Primary Prevention
WP number and title	WP5 - Lifelong prevention approaches across the lifespan
Related task(s)	Task 5.1, 5.2, 5.3
Lead beneficiary	UNIPD
Contributing beneficiaries	UNIBO, UNIPA
Dissemination level	Public, fully open
Due date	15/06/2024
Actual date of delivery	12/06/2024
Author(s)	Silvia Bressan (UNIPD), Veronica Casotto (UNIPD) Barbara Carretti (UNIPD), Elisa Cainelli (UNIPD), Irene Mammarella (UNIPD), Tiziana Pozzoli (UNIPD) Giovanna Mioni (UNIPD), Andrea Zangrossi (UNIPD), Nicola Veronese (UNIPA), Mario Barbagallo (UNIPA)
Contributors	Cristina Canova (UNIP), Anna Cantarutti (UNIMIB), Dolores Catelan (UNIPD), Marianna Costa (UNIPD), Joaquin Gutierrez De Rubalcava (UNIPD), Roberta Biundo (UNIPD), Nicola Cellini (UNIPD), Elisa di Rosa (UNIPD), Alberto Testolin (UNIPD), Giovanni Mento (UNIPD), Enrico Toffalini (UNIPD), Silvia Lanfranchi (UNIPD), Sara Onnivello (UNIPD), Camiella Orefice (UNIPD), Emanuele Amodio (UNIPA), Alessandra Simonelli (UNIPD)
Quality Assurance	Salvatore Vitabile, Francesco Prinzi (UNIPA)

Document history

Version	Date	Author(s) /Reviewer(s) (Beneficiary)	Description
0.1	12/06/2024	Silvia Bressan (UNIPD), Veronica Casotto (UNIPD), Barbara Carretti (UNIPD), Elisa Cainelli (UNIPD), Irene Mammarella (UNIPD), Tiziana Pozzoli (UNIPD), Giovanna Mioni (UNIPD), Andrea Zangrossi (UNIPD), Nicola Veronese (UNIPA), Mario Barbagallo (UNIPA)	First draft
0.2	13/06/2024	Salvatore Vitabile (UNIPA), Francesco Prinzi (UNIPA)	Revision
0.3	15/06/2024	Silvia Bressan (UNIPD), Veronica Casotto (UNIPD), Barbara Carretti (UNIPD), Elisa Cainelli (UNIPD), Irene Mammarella (UNIPD), Tiziana Pozzoli (UNIPD), Giovanna Mioni (UNIPD), Andrea Zangrossi (UNIPD), Nicola Veronese (UNIPA), Mario Barbagallo (UNIPA)	Final document
0.4			
1.0			

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.

Table of contents

Publishable summary

The objective of SPOKE 2 – Work Package 5 is to establish an epidemiological framework that will allow for the optimal exploitation of real-world and research-based biological, clinical, psychological, social, and environmental data. This exploitation will enable personalised, multidimensional disease risk prediction, as well as the implementation of more effective lifelong preventive strategies. Five pilots have been designed to fulfil the research aim of WP5. Each pilot presents its distinct structure, characterised by the data sets used and the analytical methods identified. Four pilot studies will collect data prospectively eventually providing research datasets for future integration with existing real-world or administrative data sources. These pilots have the main objective of identifying optimal digital screening instruments and early markers of disease. One pilot study focuses, instead,on the integrated use of different existing data sets including routine primary care data, environmental, socio economic and administrative data to identify disease risk factors. Depending on the research question, design, and type of variables collected varied, approaches to analysis plans have been developed including both traditional and artificial intelligence driven analysis approaches. This second deliverable entitled “Risk Analyses and Modeling” aims at defining the plan and analyses models for the study of risk factors of disease development and early markers of diseases across the pilot projects.

1. Introduction

The lifespan of the world's population is increasing, and the proportion of people over 60 years old is predicted to rise from 12% in 2015 to 22% in 2050 (World Health Organization, 2015).

Lifelong health in a rapidly changing and increasingly digitalised society is a current global challenge. It is well known that laying the ground for a healthy life starts in childhood and continues throughout adulthood. The metabolic and brain plasticity during childhood creates the ideal opportunity to implement preventive strategies, which, in interaction with psychological (individual and relational) and environmental factors can translate into sustainable and effective lifelong healthy habits. Effective solutions that foster healthy lifestyles from an early age can thus have significant long-term impact(s). Personalized disease risk prediction to guide interventions promoting healthy lifestyle from childhood to old age therefore has a key role in abating the unprecedented pandemic of communicable and non-communicable diseases (NCDs) in adulthood. This has been recognized as a priority by the European Commission, which launched the Healthier together – EU NCD initiative in December 2021, supporting actions to reduce the burden of major NCDs and improve citizens' health and well-being, with a special attention to reducing health inequalities (https://health.ec.europa.eu/non-communicable-diseases/).

Unlocking the full potential for exploitation of real-world and research-based clinical, psychological, socioeconomic and environmental data to investigate disease risk prediction from the early stages of life through innovative digital solutions will allow to effectively move the dial in health promotion, through a holistic and coordinated approach to prevention and care.

The aim of the SP2 - WP5 is to establish an epidemiological framework for optimal exploitation of real-world and research-based biological, clinical, psychological, social, and environmental data, to enable personalised multidimensional disease-specific risk prediction, identification of early markers of disease, leading to more effective preventive strategies throughout the lifespan (Figure 1.1).

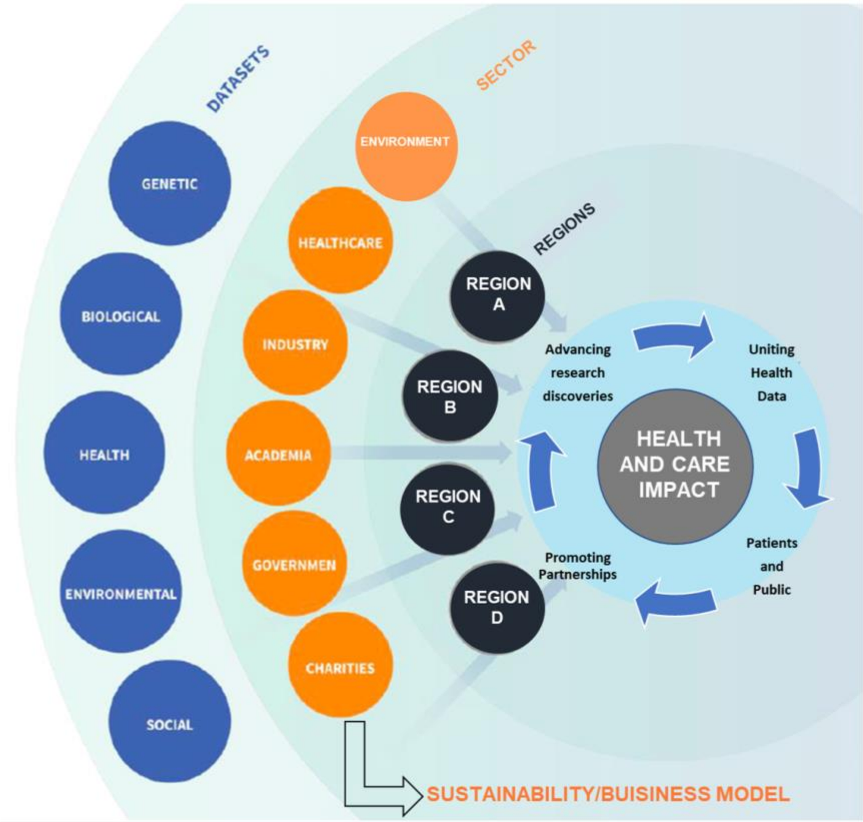


Figure 1.1 - SP2-WP5 architecture/strategy (Adapted from HDR UK)

1.1. Objectives of the deliverable

In this context the aim of the second deliverable (D5.2) is to define the plan and analyses models for the study of disease risk factors of disease development or early markers of disease across the pilot projects included in WP5. Obtaining more precise risk estimates of disease development through advanced predictive modeling will inform targeted and more effective prevention strategies.

1.1.1. Pilots SP2 - WP5

The ongoing pilot projects within WP5 are listed in Table 1.1.1.1.

