UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI AGRARIA

Department of Food, Environmental and Nutritional Sciences

Graduate School in Biochemical,

Nutritional and Metabolic Sciences

PhD Course in Experimental and Clinical Nutrition

XXVII cycle

Diet and other lifestyles factors in adulthood and incidence of chronic diseases and mortality: a passive-case finding cohort study. Evaluation of the possible use of record linkage in the analysis of weak exposures.

Federica Prinelli

2013/2014

Tutor: **Prof. Ambrogina Pagani**

Co tutor: **Dr. Massimo Musicco**

Prof. Monica Ferraroni

Thesis supervisor: **Prof. Amelia Fiorilli**

	Abstract	**3**
	Riassunto	**6**
0	**PREFACE**	**9**
0.1	**Preface**	**10**
0.2	**References**	**14**
1	**STATE OF THE ART**	**15**
1.1	**Administrative Data for Epidemiological Purposes**	**16**
1.2	**International Classification of Diseases**	**20**
1.3	**Healthy Status and Chronic Diseases: an Overview**	**23**
1.4	**Overview of Nutritional Epidemiology**	**25**
1.5	**References**	**28**
2	**AIMS OF THE STUDY**	**31**
2.1	**Previous data**	**32**
2.2	**Aims of the present study**	**36**
2.3	**References**	**37**
3	**SUBJECTS AND METHODS**	**39**
3.1	**Study Design**	**40**
3.2	**Study Sample**	**41**
3.3	**Baseline Data Collection**	**42**
3.3.1	**Structured Questionnaire**	**42**
	Section A: Socio-demographic variables	**42**
	Section B: Family history of diseases and use of drugs	**43**
	Section C: Life-style variables	**43**
3.3.2	**Laboratory Test**	**44**
3.3.3	**Blood Pressure Measurements**	**44**
3.3.4	**Anthropometric Assessment**	**44**
3.3.5	**Dietary Assessment**	**46**
3.5	**Ascertainment of Life Status and Dataset Definition**	**48**

3.6	**Outcomes Ascertainment**	**51**
3.7	**Dietary Data Analysis**	**55**
3.7.1	**Percentage of Energy from Macronutrient**	**57**
3.7.2	**A *priori* Dietary Pattern**	**59**
3.7.3	**A *posteriori* Dietary Pattern**	**62**
3.8	**Statistical Analysis: Multivariate Survival Analysis Using Cox's Regression Model**	**64**
3.8.1	**Association between Dietary Intake and Endpoints**	**69**
3.9	**References**	**71**
4	**RESULTS**	**75**
4.1	**Descriptive Analyses**	**76**
4.2	**Associations between Dietary Intake and CVD Incidence**	**79**
4.3	**Associations between Dietary Intake and All-causes Death**	**81**
5	**DISCUSSION**	**83**
5.1	**The use of Record Linkage in Observational Studies: Problems and Solutions**	**85**
5.2	**Effects of Diet and other Lifestyle Factors on CVD and All-causes Death**	**91**
5.3	**Weaknesses and Strengths of the Project**	**98**
5.4	**References**	**102**
6	**CONCLUSIONS**	**109**
7	**TABLES**	**112**
	Appendix 1	**131**
	Acknowledgments	**149**

Abstract

Introduction and Aims: Although the integrated use of administrative data in public health and epidemiological research is of increasing interest, the practice of linking data from epidemiological studies to administrative data is not very common. As far as I know, only few longitudinal studies have been carried out combining historical cohort and administrative data for evaluating the effects of life-styles factors collected in mid-life on long term health outcome, and the main evidence are related to mortality.

The aims of this thesis project were: i) to reconstruct the history of a cohort of subjects who were visited and interviewed during the period from 1991-1995, ii) to estimate the incidence of major diseases that occurred during 20-years follow-up and mortality for all causes, and iii) to study the relationship between dietary habits and other lifestyles factors in middle age, and risk of cardiovascular diseases (CVD) and death by means of passive follow-up.

Subjects and Methods: The data used for illustrating the possible application of the use of administrative data in the nutritional epidemiology field come from the cross-sectional 'Bollate Eye Study', which was carried out in a population sample of people aged 40-74 resident in Lombardy Region. At baseline, the hospital visits were attended by 1693 subjects and included an interview concerning medical and family history, the past or current use of drugs, lifestyle habits, and a food frequency questionnaire (FFQ) for estimating the usual intake of 158 foods during the year preceding the assessment. The study involved the application of algorithms developed to retrieve subjects of the cohort in the Regional Health Services databases (RHSD) of the Local Health Authority (ASL) of Mi1 by using fiscal and regional code as personal identifiers codes. The main outcomes investigated in this work were the incidence of cardiovascular diseases (CVD) and all-causes death, that were determined by means of deterministic record linkage (DRL) with the Regional

Hospital discharge forms and the Mortality registries by using an international coding system of diseases. Baseline data were thereafter aggregated into a single dataset to carry out all of the analyses.

For evaluating the role of diet on the two endpoints, the dietary data collected at baseline were processed and analyzed with different approaches. First, energy derived from macronutrient intake was calculated and then categorized in tertiles, then an *a priori* index was computed in order to assess the adherence to Mediterranean Diet (MeDi) and finally an exploratory *a posteriori* approach was applied to characterize subjects on the basis of their dietary patterns. Furthermore, the combined effect of lifestyle factors was evaluated by calculating a healthy score. The relationship between exposures (diet, physical activity, smoking habits and the healthy score) and fatal and nonfatal CVD and all-causes death was investigated by means of Cox's regression models that estimated relative risks (RRs) as hazard ratios (HRs) with their 95% confidence intervals (95%CI).

Results and Conclusion: From a methodological point of view, several issues related to the reconstruction of the cohort and to the outcomes definition have been focused and then addressed. After the record linkage procedures, 95% of subjects recruited during the period 1991-1995 were retrieved in the registries (n=1604). Excluding subjects with previous chronic diseases at baseline, a total of 530 CVD occurred and 194 persons died for all causes during follow-up. Results support the evidence of a positive effect of a healthy lifestyle, as adherence to a Mediterranean dietary pattern, the intake of specific components of this pattern, such as olive oil, abstinence from smoking and engaging in regular physical activity on the risk of CVD and all-causes death, particularly when these modifiable risk factors were considered in a healthy score.

Furthermore, this thesis showed a possible application of the use of administrative data rarely used for epidemiological analytical study in Italy, suggesting their promising role for carrying out observational cohort studies with a large amount of data collected at baseline, guaranteeing a reduction in times and costs compared to traditional studies, ensuring a minimal number of persons lost to follow-up and resulting in a high number of subjects to analyze.

Riassunto

Introduzione e Obiettivi: Sebbene l'uso integrato di dati amministrativi in sanità pubblica e nella ricerca epidemiologica sia di crescente interesse, il loro utilizzo nel mettere in relazione dati provenienti da studi epidemiologici con quelli provenienti da fonti amministrative non è molto frequente. Ad oggi, pochi studi longitudinali hanno valutato gli effetti degli stili di vita, come la dieta, nella mezza età su esiti di salute a lungo termine e comunque i principali si sono focalizzati sulla mortalità.