Table 1.1.1.1. Approved pilot SP2-WP5

Task	ID-PILOT	PILOT title
T5.2	S2-WP5-T5.2-P01	A multidimensional integrated digital prevention approach for healthy elder people
T5.3	S2-WP5-T5.3-P01	Clinical, cognitive and neuropsychological markers of healthy ageing
T5.3	S2-WP5-T5.3-P02	A multidimensional and multimodal analysis of visuospatial and socio-relational abilities in typical and atypical development
T5.3	S2-WP5-T5.3-P03	Early markers and correlates of learning disorders and ADHD
T5.3	S2-WP5-T5.3-P04	Unlocking the full potential for uniting, improving and using electronic health data: innovative pathways from health data research to better health care for all, from prenatal life into adulthood - Use case: “The burden and risk factors of obesity in children and adolescents in Italy”

2. Pilot studies

2.1. A multidimensional integrated digital prevention approach for healthy elder people

Mario Barbagallo, Nicola Veronese (UNIPA)

2.1.1. Pilot framework

From a public health perspective, geriatric syndromes (such as depression, cognitive impairment, osteoporosis) are particularly present among older people in our project (Sanford, 2020). Whilst reliable epidemiological data regarding elder abuse are not available, it is estimated that every three seconds one older person could be affected by dementia or cognitive impairment (Kukull, 2020). Dementia and cognitive impairment in general may have tremendous public health costs: it is reported that our Health Care System spends 10-12 billion euro per year for these conditions. Moreover, the consequences of osteoporosis are of importance: hip fractures, for example, are associated with costs of more than 9 billion Euros per year in Italy. Finally, depression is another common condition since it was reported that 16.8% of people older than 85 years are affected by this condition.

Considering that the early recognition of these conditions can introduce more effective interventions (pharmacological and non-pharmacological, including vaccinations suggested for older people), our project will approach the prevention of these aspects from a primary prevention perspective, particularly in those older people that appear substantially healthy, but can develop relevant geriatric problems.

Using the ICOPE (Integrated Care for Older People) tool, a tool proposed and used by the World Health Organization to investigate the intrinsic capacity, we will discover the domains most commonly affected and suggest interventions to slow or prevent the transitions from a healthy condition to a relevant clinical one. In particular, we will integrate the information derived from the ICOPE tool to the most important sources of health and non-health information such as hospital medical records, general practitioners' archives, etc. ICOPE is an easy tool formed by simple yes/no questions on six different domains, important for older people.ICOPE is available in several languages at this website: https://www.who.int/publications/i/item/WHO-FWC-ALC-19.1. ICOPE is reported in the Appendix Material as Appendix 2. Briefly the ICOPE screening tool investigates the domain of cognitive decline, limited mobility, malnutrition, visual impairment, hearing loss, and depressive symptoms. This tool integrates simple yes/no questions and some tests (audio and visual) that investigate specific domains.

2.1.2. Features/factors

The proposed experimental approach will have several objectives. First, it will realise a feasibility study for a digital version of ICOPE in a sample of Italian healthy elder individuals. Second, we will explore the possibility of implementing interventions (pharmacological and non-pharmacological) for the primary prevention of depression, cognitive impairment, elder abuse and osteoporosis/fractures. Lasty, we will develop an innovative data mining approach to define primary preventive interventions and to promote healthy ageing.

2.1.3. Statistical analysis plan/model(s)

All the analyses will be done using SPSS 26.0 or STATA 14.0. Since our pilot study will try to interconnect information deriving from different sources, we will also approach the data using artificial intelligence (AI) methods. Machine learning techniques will be used, specifying as descriptors the epidemiological, biological, clinical data and risk factors collected from the healthcare infrastructure with the aim of providing a probability of developing loss of intrinsic capacity on the healthy elderly. For specific types of data, such as imaging, convolutional neural networks will be adopted. The system will include a computerised decision support system capable of providing alerts in the event of the presence of conditions that may interfere with healthy ageing, such as depression, cognitive impairment, osteoporosis/falls, elder abuse integrating the information of the ICOPE app and electronic health records.
We will integrate the information derived from ICOPE with relevant sources of health and non-health information such as hospital medical records, general practitioner archives, etc…

2.1.4. Expected challenges and possible solutions

The following table will present an overview of the anticipated challenges and propose possible solutions that could mitigate the impact of these challenges (Table 2.1.4.1).

Table 2.1.4.1 Expected challenges and proposed solutions

Topic	Challenge	Proposed solution
Copyright App	ICOPE is a World Health Organization (WHO) product. The topic of the copyright of the official app is of importance from a legal point of view.	Some French researchers have produced an app in French that is not authorised by WHO, but accepted at international level. We will contact, again, WHO offices to solve this problem, if present.
Integration of information	Medical and social information about the outcomes must be integrated with the information derived from ICOPE	To prepare a RedCap interface or similar to overcome the integration of information derived from different sources.

2.2. Clinical, cognitive and neuropsychological markers of healthy ageing

Giovanna Mioni, Andrea Zangrossi (UNIPD)

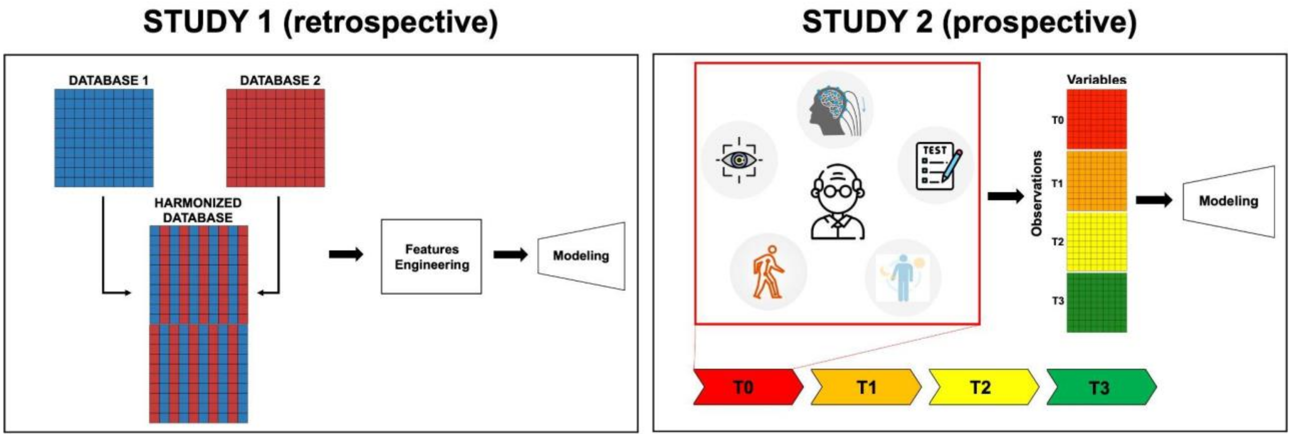
2.2.1. Pilot framework

The global population is living longer, with the percentage of individuals over 60 years old expected to increase from 12% in 2015 to 22% by 2050 (World Health Organization, 2015). Maintaining a healthy ageing is a pressing worldwide challenge, and the possibilities to monitor the health status in a fast-evolving, digitalized society are increasing rapidly. It has been recently estimated that among the causes of death in the ageing population, 13.6% can be attributed to dementia (Stokes et al., 2020), which refers to a group of neurodegenerative diseases characterized by impairment in different cognitive domains (e.g., memory, attention, language, social abilities) which have a devasting impact on the quality of life of an individual, making him/her in need of care for many daily activities. At the societal level, this has important implications also on government finances and policies. Suffice it to say that the total societal costs of dementia in Europe were estimated to 103 billions € in 2009, but the demographic forecast of costs will result in an increase to over 250 billions € in 2030. To date, effective pharmacological treatments for the different kinds of dementias are lacking. For instance, a particularly interesting story regards aducanumab, a proposed pharmacological treatment for Alzheimer's disease that was presented as a new hope for patients and approved by the FDA in 2021, has been recently retired from the market due to controversial scientific findings (Heidebrink and Paulson, 2024). This underlines the need for methods for the early detection of neurodegenerative diseases in the elderly population. Indeed, early patient management, is currently the best approach to deal with neurodegenerative pathologies in ageing (Kivipelto et al., 2018; van der Flier et al., 2023). Among the first clinical signs of dementia, cognitive impairment is often the most prominent. Existing methods for detecting cognitive decline are mainly effective when symptoms have already reached a clinical manifestation (e.g., a patient referring to a neurologist for memory complaints), nevertheless, they are not useful for the monitoring of individuals in the preclinical stages, when the cognitive performance shows only subtle alterations or does not show impairment at all. For this reason, a global challenge is to identify early markers of dementia that can be implemented in clinical practice to detect subtle dementia-related alterations before the clinical manifestations of symptoms.

Our work will contribute to this goal in two ways (see Figure 2.2.1.1):

- In Study 1 (retrospective), we will extract and test potentially sensible indices to detect early manifestation of cognitive decline from existing databases of medical, psychological, social and demographic variables (e.g., Health and Retirement Study, https://hrs.isr.umich.edu/, see Fisher et al., 2018; UK Biobank, https://www.ukbiobank.ac.uk/, see Sudlow et al., 2015).

- In Study 2 (prospective), we will run a prospective multimodal data collection with data from different domains (brain activity, oculomotor dynamics, cognition, gait analysis, and motion patterns) on a cohort of healthy elders. In each timepoint, the different data recording will be divided in two experimental sessions with a time interval of 1 week to allow for a 7-days actigraphic registration.



Database 1 - Health and Retirement Study, database 2 - UK Biobank

Figure 2.2.1.1 Description of the Pilot Study Structure

2.2.2. Features/factors

A promising approach for the identification of early markers of neurodegenerative pathology is the adoption of a multimodal approach (Wang et al., 2024), i.e., an approach based on data from different sources. In line with this idea, we will investigate multivariate patterns of information with the potential to be predictive of cognitive decline. To this end, both in Study 1 and Study 2, we will focus on variables falling within the following categories:

- Demographic and lifestyle (e.g., sex, age, education, food habits)
- Cognitive and or clinical (e.g., test scores)
- Social information (e.g., occupation)

Furthermore, in Study 2, in addition to the abovementioned variables, we will have the opportunity to collect data from different sources including:

- Brain activity (i.e., EEG data)
- Oculomotor dynamics (e.g., eye movements and pupil dilation)
- Motion dynamics (e.g., 7-days actigraphic recording, and gait analysis).

2.2.3. Statistical analysis plan/model(s)

Statistical analyses in Study 1 and Study 2 will be run by means of R (R Core Team, 2022) and Matlab (Mathworks) softwares and in-house code. Specifically, the preprocessing of EEG data will be carried out through specific packages of Matlab software, while statistical models will be mainly built with R.

In the first step we will work on features engineering to identify the most effective features starting from raw data (both for Study 1 and Study 2). This step will also include studying the correlation among variables and, if appropriate, running data reduction techniques, such as Principal Components Analysis (PCA).

Then, in a second step we will move to predictive modeling. First, we will investigate variables showing sensitivity to changes in time, in a univariate perspective, through linear regression models and mixed-effects models (R library lme4). Then, we will move towards a multivariate perspective by investigating patterns of features that can identify subtle changes in time potentially predicting cognitive decline. To this end we will start from linear methods (e.g., logistic regression) and then we will test more complex methods such as supervised machine learning models (e.g., SVM and SVR, through R library e1071) and deep learning models (R library Keras). This is in line with recent studies showing that multivariate and multimodal approaches are the most effective for the prediction of neurodegeneration (Wang et al., 2024).

2.2.4. Expected challenges and possible solutions

An overview of some possible challenges and proposed mitigation strategies is presented in the table below (Table 2.2.4.1).

Table 2.2.4.1 Expected challenges and proposed solutions

Topic	Challenge	Proposed solution
Sample	Due to the longitudinal nature of Study 2 (i.e., follow ups every 6 months) we may not be able to reach the desired sample size of healthy elders	Enlarge the recruitment to young adults, with the goal of identifying a normative sample to quantify the deviation of healthy elders from the trajectory of young adults
Devices	Breaking or malfunctioning of the devices/softwares (Study 2)	Opening negotiations with producers to replace the broken parts/sensors
Data	Inhomogeneity of data between databases (Study 1)	Selection of variables or creation of meta-variables (e.g., components) based on the whole set of variables (between databases)
Data	Need for higher computational performance than expected to deal with the amount of data (Study 1 and Study 2)	Purchase of new computational units or storage
AIMS	The probability that subjects recruited (Study 2) will show cognitive decline within DARE project time is very low. Thus we might not be able to predict the onset of dementia or mild cognitive impairment	The prediction will focus on subtle changes in cognitive performance (i.e., not necessarily indicating a formal impairment)

2.3. A multidimensional and multimodal analysis of visuospatial and socio-relational abilities in typical and atypical development

Irene Mammarella, Tiziana Pozzoli, Silvia Lanfranchi, Sara Onnivello, Camilla Orefice (UNIPD)

2.3.1. Pilot framework

Based on the trans-diagnostic approach (Jaffee, 2022; Astle et al., 2022), psychological development can be understood as changes in dimensional variability among individuals. This pilot study will focus on two key psychological dimensions: cognitive processing (primarily visuospatial skills) and socio-relational skills in children and adolescents with typical and atypical development.

First, a digital battery will be developed to measure visuospatial and socio-relational domains, and it will be administered to typically developing children aged 8 to 14. The measures will be based on the latest research findings (e.g., Hodgkiss et al., 2021; Prinstein & Giletta, 2020; Uttal et al., 2013) and will provide insights into the developmental changes in these domains for typically developing children. For reaching this aim, children with typical development will be recruited at their school. Hence, contacts with primary and middle schools will be established and the project will be presented to school majors and teachers.

Second, to enhance our understanding of the neuropsychological profiles associated with various neurodevelopmental disorders, the pilot will assess the strengths and weaknesses of individuals with Autism Spectrum Disorder (ASD), Developmental Coordination Disorder (DCD), Nonverbal Learning Disability (NLD), and genetic syndromes such as Down syndrome, which are characterized by specific challenges in visuospatial processing (APA, 2013; Semrud-Clikeman et al., 2010, 2014; Tsai et al., 2008) and socio-relational skills (Kwan et al., 2020; Musetti et al., 2019). Children with Neurodevelopmental Disorders will be recruited via specialized clinics and hospitals in the Veneto region. As inclusion criteria, children should have a previous independent clinical diagnosis, according to DSM 5 (APA, 2013) or ICD-10 (World Health Organization [WHO], 1993) criteria, made by private psychologists or child psychiatrists at clinical specialized centers. As exclusion criteria, none of the children should have neurological, or other psychopathological conditions.

Finally, our work will contribute to the overarching goal of WP 5 by identifying and integrating psychological health determinants to Pedianet and other data platforms as appropriate. The final goal is to understand how cognitive and socio-relational skills, viewed as changes in dimensional variability between children and adolescents, may help to increase our knowledge about risk factors of neurodevelopmental disorders.

2.3.2. Features/factors

The primary objective is to evaluate the visuospatial and socio-relational domains based on the latest research (Hodgkiss et al., 2021; Uttal et al., 2013) and to illuminate the developmental changes in these domains in typically developing children. This will involve administering a digital battery of tests on portable tablets to children aged 8 to 14.

We will consider several variables, including demographic information (e.g., the child's age, any previous diagnoses, parental education) and early developmental information (e.g., APGAR scores, medical or genetic conditions, birth weight).

For visuospatial abilities, we will assess visual perception, mental rotation, and perspective-taking. Visuo-constructive abilities will be evaluated using the Rey-Osterrieth Complex Figure, while fine motor skills will be measured with the Purdue Pegboard Test. Socio-relational abilities will be assessed using verbal and pictorial theory of mind tasks. In all measures, accuracy of participants will be assessed and analyzed.

For children with Down syndrome (aged 3-16), we will adapt the measures to account for their specific characteristics and intellectual disabilities.

2.3.3. Statistical analysis plan/model(s)

All the analysis within the present project will be run using R (R Core Team, 2022), along with the appropriate libraries.

Study 1 will be conducted involving children without any known diagnosis, aged 8 to 14 years old.

First of all, the features of the experimental tasks used for the measurement of visuospatial skills will be assessed. To do so, either explorative or confirmative factorial analysis will be conducted. Moreover, several simple logistic models (SLM), paradigm around the item response theory (IRT), will be used to assess the difficulty of such tasks.

In order to assess the developmental trajectories of the visuospatial abilities, regression models will be built, entering the score at each test as dependent variable and age (and/or school grade) as predictor. In this phase, not only accuracy but also reaction times will be analyzed.