Gli obiettivi di questo progetto di tesi sono stati: i) ricostruire la storia di un gruppo di soggetti che furono visitati e intervistati tra il 1991 e il 1995, ii) stimare l'incidenza delle principali malattie che si sono verificate durante 20 anni di follow-up e la mortalità per tutte le cause, e iii) studiare la relazione tra abitudini alimentari e altri stili di vita raccolti nella mezza età e il rischio di sviluppare malattie cardiovascolari (CVD) e morte attraverso l'utilizzo di un follow-up passivo.

Soggetti e Metodi: Per illustrare la possibile applicazione dell'uso di dati amministrativi in epidemiologia nutrizionale sono stati utilizzati dati provenienti dallo studio di 'Bollate Study Eye', condotto su un campione di popolazione di età compresa tra 40 e 74 anni residente nella Regione Lombardia. Al baseline, 1.693 soggetti furono visitati ed intervistati sull'utilizzo di farmaci, su malattie pregresse, e sugli stili di vita; ai soggetti fu inoltre somministrato un questionario alimentare per frequenza (FFQ) per stimare l'assunzione abituale di 158 alimenti durante l'anno precedente. Lo studio si è avvalso dell'applicazione di algoritmi sviluppati per identificare i soggetti della coorte nei database dei Servizi Sanitari Regionali (RHSD) dell'Azienda Sanitaria Locale (ASL) di Mi1, utilizzando il codice fiscale e regionale come codici identificativi personali. I due endpoints principali di

questo lavoro sono stati l'incidenza delle malattie cardiovascolari (CVD) e il decesso per tutte le cause, che sono stati determinati per mezzo di record linkage deterministici (DRL) con i registri di dimissione ospedaliera e di mortalità.

I dati originariamente raccolti al baseline sono stati integrati con quelli di follow-up in un unico dataset utilizzato per la conduzione delle analisi statistiche.

Per valutare il ruolo della dieta, i dati nutrizionali raccolti al baseline sono stati elaborati ed analizzati mediante approcci differenti. In primo luogo è stata calcolata l'energia proveniente dal consumo dei macronutrienti e categorizzata in terzili, poi è stato calcolato un indice *a priori* al fine di valutare l'aderenza alla dieta mediterranea (MeDi) ed infine è stato applicato un approccio esplorativo *a posteriori* per caratterizzare i soggetti della coorte sulla base delle loro abitudini alimentari. Inoltre, è stato valutato l'effetto combinato dei diversi fattori legati agli stili di vita, calcolando un punteggio 'salutare'. Lo studio della relazione tra le diverse esposizioni (dieta, attività fisica, abitudine al fumo e punteggio salutare) e l'insorgenza di CVD e la morte per tutte le cause è stata studiata utilizzando modelli di regressione di Cox, che permettono di stimare i rischi relativi (RRs) espressi come hazard ratios (HRs) con i relativi intervalli di confidenza al 95% (95% CI).

Risultati e Conclusioni: Da un punto di vista metodologico, si sono presentate e sono conseguentemente state affrontate diverse problematiche relative alla ricostruzione della coorte e alla determinazione degli outcomes. Attraverso il record linkage, il 95% dei soggetti reclutati nel periodo 1991-1995 è stato recuperato dai registri regionali (n=1604). Escludendo i soggetti con precedenti di malattie croniche al baseline, 530 soggetti hanno sviluppato CVD e 194 persone sono morte durante il follow-up. I risultati confermano che mantenere un corretto stile di vita, consumare alimenti tipici della dieta mediterranea, come ad esempio l'olio d'oliva, non fumare e svolgere regolare attività fisica sono condizioni che riducono il rischio di CVD e garantiscono un prolungamento della vita, soprattutto considerando questi fattori in combinazione tra di loro.

Inoltre, questa tesi ha mostrato una possibile applicazione dell'uso dei dati amministrativi, che in Italia vengono utilizzati raramente in combinazione con studi di epidemiologia. L'utilizzo di questa sorgente di informazioni può rivestire un ruolo promettente nella conduzione di studi di coorte osservazionali che dispongano di numerose informazioni al baseline, e risulterebbe molto vantaggioso sia per la riduzione dei tempi e dei costi rispetto agli studi tradizionali sia per il ridotto numero di persi al follow-up e, conseguentemente, per l'elevato numero di soggetti da poter analizzare.

0.PREFACE

0.1 Preface

This is an observational epidemiological study carried out in middle-aged subjects, residents in the north - west area of Milan (Lombardy Region) who were visited and interviewed about dietary habits and other lifestyle related factors, during the period from 1991 to 1995 (Nicolosi 1997)

The main goals of the present thesis project were to reconstruct and trace the history of the cohort of subjects recruited 20 years ago, and to obtain several information about the health status of each one using administrative data. Furthermore, the association between the behaviors collected during adulthood, as dietary habits, and the risk of diseases that were occurred during follow-up and mortality for all-causes, has been evaluated.

Generally, the longitudinal studies performed in order to investigate the relationship between diet and diseases, assume that the dietary intake is assessed at baseline and at given points in time during follow-up, to capture changes in eating behaviors. In our study, the dietary assessment has been carried out once at baseline without further evaluations as defined in the study design. The possible reason is that 20 years ago there was not a great knowledge and awareness about the relationship between diet as exposure and diseases.

With this work I focused my attention on the methodological issues related to the dietary analysis in particular conditions, for example when a single assessment has been carried out and also when an active follow-up is not possible for different reasons related to costs, time or scarcity of personnel, and furthermore, highlighting the potentiality of the use of administrative data.

The study was carried out in collaboration with the Local Health Authority of Milan (ASL MI1) who gave me the opportunity to access to the data from administrative databases.
In all the phases of the project, starting from the definition of the dataset, the diseases ascertainment and finally the dietary analysis, several aspects related to the data collection and elaboration, have been taken into account.
In the first phase relating to the dataset definition, the data sources were not completely homogeneous, so it has been necessary to check them in order to avoid possible discrepancies, errors or duplication.