Finally, the pattern of relationships between the visuospatial and socio-relational domains will be analyzed. To do so, correlational patterns will be explored and, based on these results, further analysis will be run (i.e., regression analysis, structural equation models, factorial analysis).

Study 2 will revolve around visuospatial and socio-relational skills in the comparison between typical and atypical development.

As a first step, aiming to highlight statistically significant differences between groups, descriptive statistics for each group will be computed; in addition, several univariate analyses of variance will be run, also considering measures of effect sizes. Bonferroni's correction for multiple comparisons will be used, when appropriate.

Finally, correlation matrices and regression models, built taking into account the specific distributions of our data, will be used to assess the relationships between visuospatial and socio-relational measures, as concerns both accuracy and reaction times. Such models will be run entering the groups as factors.

2.3.4. Expected challenges and possible solutions

The following table will present an overview of the anticipated challenges and propose possible solutions that could mitigate the impact of these challenges (Table 2.3.4.1).

Table 2.3.4.1 Expected challenges and proposed solutions

Topic	Challenge	Proposed solution
Sample	Different neurodevelopmental conditions may be difficult to compare in terms of statistical analyses	Children with neurodevelopmental disorders (Autism Spectrum Disorder, Developmental Coordination Disorder, Nonverbal Learning Disability) will be matched with non-diagnosed children matched for age, biological sex, and verbal IQ. Individuals with Down syndrome will be matched with non-diagnosed children for biological sex and mental age
Measures	Based on the characteristics of the groups included (neurodevelopmental and genetic conditions) some measures cannot be identical	Measures of visuospatial abilities, as well as of socio-relational skills will be adapted to the mental age of the individuals (in particular for Down syndrome), and if available, standardised materials will be used.

2.4. Early markers and correlates of learning disorders and ADHD

Barbara Carretti, Elisa Cainelli (UNIPD)

Over the last decade, a significant increase in the incidence of neurodevelopmental disorders in childhood has been reported (e.g. Cortese et al. 2023; Grigorenko et al. 2022). Among them, specific learning disabilities (SLD - reading, writing, and mathematics disabilities) and attention/hyperactivity disorder (ADHD) are the most frequently diagnosed developmental disorders in children (Cainelli & Bisiacchi, 2023). Given their widespread occurrence, these disorders significantly impact learning and academic achievements with cascading effects on society and economics.

Neurodevelopment is a slow process; problems may take years to emerge and to establish. Therefore, currently, the interest of the scientific world is turning towards indicators of risk and precursor skills for early identification of children at risk of learning difficulties (Mercugliano et al., 2024) and ADHD (e.g. Oerbeck et al., 2020). Early years are indeed a critical developmental period during which the building blocks for later success are laid, and the social, behavioural, and neuropsychological skills necessary for academic success are acquired. From this perspective, investigating prerequisites and risk indicators makes it possible to identify children with difficulties at an early stage; furthermore, performing long-term longitudinal studies allows us to understand the developmental trajectories.

2.4.1. Pilot framework

This pilot seeks to address this complex question from multiple levels of analysis, spanning different studies to highlight: i) early risk markers, ii) baseline prerequisites, and iii) trends over time, through both prospective and retrospective studies, of two of the most common neurodevelopmental disorders: SLD and ADHD.

- Retrospective study: This study will be conducted using a retrospective research analysis of the Pedianet database (see deliverable 5.1 “Concept of models and paths and mapping of eligible data sources”) and will be carried out in collaboration with pediatricians. The aim is to estimate the incidence and prevalence of ADHD and SLDs in children aged 5-14 in Italy from 2004 to 2023, describe the demographic, social, and clinical characteristics of children with ADHD/SLD from 2004 to 2023, and evaluate the associated risk factors with ADHD/SLDs. The variable selection and the extraction procedures of this study are ongoing.

- Prospective study 1: This study will prospectively evaluate the presence of SLD and ADHD and investigate the neuropsychological and psychophysiological profile of children aged 8-13 years with a history of prematurity or perinatal asphyxia. These pre-perinatal conditions have been chosen given the high incidence of ADHD, SLD, and other neurodevelopmental disorders in these populations (e.g. Radtke et al., 2023). The study leans on a previous study in which these children were recruited at birth, followed until 6 years, and the clinical/medical records, in the neonatal period and further, collected. The ethical committee for this study has been submitted. The study will start as soon as the ethical approval is available and will end in December 2024.

- Prospective study 2: This study focuses on neurophysiological and neuropsychological correlates of ADHD. It includes data collection within schools and in clinical centers in children with a diagnosis of ADHD, in their siblings, and healthy controls. The ethical approval for this study has already been obtained, and the study has started and will continue throughout 2025.

- Prospective study 3: This study focuses on the learning and cognitive trajectories of children from preschool to the last year of primary school. The aim is to identify and analyze, in the general population, the preschool learning prerequisites, the learning trajectory over time, the transversal cognitive functions involved, and patterns of activity/movements and heart rate. The ethical committee for this study has been submitted. The study will start in autumn 2024 and will end in spring 2026.

2.4.2. Features/factors

In this research, the "features" or "factors" are the various attributes that i) characterize SLD and ADHD (at cognitive, behavioural and neurophysiological/psychophysiological level) and ii) influence and contribute to this condition (risk factors, prerequisites, factors influencing the course). The pilot was subdivided into sub-studies to investigate the different specific aspects involved and to allow for a more complete and integrated view at the conclusion of the research. Given that each sub-study will focus on specific aspects and it is not possible here to deal specifically with each study, we list below all the variables that will be considered, taking the sub-studies that make up the pilot as a whole.

1. <u>Features/factors characterizing</u>: SLD and ADHD are well-defined conditions described in the diagnostic and statistical manual of mental disorders (DSM-5, APA, 2013) and with national guidelines addressing the diagnosis. SLD involves the administration of specific tests that assess the child's abilities in the domains of reading, writing, and calculation. The administration of tests and questionnaires assessing skills and environmental aspects strictly connected is also indicated and will be performed. ADHD requests clinical evaluations and tests administration to the child and questionnaires to the parents/caregivers and the teachers to evaluate the behaviour in different contexts. Knowledge about the neural correlates of these conditions is relatively less extensive. Among different neurobiological aspects, our research will focus, in particular, on the neurophysiological/psychophysiological correlates: electroencephalography and heart rate variability. Furthermore, this pilot aims to objectively measure characteristic symptoms, such as hyperactivity, by analyzing movements through portable sensors (smartwatches).

2. <u>Features/factors contributing</u>: Research on factors influencing the course and emergence of SLD and ADHD has identified some key factors that need to be explored to understand these neurodevelopmental disorders better.
    - Precursors: all those skills, processes, or mechanisms that precede and underlie the acquisition of another skill, in this case, learning. Precursors that will be investigated will be general (verbal and visuospatial memory, rapid naming, processing speed and attention, reasoning, behavioral regulation) and specific (phonological awareness, letter recognition, letter writing, language, quantity recognition, number recognition, and writing) prerequisites in preschool (e.g. Mercugliano et al., 2024).
    - Cognitive transversal abilities. Learning is a complex process involving general and domain-specific skills. Their acquisition is supported by several more or less specific

skills that support and influence the acquisition and consolidation of learning over time. Transversal cognitive processes that will be assessed: short-term memory, rapid naming, processing speed and attention, vocabulary, non-verbal reasoning, and behavioural regulation.

- Risk factors. There are many known risk factors for SLD and ADHD, but also factors that could modify the course of the disorder positively or negatively. Factors related to pregnancy and childbirth (such as premature birth and low weight), perinatal clinical examinations (such as Apgar, pH, and neuroimaging), maternal lifestyles, health status of the child in the first years of life and medications, sex and familiarity will be considered.

2.4.3. Statistical analysis plan/model(s)

The pilot's substudies have very different study designs (retrospective vs. prospective, case-control vs. correlational, cross-sectional vs. longitudinal), and therefore, the statistical analyses envisaged are extensive and highly differentiated.