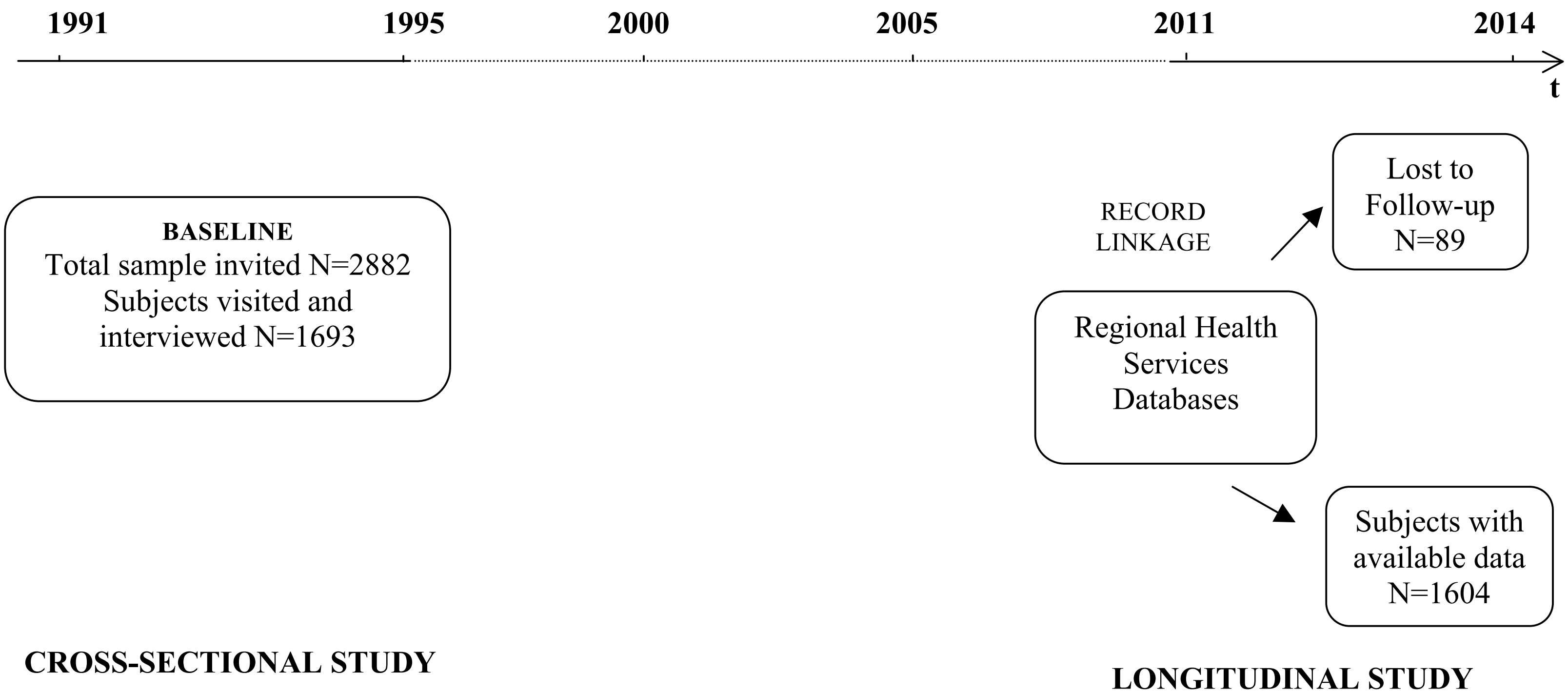
The administrative databases include different data sources that can be interrogated through record linkage procedure. The International Classification of Diseases (ICD) was used to identify and classify the major diseases occurred during the follow-up, such as cardiovascular diseases (CVD), cancer, diabetes and neurodegenerative disorders (Alzheimer's and Parkinson's Disease).
In particular, I decided to explore the role of dietary habits and other lifestyles behaviors, considering two primary endpoints as CVD because they were more represented in this cohort, and mortality for all-causes.
For the dietary data processing, I studied in detail the application of different approaches for analyzing them.

This population represents an interesting cohort that could be followed over time and used for performing several epidemiological studies without the need to design ad hoc studies for collecting information.
Thus, the results of this study could be, in the future, compared with those provided by other historical cohort such as the Framingham Study (Block et al. 2013) the NANHES Study (Hamner et al. 2009), the EPIC Study (Duarte-Salles et al. 2014) or the SUN Projects (Martinez-Gonzalez et al. 2002).

This thesis project has set itself the goal of illustrating a possible application of the use of administrative data for epidemiological purposes.

In order to realize it, several methodological and non methodological issues, have been taken into account.

Figure 1. Scheme of the steps involved in the study sample definition



0.2 References

Block JP, Subramanian SV, Christakis NA, et al. (2013). Population trends and variation in body mass index from 1971 to 2008 in the Framingham Heart Study Offspring Cohort. PLoS One **8**(5): e63217.

Duarte-Salles T, Fedirko V, Stepien M, et al. (2014). Dairy products and risk of hepatocellular carcinoma: the European Prospective Investigation into Cancer and Nutrition. Int J Cancer **135**(7): 1662-1672.

Hamner HC, Mulinare J, Cogswell ME, et al. (2009). Predicted contribution of folic acid fortification of corn masa flour to the usual folic acid intake for the US population: National Health and Nutrition Examination Survey 2001-2004. Am J Clin Nutr **89**(1): 305-315.

Martinez-Gonzalez MA, Sanchez-Villegas A, De Irala J, et al. (2002). Mediterranean diet and stroke: objectives and design of the SUN project. Seguimiento Universidad de Navarra. Nutr Neurosci **5**(1): 65-73.

Nicolosi A (1997). Le indagini di popolazione sulla prevalenza delle principali malattie oculari in Italia. In L'Epidemiologia oftalmica in Italia. Cerulli L, Miglior M and Ponte F. Roma.

1.STATE OF THE ART

1.1 Administrative Data for Epidemiological Purposes

In the last decades, routinely collected and electronically stored healthcare information has spread throughout the world (Motheral 1997).

Large computerised databases with millions of observations regarding the use of drugs, vaccines, devices, and procedures along with health outcomes may be useful in assessing which treatments are most effective and safe in routine care without long delays and the prohibitive costs of randomised clinical trials (RCTs) (Berger 2002).

The integrated use of electronic records in public health and epidemiological research is of increasing interest.

The administrative systems have, however, become exceedingly complex and typically are intricate relational databases. The relational database process has been widely used to combine information from individuals or entities from varied sources, using record linkage (RL). The term "record linkage" is defined as the process of comparing two or more records that contain identifying information to determine whether these records refer to the same entity (Suzuki et al. 2013). RL - i.e. the use and linkage of electronic, mainly administrative databases – has been utilized in epidemiological research over three decades now, starting from the system implemented in a few US health mantainance organization (HMO, i.e. the Kaiser Permanente network), a few Nordic Countries system, or the UK general practice research database (UK GPRD, now clinical practice research datalink, CPRD) (Corrao 2013, Schneeweiss et al. 2013).

RL techniques are widely used in epidemiological studies to obtain comprehensive information and conduct more robust analyses and essentially to investigate the implications and side effects of drugs and, more in general, medical interventions.

There are two commonly used computerized record linkage approaches: the

probabilistic and deterministic. The probabilistic record linkage (PRL) approach creates links between two databases based on a calculated statistical probability of a set of common identifiers. The probability is used to determine whether a pair of records approximately refers to the same individuals(Li et al. 2006). The deterministic record linkage (DRL) approach generates links on the basis of a full agreement of a unique identifier. These methods minimize the uncertainties in the match between two databases since only complete match on a set of personal variables is accepted at the cost of lowering the linkage rate.
Both the deterministic and probabilistic methods are prone to two types of errors: false negatives (the procedure does not match together records that do belong to the same person) and false positives (records that have been matched together erroneously, due to the quality or commonality of the identifying items). In the trade-off between the two types of errors most researchers prefer to allow more false negatives than false positives, on the premise that the file can be re-matched at some future time. The extent of the mismatch can be estimated by extracting and clerically inspecting a random sample of the matches. A second option employs assumptions regarding the dependency of mismatch on the amount of potential matches before the linkage decision is taken, and the estimation of error on the basis of the actual frequency of ambiguous matches (Blakely & Salmond 2002).