- Retrospective study: In this study, the incidence rate of children with SLD and ADHD in the Italian population will first be calculated. Thereafter, a multivariate Cox proportional hazard regression model analysis will be applied to evaluate the association with possible risk factors.
- Prospective study 1: In this study, the incidence rate of children with SLD and ADHD in the population of children with a history of prematurity and perinatal asphyxia will first be calculated. Thereafter, a multivariate regression model analysis will be applied to evaluate the association with possible neonatal risk factors and neuropsychological and psychophysiological (heart rate variability) profiles.
- Prospective study 2: This study will compare the neuropsychological and neurophysiological (electroencephalogram) profile of children with ADHD with that of siblings and controls. Generalized mixed models will be employed to investigate the effect of both between (group) and within (experimental task conditions) factors as predictors. Both behavioural (reaction times and accuracy) and neurophysiological measures (event-related potentials and time-resolved oscillatory activity) collected during a computerized cognitive control task will be considered dependent variables.

- Prospective study 3: In this study, the development over time of abilities in the domains of reading, writing, and mathematics will be assessed using a repeated measures analysis. The role of prerequisites and transversal cognitive functions in influencing school learning abilities will also be assessed using a linear regression model.

All the analyses within the present project will be run using R (R Core Team, 2022) and the appropriate libraries.

2.4.4. Expected challenges and possible solutions

The following table presents an overview of the anticipated challenges and proposes possible solutions to mitigate their impact (Table 2.4.4.1)

Table 2.4.4.1 Expected challenges and proposed solutions

Topic	Challenge	Proposed solution
Sample	Dropout in longitudinal studies	We will work on two levels: 1. making clear the significance of the study from an applied point of view when presenting the study; 2. using statistical methods for missing data
Measures	Given the type of population studied (in particular ADHD children) and the need to complete more then one session of data collection, it might be difficult to complete the data collection of some participants	The solutions are similar to those from the previous challenge. Additionally, we will include monetary rewards for participation in the studies.
Timeline	Given the studies depend on family's and schools/clinical centers' agreement, some delays in the timeline may occur	The various steps of the studies were planned so as to foresee possible delays or temporary blockages in the activity

2.5. Unlocking the full potential for uniting, improving and using electronic health data: innovative pathways from health data research to better health care for all, from prenatal life into adulthood - Use case: "The burden and risk factors of obesity in children and adolescents in Italy"

Veronica Casotto, Silvia Bressan (UNIPD)

2.5.1. Pilot framework

The pilot project, entitled "The Burden and Risk Factors of Obesity in Children and Adolescents in Italy," was designed to assess the feasibility, preliminary effectiveness, and potential implementation issues of uniting, improving, and using electronic health data from existing real-world, environmental or administrative sources. The ultimate goal of this project is to identify innovative pathways from health data research that can lead to better health care for all, from prenatal life into adulthood.

In order to reach these goals, several approaches will be implemented. (i) Using data on maternal health and environmental factors, researchers can develop models to predict potential health risks for babies before birth. This allows for early intervention and improved prenatal care. (ii) Analyzing electronic health records can allow for (a) earlier detection of chronic diseases like asthma or diabetes in children, (b) tailor educational programs to individual needs and identify children who might benefit from additional support or interventions, and (c) identify trends in disease prevalence and risk factors, allowing for targeted public health interventions and improved resource allocation. Robust data and security protocols will build trust in health data research. Leveraging artificial intelligence and machine learning can help extrapolate specific diagnoses not always coded in the same way and identify patterns and trends that humans might miss, leading to new discoveries and insights.

A micro-level approach dealing with data on healthcare pathways experienced by each beneficiary of the health system will be used. Because of its "patient-centered" nature, the approach of the project will be conducted following the basic rules of good clinical research practice. This means that a clear definition of the questions the study will have to answer, of the study design, of the methods for overcoming and taking into account the sources of uncertainty, and of the models for data analysis must be predefined in a protocol. The burden and risk factors of obesity in children and adolescents in Italy project will focus on:

- Dataset identification: (i) an electronic primary-care dataset derived from a systematic collection of information on the paediatric population (the identified data source was Pedianet: see deliverable 5.1 "Concept of models and paths and mapping of eligible data sources"), (ii) an environmental dataset derived from ARPAV dataset from ISTAT (http://dati.istat.it) and (iii) administrative data (see deliverable D5.1 entitled "Concept of models and paths and mapping of eligible data sources").

- Definition of the outcome: the primary outcome will be childhood obesity which will be defined based on several criteria. The most common is the Body Mass Index (BMI), which is calculated as weight in kilograms divided by the square of height in metres. For children, the BMI is interpreted with respect to age- and sex-specific growth curves - BMI z scores (De Onis et al., 2007; www.who.int, Canova et al., 2021). According to the World Health Organisation (WHO) criteria, a child is considered obese if their BMI is at or above the 95th percentile for their age and sex: obesity was defined as a z score >3 SDs above the mean for children aged ≤5 years and a z score >2 SDs for children aged >5 years (Valerio et al., 2018). Contrary to the Centers for Disease Control and Prevention (CDC) Growth Charts, identifying children as obese for BMI z score >2 values (https://www.cdc.gov). Single BMI measurements, as well as BMI trajectories over a specified follow-up period, will be used for analysis.

- Identification of variable of interest: there are several potential variables of interest that we can explore using the provided data sources to identify risk factors for childhood obesity: (i) from the electronic primary-care dataset, we will be able to assess demographic features (e.g., age, sex, region of resident, socioeconomic status identified from area deprivation index), anthropometric measurements, medical history (e.g., type 2 diabetes, fatty liver disease, others comorbidities), and delivery characteristics (e.g., gestational age at delivery, birth weight); (ii) from the environmental dataset, we will able to assess air quality based on gaseous (NO2, SO2, CO) and particulates (PM10 and PM2.5)(Appendix 1), and (iii) from administrative data we will able to assess exposures before and during pregnancy, as well as the medical history of the mother and several characteristics of pregnancy, delivery and birth. The most appropriate variable(s) of interest will depend on the specific research question.

2.5.2. Features/factors

In the field of childhood obesity research, the "features" or "factors" are the various attributes and influences that contribute to this condition. These include demographic characteristics such as age and sex, and behavioural aspects including dietary habits and physical activity levels. It is also important to consider the role of environmental, and socioeconomic factors. An understanding of these features is essential for the identification of children at risk, the development of effective prevention strategies, and the tailoring of interventions to individual needs.

The potential for electronic data collection on childhood obesity characteristics and contributing factors is significant due to the advances in computerisation that have been made. However, the immediate analysis of the collected data is not always guaranteed to be straightforward. Diaries kept by paediatricians, for example, are recorded with dedicated software that allows the systematic and continuous collection of data, but the variability of clinical data documentation and the lack of standardisation require specific analytical consideration and data pre-processing.

In the context of this study, the key variables of interest that can be extracted from the paediatric dataset are as follows:

- Demographic characteristics, such as biological sex and age
- Gestational variables, encompassing maternal health factors such as diabetes and smoking, birth weight, type of delivery (vaginal or caesarean), and infant nutrition.
- Comorbidities, including intellectual and physical chronic diseases and disabilities,
- Medications taken (occasional and chronic).

2.5.3. Data interoperability

The concept of interoperability has been around since 2004 with the “European Interoperability Framework for eGovernment Services” (European Commission. 2004). The interoperability is the “ability of information and communication technology (ICT) systems and of the business processes they support to exchange data and to enable the sharing of information and knowledge". The three key areas of interoperability are: 1) organisational Interoperability "[…] defining business goals, modelling business processes and bringing about the collaboration of administrations that wish to exchange information and may have different internal structures and processes", 2) semantic interoperability "[…] ensuring that the precise meaning of exchanged information is understandable by any other application that was not initially developed for this purpose" and 3) technical interoperability "[…] technical issues of linking computer systems and services". Interoperability is the ability to access, process, and integrate data from multiple sources for mapping, visualisation, and other forms of presentation and analysis. It allows finding, exploring, and understanding data. Essentially, it is the ability to 'connect' data from different sources to create a contextual and holistic picture for easier (sometimes automated) analysis, decision making and accountability.

In our context the interoperability results are essential to integrate several pieces of information into the electronic primary-care dataset from the other source of data presented above.