Several aspects related to the use of administrative data have to be considered. The principle disadvantage of administrative data is the limitation to the data elements that were introduced, almost always for a totally different purpose. If there is a change in the administrative procedures (e.g. a change in entitlement) then the definition of the information collected could change. This in turn could make comparisons over time problematic (Smith 2004). Another criticisms addressed to these resources concerns the lack of control the researcher has during the data collection stage and how this affects what can be done with the data and more general, the lack of well established theory and methods to guide the use of administrative data in social science research (Wallgren 2007). A related problem is the absence of contextual/background information, related to the "culture" of a particular region of the country that might provide some "explanation" of the large observed differences in service use patterns. A limit of the use of this kind of data is the frequent lack of individual-level socio-economic status information as educational level, or job description that often are not recorded at the population level with few exception (Black 2005).

Data on social supports and interpersonal relationships are another type of information not often available in population-wide administrative data. Studies have shown these to be important components of health and wellbeing. Race/ethnicity is also often unavailable in administrative data (Jutte et al. 2011).

Administrative data have several strengths also. The primary advantage of administrative data systems is that they are already being collected and therefore there are no additional costs of collection; they are regularly (sometimes continuously) updated; they can also (if historical extracts are retained) provide information about past periods of time that pre-date the formation of the research question; if they forms part of a national system it is likely to be collected in a consistent way; they are subject to rigorous quality checks; the collection process is not intrusive to the target population; administrative systems will contain 100% of the records in question.

Furthermore, such data are free from many of the measurement and loss-to-follow-up problems associated with longitudinal surveys. The flexibility of continuously collected administrative data presents some advantages over surveys dependent on expensive, repeated interviews of the same respondents at fixed intervals. Researchers are able to study the health or health care utilization of a population for any length of time and for any year or set of years. This flexibility can be useful for natural experiments such as abrupt economic changes or the implementation of new policies (Roos 2008).

Finally, administrative data are useful to compare regions, areas, and hospitals; study policy interventions longitudinally; combine information related to physicians and their patients; sum expenditures across different services within the health care system; and examine social determinants of health using education and family services data in conjunction with health-related information (Roos 2004).

1.2 International Classification of Diseases

Passive surveillance for disease is a public health approach that relies on documentation available within existing health records for the region or communities being studied. The effectiveness of passive case-finding depends on the comprehensiveness of the healthcare coverage in a given community and the adequacy of the available medical records (Alberti et al. 2006, Knopman et al. 2011).

Specific disease coding must be used to extract patients experiencing a given outcome and to characterize their clinical profile.

The International Classification of Diseases 9th revision, Clinical Manifestation (ICD-9-CM) and the ICD-10th revision in some countries, is the classification system of diseases and medical procedures; it is the standard diagnostic tool for epidemiology, health management and clinical purposes that include the analysis of the general health situation of population groups. It is used to monitor the incidence and prevalence of diseases and also to classify diseases and other health problems recorded on many types of health and vital records including death certificates and health records. In addition to enabling the storage and retrieval of diagnostic information for clinical, epidemiological and quality purposes, these records also provide the basis for the compilation of national mortality and morbidity statistics by World Health Organization (WHO) Member States (WHO 2002).

The 9th classification of diseases and injuries includes 17 chapters or broad groups, of them 10 are dedicated to specific organs or anatomical apparatus, while the other 7 describe specific types of conditions that affect the entire body:

- Infectious and parasitic diseases (001-139)
- Neoplasms (140-239)

- Endocrine, nutritional and metabolic diseases and immunity disorders (240-279)
- Diseases of the blood and blood-forming organs (280-289)
- Mental disorders (290-319)
- Diseases of the nervous system and sense organs (320-389)
- Diseases of the circulatory system (390-459)
- Diseases of the respiratory system (460-519)
- Diseases of the digestive system (520-579)
- Diseases of the genitourinary system (580-629)
- Complications of pregnancy, childbirth and the puerperium (630-679)
- Diseases skin and subcutaneous tissue (680-709)
- Diseases of the musculoskeletal and connective tissue (710-739)
- Congenital anomalies (740-759)
- Newborn (perinatal) guidelines (760-779)
- Signs, symptoms and ill-defined conditions (780-799)
- Injuries and poisonings (800-999)

Each chapter includes a set of codes that classify diseases related the same anatomical apparatus or the same clinical typology. Each of the 17 chapters is divided in the following parts:

 block: a set of closely related conditions (eg. Diseases Intestinal infectious, 001-009);

 category: three-character codes, some of which are very specific and not further subdivided (eg. 462 acute pharyngitis), while others are further subdivided, with the addition of a fourth character after point decimal;

 Subcategory: four-character codes; the fourth character provides further specificity or information in relation to etiology, location or clinical manifestation.

The additional Classification of factors influencing health status and contact with health services (V01-V91) is composed of alphanumeric codes starting with the letter V and are used to describe clinical problems, or services provided under special circumstances.

The ICD-10 was endorsed by the Forty-third World Health Assembly in May 1990 and came into use in WHO Member States as from 1994. ICD-10-CM diagnosis code set differ from the ICD-9-CM for several aspects:

- The code set has been expanded from five positions (first one alphanumeric, others numeric) to seven positions. The codes use alphanumeric characters in all positions, not just the first position as in ICD-9.
- As of the latest version, there are 68,000 existing codes, as opposed to the 13,000 in ICD-9.
- The new code set provides a significant increase in the specificity of the reporting, allowing more information to be conveyed in a code.
- The terminology has been modernized and has been made consistent throughout the code set.
- There are codes that are a combination of diagnoses and symptoms, so that few codes need to be reported to fully describe a condition.

1.3 Healthy Status and Chronic Diseases: an Overview

In 1900, the three leading causes of death were pneumonia and influenza; tuberculosis and gastritis, enteritis and colitis. These diseases accounted for nearly one-third of all deaths. Today heart disease, malignant neoplasm and cerebrovascular diseases are the three leading causes of death, accounting for over half of all deaths. These changes are known as the "epidemiological transition" and the predictably occurs in populations that move through stages of technological and economic development (Omran 1971). With these changes in the leading causes of mortality, the primary emphasis of societal systems designed to address health concerns has generally shifted from microbiologic investigation of communicable diseases to a focus on the role of behavioural and environmental risk factors and methods for preventing diseases, disabilities and death in population.

Chronic disease is a term referred to a chronic illness, non-communicable diseases and degenerative diseases. They are generally characterized by uncertain etiology, multiple risk factors, a long latency period, a prolonged course of illness, non-contagious origin, functional impairment or disability, and incurability. In more general terms, a chronic disease can be defined as a disease that has a prolonged temporal course, that does not resolve spontaneously, and for which a complete cure is rarely achieved (Torres 2011).