In particular, regarding the identification of variables of interest, we will use several data sources to answer our research questions.

i. The socioeconomic status will be identified from the area deprivation index (ADI). This information was integrated through the patient addresses, which were georeferenced and linked to the census block of each Italian municipality. The ADI is based on 5 items that describe social and material deprivation: (1) low education, (2) unemployment, (3) living on rent, (4) crowded households, and (5) single-parent families. The index is calculated as the sum of standardized indicators. The index was then categorized in quintiles based on the regional ADI level to ensure within-region appropriately represented categories.
ii. Environmental data on air pollution will be retreived from the Regional Agency for Environmental Prevention and Protection of the Veneto Region (ARPAV). In particular we will consider data on daily mean level for PM10 and PM2.5 and hourly series for Benzene, NO2 CO and SO2 from 33 monitoring stations located in the Veneto Region. We will also consider PM10 and PM2.5 daily series from the photochemical deterministic CAMx model which predict for all the Veneto Region on a regular grid (1x1 Km) daily mean level of the most important pollutants. Daily data on meteorological variables (in particular temperature and humidity) will be constructed starting from hourly data made available from ARPAV. A

Geographicl Information System (GIS) will be constructed with all this information (location of monitors, ARPA grid of the deterministic model, daily level of pollutants) in Q-GIS.

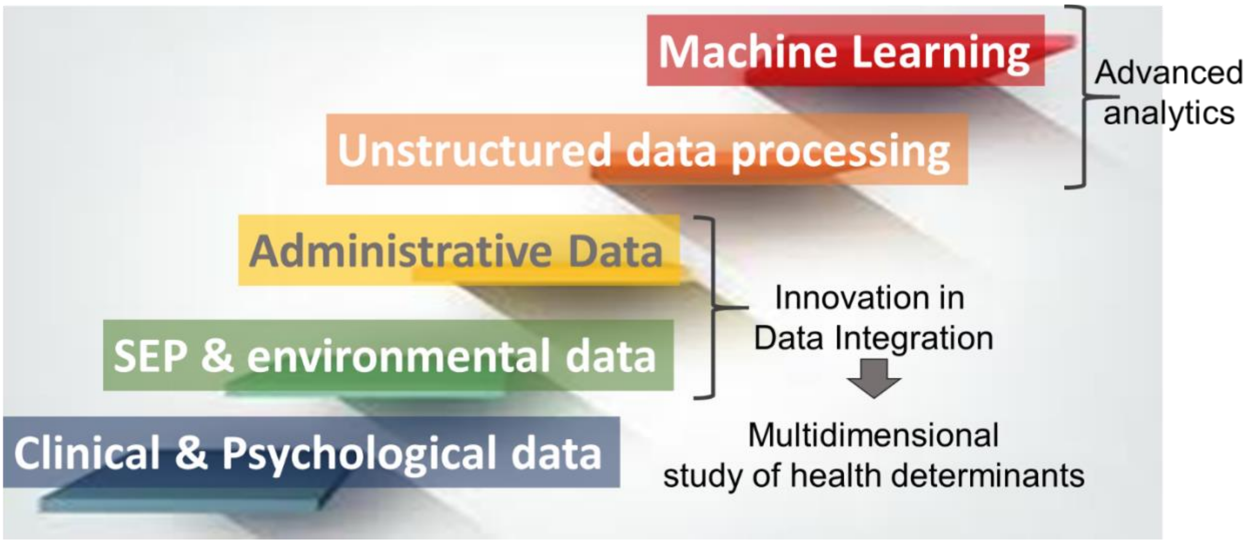
iii. The administrative databases record demographic and administrative data for all beneficiaries of the National Health Service and their health care use. Of our interest will be (i) the hospital discharges registry, which reports all diagnoses released from public or private hospitals, and (ii) the Certificates of Delivery Assistance (CeDAP), which provides detailed information on pregnancy, childbirth, and child presentation at delivery. In Veneto, since 2019, Pedianet data have been integrated with data from the Electronic Health Record (EHR) of the Veneto Region. This has been an important resource that allows us to evaluate the hospital discharges of a child who approves the consent of the family paediatrician to download these data. Moreover, a step forward towards interoperability for integrating different other data sources, such as CedAP, would be very important.

iv. Omics data – as a last step in data integration from different data sources we will endeavour to test the integration, as a feasibility pilot project, of omics data from a newly established biobank within PEDIANET.

2.5.4. Statistical analysis plan/model(s)

In this section, we delineate the statistical analysis plan and the models that will be employed to assess the data.

Our focus is on childhood obesity as the outcome variable, and our primary objective is to investigate the relationships between childhood obesity and potential risk factors, including clinical variables, the deprivation index, and air pollution data. The steps for identifying health risk factors are described in Figure 2.5.4.1.



(SEP: Socio Economic Position)

Figure 2.5.4.1. Steps for identifying health risk factors

We will also employ several study protocols to: (1) identify the known risk factors of childhood obesity and assess which are still open in this field and (2) evaluate the comprehensive understanding of the relationships between such risk factors and the onset of childhood obesity.

For aim (1), a systematic literature review (umbrella review) is being conducted to assess the risk factors associated with the development of pediatric obesity. This research is conducted comprehensively, aiming to identify the widest range of biological, social, economic, and other factors related to its development.

The initial stages of screening the identified literature were carried out with the support of an innovative artificial intelligence platform, ASReview (https://asreview.nl/about), which assists the researcher in making decisions regarding the paper selection. More precisely, it uses active learning to influence the order of articles based on relevance for the inclusion process, and when many articles (around 150 papers) are excluded in a row, it can be assumed that the articles listed after them can be labelled as irrelevant.

The results of this investigation will allow for the identification of current knowledge gaps in this field, useful for targeting new studies, and for the optimal direction of research to be conducted on available datasets for a multidimensional study of risk factors for pediatric obesity integrating the identified data sources.

For aim (2) several protocols will be implemented focus on:

A. Data Preparation. The first step of this study involves the meticulous selection of variables that will serve as indicators of potential risk factors for childhood obesity. We will use large language models (LLMs) to extract information from the free text and integrate and enrich data derived from the structured fields. These models will allow us to identify and integrate additional relevant information that may be missing from the encoded fields, ensuring a more complete dataset. By leveraging the advanced natural language processing capabilities of LLMs, we can extract insights and details from clinical notes. This integrated approach will improve the completeness and depth of our analysis, leading to more accurate and insightful results. The completeness of the chosen variables will be rigorously assessed. Missing data is a common challenge in research, and neglecting it can introduce bias. We will employ a two-pronged approach to address missing data: (1) Missing Information Due to Non-Reporting: in cases where data (e.g., gestational age or birth weight) is simply not reported, we will implement well-established predictive models to impute the missing values (Cantarutti et al., 2022). This imputation will be conducted only when necessary and with careful consideration of the specific model's assumptions and limitations. (2) Missing Data Due to Uncoded Diagnoses: approximately 30% of diagnoses may lack the standardized ID-9-CM code. To address this challenge, we will leverage cutting-edge Artificial Intelligence (AI) and Machine Learning (ML) techniques, including Large Language Models (LLMs). These advanced tools will be trained on the existing coded diagnoses to identify patterns and predict the most likely missing codes. This process will be implemented cautiously and with a focus on maintaining data accuracy. By carefully selecting relevant variables and employing a comprehensive strategy to handle missing data, this study aims to establish robust and reliable risk factors for childhood obesity.

B. Identification of the outcome of interest. The primary outcome variable of interest in this study is childhood obesity. This will be assessed using two complementary approaches: (1) *incident measure*: we will classify childhood obesity based on established World Health Organization (WHO) criteria. This approach will identify children who meet the criteria for obesity at a specific point in time. (2) *Longitudinal BMI Trajectories*: to gain a deeper understanding of weight patterns over time, we will employ a robust statistical technique called Group-Based Trajectory Modelling (GBTM), also known as Latent Class Growth Analysis (LCGA). This semi-parametric finite mixture model is specifically designed for analyzing longitudinal data. It allows us to identify distinct subgroups within the study population. Each subgroup will be characterized by a unique trajectory of Body Mass Index (BMI) z-scores across the study period. This approach will reveal how BMI patterns differ among subgroups of children, providing valuable insights into the development of childhood obesity. By utilizing both incident and longitudinal measures, this study will capture a comprehensive picture of childhood obesity prevalence and its trajectories within the study population.

C. Descriptive Analysis. The descriptive analysis aims to gain a better understanding of the available variables. Trends, correlations, and possible anomalies in the data are being explored in order to obtain a clear and comprehensive overview of the situation under consideration. This descriptive analysis phase is crucial for identifying the underlying patterns and providing a solid foundation for the development of subsequent statistical models. The information that emerges from this analysis will be critical to the construction of accurate and meaningful predictive models. The descriptive analyses employed include the calculation of averages, standard deviation analysis, distribution analysis (through the use of graphs or diagrams to identify any patterns or anomalies), frequency analysis (to understand the distribution of the data), and correlation analysis (to assess whether there is a relationship between two or more variables to ascertain whether there is an association between them).