An estimated 36 million deaths, or 63% of the 57 million deaths that occurred globally in 2008, were due to non-communicable diseases, comprising mainly cardiovascular diseases (48% of non communicable diseases), cancers (21%), chronic respiratory diseases (12%) and diabetes (3.5%). In 2008, 80% of all deaths (29 million) from non-communicable diseases occurred in low- and middle-income countries, and a higher proportion (48%) of the deaths in the latter countries are premature (under the age of 70) compared to high income countries (26%). According to WHO's projections, the total annual number of deaths from non communicable diseases will increase to 55 million by 2030 if "business as usual" continues (WHO 2013).

Chronic diseases are largely preventable diseases. While age, sex and genetic susceptibility are non modifiable risk factors, many of the conditions associated with age and sex are modifiable. Such risks include behavioural factors (e.g. diet, physical inactivity, tobacco use, alcohol consumption); biological factors (e.g. dyslipidaemia, hypertension, overweight, hyperinsulinemia); and finally societal factors, which include a complex mixture of interacting socioeconomic, cultural and other environmental parameters (WHO 2002).

The goals of chronic diseases prevention and control, are to reduce the incidence of diseases, delay the onset of disease and disability, alleviate the severity of disease and improve the health-related quality and duration of the individuals life (Torres 2011).

1.4 Overview of Nutritional Epidemiology

The public has a heightened sense of interest in the relation between diet and the risk of disease. Nutritional epidemiology is a specialized field within epidemiology and concerns the study of the nutritional determinants of diseases. Most of the nutritional epidemiology is concerned with effects of diet on chronic disease that are multifactorial and that take years, if not decades, to develop. The methods used in nutritional epidemiology are designed to take into account different features focused on measuring the exposure to nutritional factors, the frequency and distribution of disease, and the exposure to other factors that could confound the hypothesize association (Sempos et al. 1999).

Although the formal study of diet and health is only a few decades old, the importance of diet to maintain health was already known to the ancient Greeks. As Hippocrates (460–377 BC) said: 'If we could give every individual the right amount of nourishment and exercise, not too little and not too much, we would have found the safest way to health" (Hippocrates. Adams 1955).

The relationship between diet and the occurrence of the major diseases of our civilization are both scientific and practical importance to public health nutritionist (Willet 1990).

A major difficult in nutritional epidemiology is the complex nature of diet and the difficulties in measuring diet as an exposure. Diet represents an unusually complex set of exposures that are strongly correlated. We all eat, we all eat many different foods, we tend to forget rather quickly what we ate, and we often do not know the ingredients of the dishes we consume. Hence we are all exposed, and the variation may be most subtle than with other, more distinct exposures such as smoking or use of hormone replacement therapy. Few people maintain extreme diets; thus assessing diet within fairly homogeneous populations makes it difficult to detect associations between dietary patterns (or particular foods and nutrients) and health or disease due to the lack of sufficient variation. All individuals are exposed to hypothesize causal factors; everyone eat fats, fibre and vitamin A. Thus exposure cannot be characterized as present or absent; rather they are continuous variables, often with rather limited range of variation. Furthermore, individuals rarely make clear changes in their diet at identifiable points in time; more typically, eating patterns evolve over periods of years.

The most serious limitation to research in nutritional epidemiology has been the lack of practical methods to measure diet. Because such epidemiologic studies usually involve at least several hundred and sometimes tens of thousands of subjects, dietary assessment methods must be not only reasonably accurate, but also relatively inexpensive (Willet 1990).

The interpretation of positive (or inverse) and null association in epidemiologic studies has received considerable attention. In the nutritional epidemiology field, true association are not likely to be strong although relative risks of 0.7 or 1.5 could potentially be important because the dietary exposure are common. In a study of diet and disease failure to observe a statistically significant association when such an association truly exist can occur in several circumstances alluded to earlier. One possibility is that the variation in diet is insufficient; in the extreme, no association can occur if everyone in the study population eats the same diet. Second variation may exist for the study population, but only within a flat portion of the dose-response relationship. A thirds possibility is that the method of measuring dietary intake is not sufficiently precise to measure differences that truly exist. Fourth an association may be missed because of low statistical power due to inadequate number of diseases and non diseased subjects. Fifth a relationship can be undetected because the temporal relationship between the measured exposure and the occurrence of disease did not encompass the true latent period; this could easily happen if the critical dietary exposure occurred during childhood and the disease was diagnosed during adulthood. Sixth, an association could be undetected because an unmeasured third variable was related to exposure and disease in opposite directions, in other words negative confounding exist. Finally, methodological sources of bias could obscure a relationship (Willet 1990).

Types and amounts of food eaten may be related to important non dietary determinant of disease, such as age, smoking, exercise and occupation, which may distort or confound and modify relationship with diet. Due to these complexities, it is almost always necessary to employ multivariate techniques to adjust for potentially confounding variables and examine interactions.

Another aspect is that individuals who try to eat a healthy diet are likely to lead a healthy lifestyle in general. It is probably not possible to measure all the important markers of a healthy lifestyle sufficiently to eliminate confounding. The inability to distinguish the effect of diet from that of other lifestyle factors may pose a threat to the validity of diet—disease associations observed in epidemiologic studies (Michels 2003).

1.5 References

Alberti KG, Zimmet P & Shaw J (2006). Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med **23**(5): 469-480.

Berger MM, M. Atkins, D. et al. (2002). Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force. Report-Part I. Value in Health **12**: 1044-1052.

Black CR, LL. (2005). Linking and combining data to develop statistics for understanding the population’s health. In Heatlth Statistics: Shaping Policy and practice to Improve the Population's Health. Friedman, DJ. Hunter, EL. Parrish II, RG. (Eds). New York, NY: Oxford University Press.

Blakely T & Salmond C (2002). Probabilistic record linkage and a method to calculate the positive predictive value. Int J Epidemiol **31**(6): 1246-1252.

Corrao G (2013). Building reliable evidence from real-world data: methods, cautiousness and recommendations. Epidemiology Biostatistics and Public Health **10**(3): e8981 8981-8940.

Hippocrates. Adams FB, AJ. Galen. (1955). Hippocratic Writings. Chicago, Encyclopædia Britannica.

Jutte DP, Roos LL & Brownell MD (2011). Administrative record linkage as a tool for public health research. Annu Rev Public Health **32**: 91-108.

Knopman DS, Petersen RC, Rocca WA, et al. (2011). Passive case-finding for Alzheimer's disease and dementia in two U.S. communities. Alzheimers Dement **7**(1): 53-60.

Li B, Quan H, Fong A, et al. (2006). Assessing record linkage between health care and Vital Statistics databases using deterministic methods. BMC Health Serv Res **6**: 48.

Michels KB (2003). Nutritional epidemiology--past, present, future. Int J Epidemiol **32**(4): 486-488.