D. Modelling. The modelling process involves a number of steps, including the selection of variables, the choice of model type, the estimation of parameters and the validation of the model. Selection of variables: the variables are selected based on theoretical assumptions, preliminary analysis and dimensionality reduction techniques, such as factor or principal component analysis. The choice of model depends on whether the outcome variable is dichotomous or categorical. We will implement a mixed-effect Cox proportional-hazards model to estimate the Hazard Ratio (HR) and 95% Confidence Interval (CI) for the association between exposures of interest and child obesity between 24 months to 14 years of age considering the family paediatricians as a random factor. The proportional hazard assumption for the time-fixed covariates will be tested using Schoenfeld residuals (Grambsch et al., 1994). Follow-up will began after two years of age and ended with the last anthropometric measure available by the end of the study, the completion of the 14th year of life, or the end of paediatric assistance, whichever comes first. Furthermore, to examine, for example, differences in the identified trajectories, we will test their associations with the considered set of sociodemographic/clinical factors using a multinomial logistic regression with a random intercept on the family paediatricians (Mlogit). We will estimate the odds ratios (ORs) for membership in each trajectory vs the reference group exhibiting normal growth, adjusting for several characteristics of the children. Each residential address of childrens will be georeference in the GIS. We then will apply Bayesian Kriging models to estimate exposure to air pollutants at the children's residential address. We will use data on air pollution from ARPAV monitoring stations and the output of the ARPA photochemical deterministic model. We will also test a comprehensive set of machine learning techniques to identify obesity predictors, including tree-based methods (e.g., decision trees, random forests), boosting algorithms (e.g., gradient boosting, XGBoost, LightGBM), support vector machines (SVMs), and regularized regression models (e.g., Lasso and Elastic-Net Regularized Generalized Linear Models). We will also test ensemble methods (e.g., SuperLearner) which combine multiple models to improve prediction accuracy and robustness by exploiting the strengths of different algorithms.To ensure the stability and generalizability of our models, we will use robust model validation techniques, such as k-fold cross-validation and bootstrap resampling, which help mitigate overfitting and provide reliable estimates of model performance. Model selection will involve a thorough evaluation of performance metrics, including accuracy, precision, recall, and area under the ROC curve (AUC), as well as model interpretability. To improve clinical interpretability, we will implement SHAP (SHapley Additive exPlanations), which assigns an importance value to each characteristic for individual predictions, clarifying the contribution of each factor to obesity risk. This approach allows us to identify key predictors while ensuring clinically meaningful results

E. Model Interpretation and utilisation. Finally, the model results will be interpreted based on the evaluation of the model coefficients and their statistical significance, as well as the analysis of the overall goodness of fit of the model.

2.5.5. Expected challenges and proposed solutions

The study of the relationship between childhood obesity, personal variables, area index deprivation and air pollution presents several challenges. It is therefore important to address these challenges in order to ensure the validity and reliability of the results. The following table will present an overview of the anticipated challenges and propose possible solutions that could mitigate the impact of these challenges (Table 2.5.5.1).

Table 2.5.5.1 Expected challenges and proposed solutions

Topic	Challenge	Proposed solution
Exploitation of free-text data fields	Free-text variables can be difficult to standardize and analyze statistically due to their unstructured nature	To develop coding algorithms that will enable the categorisation of free text responses into meaningful information that can be analysed quantitatively.
Data	The current unavailability of OMICS data	Establish a biobank for a predefined subset of pediatric patients to pre-test integration and exploration of omics data in multidimensional risk factor analysis

3. References

American Psychiatric Association, DSM-5 Task Force. (2013). Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). *American Psychiatric Publishing*, Inc..https://doi.org/10.1176/appi.books.9780890425596

Astle, D. E., Holmes, J., Kievit, R., & Gathercole, S. E. (2022). Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry,* 63(4), 397–417. https://doi.org/10.1111/jcpp.13481

Cainelli E, Bisiacchi P. (2023). Neurodevelopmental Disorders: Past, Present, and Future. *Children*. 10(1):31. https://doi.org/10.3390/children10010031

Canova C, Di Nisio A, Barbieri G, Russo F, Fletcher T, Batzella E, Dalla Zuanna T, Pitter G. PFAS Concentrations and Cardiometabolic Traits in Highly Exposed Children and Adolescents. *Int J Environ Res Public Health.* 2021 Dec 7;18(24):12881

Cantarutti A, Amidei CB, Stella Bonaugurio A, Rescigno P, Canova C. Early-life exposure to antibiotics and subsequent development of atopic dermatitis. *Expert Rev Clin Pharmacol* 2022 Jun;15(6):779-785. doi: 10.1080/17512433.2022.2092471. Epub 2022 Jun 21.

Cortese, S., Song, M., Farhat, L.C. et al. Incidence, prevalence, and global burden of ADHD from 1990 to 2019 across 204 countries: data, with critical re-analysis, from the Global Burden of Disease study. *Mol Psychiatry* 28, 4823–4830 (2023). https://doi.org/10.1038/s41380-023-02228-3

De Onis M. Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Org*. 2007;85: 660-667

European Commission. 2004. "European Interoperability Framework for eGovernment Services". Brussels: IDABC. https://ec.europa.eu/idabc/en/document/2319/5585.html

European Commission 2022. Non-communicable diseases. Retrieved from https://health.ec.europa.eu/non-communicable-diseases/

Fisher GG, Ryan LH. Overview of the Health and Retirement Study and Introduction to the Special Issue. *Work Aging Retire.* 2018 Jan;4(1):1-9. doi: 10.1093/workar/wax032. Epub 2017 Dec 19. PMID: 29423243; PMCID: PMC5798643. Website of the Health and Retirement Study: https://hrs.isr.umich.edu/

Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81 (3):515–526

Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *American Psychologist*, 75(1), 37–51. https://doi.org/10.1037/amp0000452

Health and Retirement Study, public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740*). Ann Arbor, MI,* (2024). https://hrs.isr.umich.edu/about

Heidebrink, J. L., & Paulson, H. L. (2024). Lessons Learned from Approval of Aducanumab for Alzheimer's Disease. *Annual review of medicine*, 75, 99–111. https://doi.org/10.1146/annurev-med-051022-043645

Hodgkiss, A., Gilligan-Lee, K. A., Thomas, M. S. C., Tolmie, A. K., & Farran, E. K. (2021). The developmental trajectories of spatial skills in middle childhood. *British Journal of Developmental Psychology*, 39(4), 566–583. https://doi.org/10.1111/bjdp.12380

https://www.who.int/nutgrowthdb/about/introduction/en/index4.html, accessed on June 2023 https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm accessed on June 2023

https://asreview.nl/about

Jaffee, S. R. (2022). Editorial: The times they are a-changin': paradigm shifts in child and adolescent psychology and psychiatry. *Journal of Child Psychology and Psychiatry,* 63(4), 357–359. https://doi.org/10.1111/jcpp.13604

Kivipelto, M., Mangialasche, F., & Ngandu, T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nature reviews. *Neurology,* 14(11), 653–666. https://doi.org/10.1038/s41582-018-0070-3

Kwan, C., Gitimoghaddam, M., & Collet, J.-P. (2020). Effects of Social Isolation and Loneliness in Children with Neurodevelopmental Disabilities: A Scoping Review. Brain Sciences, 10(11), *Articolo* 11. https://doi.org/10.3390/brainsci10110786

Kukull, W. A., & Ganguli, M. (2000). Epidemiology of dementia: concepts and overview. *Neurologic clinics*, 18(4), 923-949.

Mercugliano, A., Bigozzi, L., De Cunto, A., Graziani, D., Pecini, C., Usai, M. C., Ruffini, C. (2024). Which childhood predictive indices forecast reading and writing skills in school-age children: a systematic review*. Child Neuropsychology,* 1–36. https://doi.org/10.1080/09297049.2024.2347381

Musetti, A., Eboli, G., Cavallini, F., & Corsano, P. (2019). Social Relationships, Self-Esteem, and Loneliness in Adolescents with Learning Disabilities. *Clinical Neuropsychiatry,* 16(4), 165–172.