Motheral BF, KA. (1997). The use of claims databases for outcomes research: rationale, challenges, and strategies. . Clin Ther **19**: 346-366.

Omran AR (1971). The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. Milbank Mem. Fund. Q. **49**: 509 – 538.

Roos LB, M. Lix, L,. Roos, NP. Walld, R. MacWilliam, L. (2008). From health research to social research: privacy, methods, approaches. Soc. Sci. Med. **66**: 117-129.

Roos LM, V. Currie, RJ. (2004). Policy analysis in an information-rich environment. Soc. Sci. Med. **58**(11): 2231-2241.

Schneeweiss S, Seeger JD, Jackson JW, et al. (2013). Methods for comparative effectiveness research/patient-centered outcomes research: from efficacy to effectiveness. J Clin Epidemiol **66**(8 Suppl): S1-4.

Sempos CT, Liu K & Ernst ND (1999). Food and nutrient exposures: what to consider when evaluating epidemiologic evidence. Am J Clin Nutr **69**(6): 1330S-1338S.

Smith GN, M. Anttilla, C. Gill, L. Zaidi, A. Wright, G. Dibben, C and Barnes, H. (2004). The Value of Linked Administrative Records for Longitudinal Analysis, Report to the ESRC National Longitudinal Strategy Committee. http://www.thepresidency.gov.za/learning/reference/analysis.pdf.

Suzuki KM, Porto Filho CH, Cozin LF, et al. (2013). Deterministic record linkage versus similarity functions: a study in health databases from Brazil. Stud Health Technol Inform **192**: 562-566.

Torres K (2011). Chronic disease epidemiology and control, 3rd edition. Remington, PL. Brownson, RC. Wegner, MV. (Eds.)

Wallgren AW, B. (2007). Register-based Statistics: Administrative Data for Statistical Purposes. John Wiley & Sons, Ltd.

WHO (2002) "International Classification of Impairments, Disability and Handicap: a manual of classification relating to consequences of diseases, Geneva, Switzerland: 2002."

WHO (2002). Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases.

WHO (2013) "Global Action Plan for the Prevention and Control of NCDs 2013-2020."

Willet WC (1990). Nutritional Epidemiology, 2nd ed. Oxford University Press, Inc.

2.AIMS OF THE STUDY

2.1 Previous data

The data used to illustrate the possible application of the use of administrative data in nutritional epidemiology, come from the cross-sectional 'Bollate Eye Study' (Nicolosi 1997), which was carried out since January 1991 until January 1995, in order to investigate the prevalence and risk factors for the major eye diseases in a population sample of people aged 40-74 retrieved form the residents' list of Bollate and Baranzate (Milan, Lombardy Region). Several data are available on this study (Leite & Nicolosi 2006, Leite et al. 2007, Leite & Nicolosi 2009, Leite 2011).

During the period from 2010-2011, I evaluated the relationship between dietary pattern and metabolic syndrome emphasizing the nutrients intake and their food sources that characterized this adult population.

Background

The metabolic syndrome (MS) is a cluster of abnormalities - abdominal obesity, atherogenic dyslipidemia, increased blood pressure and high plasma glucose - that occur together increasing the risk of CVD, cerebrovascular diseases and mortality (Grundy 2006). Among the components of MS, increased abdominal adiposity has been found associated with several metabolic pathologies, as insulin resistance, dyslipidemia and hypertension that are key components of the MS (Grundy 2008). Physical activity, smoking habits, low socio-educational status and high grade of chronic inflammation have been associated with an increased risk of MS (Villegas et al. 2004). A whole body of scientific evidence suggests that healthy lifestyles and adherence to a diet rich in fruit and vegetables, fish, poultry and low fat dairy products, are protective factors for MS. Conversely, unhealthy dietary pattern (rich in red meat and processed red meat, refined carbohydrates and low in dietary fiber) is directly related to a high risk of this condition (Williams et al. 2000, Lutsey et al. 2008).

Subjects and Methods

In the 1991-1995 period, the hospital visits were attended by 1693 subjects and included an interview concerning medical and family history, the past or current use of drugs, lifestyle habits, anthropometric and blood pressure measurements and a FFQ. Complete anthropometric, lifestyle, laboratory and dietary data were obtained for 1149 subjects. We excluded 74 subjects with diabetes (plasma glucose >125 mg/dl and/or auto-referred diabetes) for a total of 1075 subjects (543 men and 532 women). FFQ was based on 158 items grouped into 33 food subgroups. Nutrient intake was calculated using the food composition database compiled for epidemiological studies in Italy (Salvini 1998). Principal components analysis (PCA) was used to define dietary pattern and logistic regression models were performed in order to evaluate the association between MS and its single components, and dietary patterns by means odds ratios (ORs) accompanied by their 95%CI.

MS was diagnosed on the basis of the National Cholesterol Education Program (NCEP-ATP III), which is defined as the presence of three or more the following determinant: high waist circumference [WC (≥102 cm in men and ≥ 88 in women)]; impaired fasting glucose level [FG (≥100 mg/dl)]; high triglycerides [TG (≥150 mg/dl)]; low HDL-cholesterol [HDL (<40 mg/dl in men and <50 mg/dl in women)] and increased blood pressure [BP (≥130/85 mmHg)]. According to the recent International Diabetes Federation (IDF) definition, MS was diagnosed if there was central obesity (≥94 in men and ≥80 in women) plus two or more of the previous criteria (Alberti et al. 2006).

Results and Discussion

In this population the prevalence of MS was 31% in males and 22% in females using the NCEP classification (40% in males and 26% in females using the IDF definition).

Population was overweight, the age range was 50-69 years, smoking and alcohol consumption were significantly higher in males, whereas female were prevalently no smokers and abstainers. Both, had a low educational level and had a sedentary lifestyle.

Five food patterns were obtained by PCA: “Vitamins and fiber”, “Animal products”, “Vegetal fats”, “Starch products” e “PUFAs and vitD.” Multivariate models were adjusted for age, body mass index (BMI), smoking, alcohol intake, educational level, physical activity, total energy intake (Kcal/d), time spent watching television (h/d), and dietary patterns, stratifying by sex and BMI categories. MS was positively associated with the food pattern labeled ‘Animal products’ that was dominated by red meat and saturated fatty acids intakes, and also with the ‘Vegetal fats’ pattern only in males. Moreover, for males the ‘Animal products’ pattern was associated with impaired FG (≥ 100mg/dl) and with high WC (≥ 102 cm) only for those obese. In females a positive association was observed between ‘Starch products’ pattern and high WC (≥88 cm), high BP and low HDL. A negative association was observed between ‘Vitamins and fiber’ pattern and WC (≥88 cm’), and also between ‘PUFAs and vitD’ pattern and impaired FG only in females.

Results of this previous analysis showed a risk increase of MS in association with high intake of animal products in males, whereas in females we observed an increased risk with high intake of starch products characterized by high glycemic index (as refined-grain bread), while the high consumption of vitamins and fiber was inversely associated with MS. Our findings suggested that differences between males and females in the observed effects of food patterns on health depended not only on sex differences in cell metabolism but also on preferences related to food selection. The analysis of the food sources provided important elements useful to the interpretation of the results. Finally, we found that nutritional habits of a North Italy population combine the typical elements of the Mediterranean diet with those of the 'western' diet of the industrialized countries.

Conclusion

Since obesity is a key etiological factor in the development of MS, the maintenance of normal body weight should be a primary goal in preventing it. However, even this objective is not achieved, the adoption of beneficial dietary habits is important for preserving a healthy metabolic profile.

2.2 Aims of the present study

In order to show an innovative application of the use of administrative data in longitudinal observational studies, the data from the study 'Bollate Study Eye' were used.

Thus, after 20 years of follow-up, the objectives of this PhD thesis were:

i) to reconstruct the history of a cohort of subjects who were visited and interviewed during the period from 1991-1995, by means administrative databases of the Regional Health Registries

ii) to estimate the incidence of the major chronic diseases occurred during follow-up and all-causes death by using Hospital discharge and Mortality registries

iii) to process and to analyse dietary data collected at baseline using different methodological approaches

iv) to investigate the relationship between dietary habits and other lifestyle factors in middle age, and the occurrence of CVD and all-causes death

2.3 References

Alberti KG, Zimmet P & Shaw J (2006). Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. Diabet Med **23**(5): 469-480.

Grundy SM (2006). Metabolic syndrome: connecting and reconciling cardiovascular and diabetes worlds. J Am Coll Cardiol **47**(6): 1093-1100.

Grundy SM (2008). Metabolic syndrome pandemic. Arterioscler Thromb Vasc Biol **28**(4): 629-636.

Leite ML (2011). Uric acid and fibrinogen: age-modulated relationships with blood pressure components. J Hum Hypertens **25**(8): 476-483.

Leite ML & Nicolosi A (2006). Lifestyle correlates of anthropometric estimates of body adiposity in an Italian middle-aged and elderly population: a covariance analysis. Int J Obes (Lond) **30**(6): 926-934.

Leite ML & Nicolosi A (2009). Dietary patterns and metabolic syndrome factors in a non-diabetic Italian population. Public Health Nutr **12**(9): 1494-1503.

Leite ML, Nicolosi A, Firmo JO, et al. (2007). Features of metabolic syndrome in non-diabetic Italians and Brazilians: a discriminant analysis. Int J Clin Pract **61**(1): 32-38.

Lutsey PL, Steffen LM & Stevens J (2008). Dietary intake and the development of the metabolic syndrome: the Atherosclerosis Risk in Communities study. Circulation **117**(6): 754-761.

Nicolosi A (1997). Le indagini di popolazione sulla prevalenza delle principali malattie oculari in Italia. In L'Epidemiologia oftalmica in Italia. Cerulli L, Miglior M and Ponte F. Roma.

Salvini SP, M. Gnagnarella, P. (1998). Banca dati di composizione degli alimenti per studi epidemiologici in Italia. Milano: Istituto Europeo di Oncologia.

Villegas R, Creagh D, Hinchion R, et al. (2004). Prevalence and lifestyle determinants of the metabolic syndrome. Ir Med J **97**(10): 300-303.

Williams DE, Prevost AT, Whichelow MJ, et al. (2000). A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome. Br J Nutr **83**(3): 257-266.

3.SUBJECTS AND METHODS

3.1 Study Design

This is an observational prospective cohort study on a population of persons living in two towns of Northern Italy (Lombardy Region) during the period from 1991 to 1995. The Italian 'Bollate Eye Study', was designed to investigate the prevalence and risk factors of major eyes diseases (Nicolosi 1997). National Research Council (Milan, Italy) approved all study procedures and the ethical aspects. All subjects visited and interviewed at baseline was informed of the scope of the study and gave their informed consent to participate.

METHODOLOGICAL NOTE

Prospective cohort study is an observational study that look at the distribution or determinant of an outcome without attempting to change the risk factor, so nature is allowed to take its course without any manipulation by the researchers. If an observational study doesn't include a control group is a descriptive study, otherwise is an analytical study. In the last case the temporal direction of the study needs to be identify. Analytical studies include prospective, retrospective and cross-sectional studies. Prospective cohort studies proceed in a logical sequence: from exposure to outcome. Investigators identify a group with an exposure of interest and another group or groups without the exposure. The investigators then follow the exposed and unexposed groups forward in time to determine outcomes. If the exposed group develops a higher incidence of the outcome than the unexposed, the exposure may be considered associated with an increased risk of the outcome (Grimes & Schulz 2002). The question being asked in the prospective studies is: "Do persons with the risk factor develop or die from the disease more frequently or sooner than those who not have the risk factor?". Cohort studies begin with a cross-sectional survey that ascertains the presence or absence of the disease of *interest as well as measures of exposure (i.e. dietary intake). After persons diagnosed with the disease being studied are excluded, the population is then followed over time to see who develops the diseases and when the disease develops. The persons remaining in the study are those who are at risk of developing the disease (Sempos et al. 1999).*

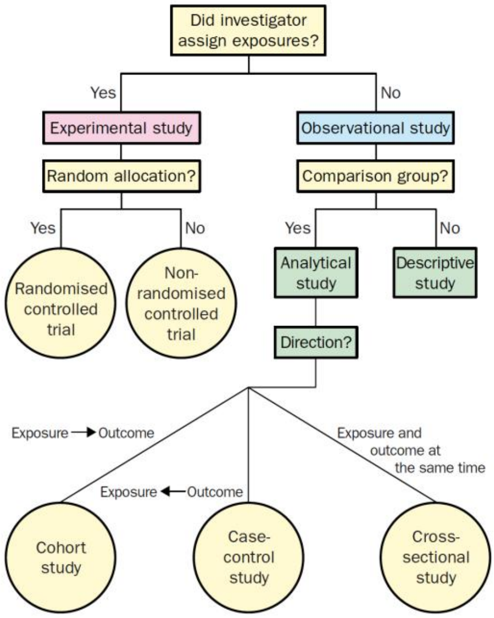


Figure 1: **Algorithm for classification of types of clinical research**

3.2 Study Sample

The study population was a simple random sample of all individuals aged 40–74 years drawn from the residents' list of the towns of Bollate and Baranzate (Milan, Italy). A letter was sent to the sampled subjects (2882) to explain the objectives and operatives modalities of the study and invite their participation. Subsequently each subject was contacted by phone to collect consent and booking an appointment at the local Hospital. Of the contacted individuals, 1693 (58,7%) were visited and interviewed.

METHODOLOGICAL NOTE

There are four major types of probability sample designs: simple random sampling, stratified sampling, systematic sampling, and cluster sampling. Simple random sampling is the most recognized probability sampling procedure where a sample has a known probability of being selected. It is the basic sampling technique where researchers select a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and every possible sample of a given size has the same chance of selection; i.e. each member of the population is equally likely to be chosen at any stage in the sampling process (Daniel 2012).