Oerbeck, B., Overgaard, K. R., Pripp, A. H., Reichborn-Kjennerud, T., Aase, H., & Zeiner, P. (2020). Early Predictors of ADHD: Evidence from a Prospective Birth Cohort. *Journal of Attention Disorders*, 24(12), 1685-1692. https://doi.org/10.1177/1087054717696765

Prinstein, M. J., & Giletta, M. (2020). Future Directions in Peer Relations Research. *Journal of Clinical Child & Adolescent Psychology,* 49(4), 556–572. https://doi.org/10.1080/15374416.2020.1756299

Radtke, B. M., Sajewicz-Radtke, U., Bieleninik, Ł., & Lipowska, M. (2023). Developmental learning disorders in children with prenatal/perinatal exposure to hypoxia: A systematic review protocol. *Plos one*, 18(10), e0293067.

R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Sanford, A. M., Morley, J. E., Berg-Weger, M., Lundy, J., Little, M. O., Leonard, K., & Malmstrom, T. K. (2020). High prevalence of geriatric syndromes in older adults. *PloS one*, 15(6), e0233857

Semrud-Clikeman, M., Walkowiak, J., Wilkinson, A., & Christopher, G. (2010). Neuropsychological Differences Among Children With Asperger Syndrome, Nonverbal Learning Disabilities, Attention Deficit Disorder, and Controls. *Developmental Neuropsychology*, 35(5), 582–600. https://doi.org/10.1080/87565641.2010.494747

Semrud-Clikeman, M., Fine, J. G., & Bledsoe, J. (2014). Comparison among children with children with autism spectrum disorder, nonverbal learning disorder and typically developing children on measures of executive functioning. *Journal of Autism and Developmental Disorders*, 44(2), 331–342. https://doi.org/10.1007/s10803-013-1871-2

Stiles, J., Akshoomoff, N. A., & Haist, F. (2020). Chapter 17—The development of visuospatial processing. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (A c. Di), Neural Circuit and Cognitive Development (Second Edition) (pp. 359–393). *Academic Press*. https://doi.org/10.1016/B978-0-12-814411-4.00017-2

Stokes, A. C., Weiss, J., Lundberg, D. J., Xie, W., Kim, J. K., Preston, S. H., & Crimmins, E. M. (2020). Estimates of the Association of Dementia With US Mortality Levels Using Linked Survey and

Mortality Records. *JAMA neurology*, 77(12), 1543–1550. https://doi.org/10.1001/jamaneurol.2020.2831

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine,* 12(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779. Website of the UK Biobank: https://www.ukbiobank.ac.uk/

Tsai, C.-L., Wilson, P. H., & Wu, S. K. (2008). Role of visual–perceptual skills (non-motor) in children with developmental coordination disorder. *Human Movement Science*, 27(4), 649–664. https://doi.org/10.1016/j.humov.2007.10.002

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402. https://doi.org/10.1037/a0028446

Valerio G, Maffeis C, Saggese G, et al. Diagnosis, treatment and prevention of pediatric obesity: consensus position statement of the Italian Society for Pediatric Endocrinology and Diabetology and the Italian Society of Pediatrics. *Ital J Pediatr.* 2018 Jul 31;44(1):88

van der Flier, W. M., de Vugt, M. E., Smets, E. M. A., Blom, M., & Teunissen, C. E. (2023). Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. *Nature aging,* 3(5), 494–505. https://doi.org/10.1038/s43587-023-00404-2

Wang, C., Tachimori, H., Yamaguchi, H., Sekiguchi, A., Li, Y., Yamashita, Y., & for Alzheimer’s Disease Neuroimaging Initiative (2024). A multimodal deep learning approach for the prediction of cognitive decline and its effectiveness in clinical trials for Alzheimer's disease. *Translational psychiatry,* 14(1), 105. https://doi.org/10.1038/s41398-024-02819-w

World Health Organization [WHO]. (1993). The ICD-10 classification of mental and behavioural disorders. *World Health Organization.*

World Health Organization. (2015). World report on ageing and health. Retrieved from https://apps.who.int/iris/handle/10665/186463

Appendix 1

Air quality monitoring stations in Veneto Region

Name	Class	Subclass	Prov	Latitude	Longitude	Altitude
BL - Parco Città di Bologna	Fondo	Urbano	BL	46,14193	12,21758	394
BL - La Cerva	Traffico	Urbano	BL	46,14376	12,21279	403
Area Feltrina	Fondo	Suburbano	BL	46,03127	11,90675	295
Pieve d'Alpago	Fondo	Rurale	BL	46,16253	12,36131	612
PD - Mandria	Fondo	Urbano	PD	45,37101	11,84084	12
PD - Arcella	Traffico	Urbano	PD	45,43291	11,8897	12
PD - Granze	Industriale	Urbano	PD	45,37757	11,93996	9
Alta Padovana	Fondo	Rurale	PD	45,60158	11,90355	28
Este	Industriale	Suburbano	PD	45,22693	11,66618	11
Parco Colli Euganei	Fondo	Rurale	PD	45,28937	11,64239	13
Adria	Fondo	Urbano	RO	45,04583	12,06087	1
RO - Largo Martiri	Traffico	Urbano	RO	45,07387	11,78255	7
RO - Borsea	Fondo	Urbano	RO	45,03876	11,79015	4
Badia Polesine	Fondo	Rurale	RO	45,10408	11,55324	7
Conegliano	Fondo	Urbano	TV	45,88973	12,30712	62
Mansue'	Fondo	Rurale	TV	45,83694	12,51036	11
TV - Via Lancieri	Fondo	Urbano	TV	45,67172	12,23781	15
TV - S. Agnese	Traffico	Urbano	TV	45,6588	12,21618	15
S.Dona' di Piave	Fondo	Urbano	VE	45,62909	12,59068	1
VE - Malcontenta	Industriale	Suburbano	VE	45,43826	12,20553	2
VE - Parco Bissuola	Fondo	Urbano	VE	45,49962	12,26125	1
VE - Sacca Fisola	Fondo	Urbano	VE	45,42842	12,31293	1
VE - Via Tagliamento	Traffico	Urbano	VE	45,4896	12,21753	3
Bassano del Grappa	Fondo	Urbano	VI	45,75927	11,73585	114
Schio	Fondo	Urbano	VI	45,71356	11,36766	189
VI - Quartiere Italia	Fondo	Urbano	VI	45,55956	11,53865	37
VI - S. Felice	Traffico	Urbano	VI	45,54499	11,53318	34
Boscochiesanuova	Fondo	Rurale	VR	45,5892	11,03691	824
Legnago	Fondo	Urbano	VR	45,18263	11,31051	15
S. Bonifacio	Traffico	Urbano	VR	45,39883	11,2853	31
VR- Borgo Milano	Traffico	Urbano	VR	45,44522	10,95395	68
VR-Giarol Grande	Fondo	Suburbano	VR	45,43341	11,03057	48
VE- Rio Novo	Traffico	Urbano	VE	45,43557	12,32313	1

Source: Regional Agency for Environmental Prevention and Protection Veneto Region

Appendix 2

WHO ICOPE tool

Priority conditions associated with declines in intrinsic capacity	Tests	Assess fully any domain with a checked circle
COGNITIVE DECLINE (Chapter 4)	1. Remember three words: flower, door, rice (for example) 2. Orientation in time and space: What is the full date today? Where are you now (home, clinic, etc)? 3. Recalls the three words?	○ Wrong to either question or does not know ○ Cannot recall all three words
LIMITED MOBILITY (Chapter 5)	Chair rise test: Rise from chair five times without using arms. Did the person complete five chair rises within 14 seconds?	○ No
MALNUTRITION (Chapter 6)	1. Weight loss: Have you unintentionally lost more than 3 kg over the last three months? 2. Appetite loss: Have you experienced loss of appetite?	○ Yes ○ Yes
VISUAL IMPAIRMENT (Chapter 7)	Do you have any problems with your eyes: difficulties in seeing far, reading, eye diseases or currently under medical treatment (e.g. diabetes, high blood pressure)?	○ Yes
HEARING LOSS (Chapter 8)	Hears whispers (whisper test) ***or*** Screening audiometry result is 35 dB or less ***or*** Passes automated app-based digits-in-noise test	○ Fail
DEPRESSIVE SYMPTOMS (Chapter 9)	Over the past two weeks, have you been bothered by - feeling down, depressed or hopeless? - little interest or pleasure in doing things?	○ Yes ○ Yes