3.3 Baseline Data Collection

The Hospital visit included an interview on personal and family medical history, past or current use of drugs, socio-demographic and lifestyle habits information and a dietary questionnaire. The subject underwent a general clinical assessment including anthropometrics measurements and arterial blood pressure measurement on both arms, and a fasting blood sample for biochemical assessment.

3.3.1 Structured Questionnaire

The comprehensive structured questionnaire was developed by a team of clinicians and consisted of different sections.

Section A : Socio-demographic variables

Section A was dedicated to personal and socio-demographic data as: age, sex, employment, marital status, and educational level. For this work purposes, education was classified as two levels: (i) primary school or less and (ii) medium school or high.

Section B: Family history of diseases and use of drugs

Section B included questions related to the major diseases occurred: diabetes, cardiovascular diseases, hypertension, endocrine diseases, respiratory diseases, immune system diseases, hypercholesterolemia, cancer, nervous system diseases, liver diseases, urinary diseases, blood diseases, gastric diseases, osteoporosis.

Clinicians collected data also about drugs assumption: diuretics, anti-hypertensive, cardiovascular, anti-diabetics, insulin, non-steroidal anti-inflammatory, corticosteroids, hormones, therapeutic radiation, chemotherapeutic, hypo-cholesterol, antibiotics, anti-acids, sedatives, vitamins.

Physician ascertained the presence of chronic diseases at baseline during the hospital visit. Participants who self-reported to have a disease or to use drugs for a specific disease were considered as prevalent cases. This baseline information, together with additional information obtained through medical record linkages, allowed us to identify and exclude prevalent cases of major chronic diseases.

Section C: Life-style variables

Section D included questions about life-style related factors such as smoke habits, alcohol intake and physical activity practicing.

Subjects were categorized in three levels: (i) no smokers, (ii) former smokers or (iii) current smokers. For alcohol consumption, subjects were categorized as (i) abstainers, (ii) subjects who consumed up to 30 g of alcohol per day and (iii) subjects who consumed more than 30 g of alcohol per day. Physical activity was derived from a series of questions about activity related to exercise and sports. Physical activity at work was measured by calculating the difference (in h/day) between standing/walking and sitting (31). For this analysis subjects were classified in two categories: (i) physically active (if they were engaged in at least one sport) and (ii) physically inactive.

3.3.2 Laboratory Test

Blood samples were drawn in the morning after overnight fasting. The biochemical assessment included the hematology, glycaemia, lipids (total cholesterol, HDL-cholesterol, triglycerides), urea, creatinine, creatine phosphokinase, sodium, potassium, chlorine, calcium, cholinesterase, lactate dehydrogenase, uric acids, prothrombin, partial time of protrombine, bilirubin, glucose 6 phosphate dehydrogenase, aspartate aminotransferase, alanine aminotransferase, fibrinogen, immunoglobulin (E, G, A, M).

All of these parameters were assessed by means traditional laboratory techniques.

Complete laboratory data were obtained relating to 1361 subjects (740 men and 621 women).

3.3.3 Blood Pressure Measurements

Systolic and diastolic blood pressure measurements were performed by specifically trained operators with the use of a mercury sphygmomanometer following standardized procedures. Subjects with a systolic blood pressure ≥140 mm Hg and/or a diastolic blood pressure ≥90 mm Hg or reporting a clinical diagnosis of hypertension or receiving any diuretics or antihypertensive treatment at baseline were considered hypertensive.

3.3.4 Anthropometric Assessment

The subjects were weighted dressed only in their underclothes, and their height was measured as they stood fully erect without shoes. A tailor’s tape measure was used to record the circumferences of their abdomens (at the height of the waist) and hips (at the height of the upper iliac crest); arm circumference was measured bilaterally in the middle third portion. The skinfold thickness measurements were made bicipitally (vertically on the anterior median line of the arm, midway between the acromion and the bend of the elbow), suprailiacally (vertically on the iliac crest at the level of the axillary media), tricipitallly (vertically on the posterior line of the arm at the same level as the bicipital fold measurement) and subscapularly (laterally about 301 from the vertical, passing through the lower angle of the scapula) (Leite & Nicolosi 2006).

As obesity indices we calculated Body Mass Index (BMI) from person's weight and height (weight (kg) / [height (m)]^2). Complete anthropometric information was obtained relating to 1415 subjects (705 men and 710 women).

3.3.5 Dietary Assessment

Usual dietary intake during the year preceding recruitment was assessed with a FFQ administered by a trained interviewer. This quantitative questionnaire was a modification of Willett's questionnaire of Nurses' Health Study (Willett 1990) adapted for the Italian population. The FFQ used for this study concerned the usual intake of foods during the past year and consisted of three parts.

The first one was a list of 158 foods arranged in thirteen categories: 1-bread/cereal products; 2-eggs/meats, crude pork; 3-fish; 4-dairy products; 5-cheese; 6-vegetables; 7-oli/condiments/sauces; 8-fruits; 9-sweet and candies; 10-beverage/coffee; 11-alcohol; 12- sugar; 13-miscellanous. The amount of foods was assessed with the support of photographic pictures of specific food portion.

For each food picture, participants indicated on a 7- categories frequency scale the number of time a given item was consumed: never, 1-3 times per month; 1-2 times per week; 3-4 times per week; 1 time per day; 2-3 times per day and 4-5 times per day. The amount of foods was assessed by selection of a picture of a food portion.

Complete dietary information was obtained for 1427 subjects (713 men and 714 women).

METHODOLOGICAL NOTE

FFQs are commonly used in epidemiological studies to assess the dietary intake of large populations. The underlying principle of this approach is that the long-term diet (weeks, months, years) or the dietary habit is more important than the intake on a few specific days. Therefore, it may be advantageous to sacrifice precise intake measurements obtainable on 1 or a few days in exchange for more crude information related to an extended period of time.

Moreover, it is typically easier to describe one's usual frequency of consuming a food that to describe what foods were eaten at any specific meal in the past. The FFQ depends from the individual recall and can be self-administered or compiled by a trained interviewer.

The basic FFQ consists of two components: a food list (item) and a frequency response section where subjects report how often each food was consumed. This is a simply or qualitative frequency questionnaire because no additional information of portion is collected.

A second possibility is to specify a portion size as part of the question on frequency. If we are interested to assess the milk intake, the question will include the portion as "How often do you drink a glass of milk?" rather than "How often do you drink milk?". This is called semi-quantitative food frequency. This specification adds clarity to the question for foods that come in a natural unit (as a slice of bread, one egg). A third alternative is to include and additional part for each food in order to describe the usual portion of intake. This information can be assessed describing a typical portion of intake in words and then asking subjects to define their usual portion as multiple of the specified portion; using a food model or providing pictures of different portion sizes as multiple-choice question.

Because diet tends to be reasonably correlated from year to year and in order to take into account for the long term dietary history, researchers ask subjects to describe their frequency of using foods in reference to the preceding year. This provides a full cycle of seasons so that, in theory, the responses should be independent of the time of year (Willet 1990).

3.5 Ascertainment of Life Status and Dataset Definition

In order to trace the personal history of subjects visited and interviewed 20 years ago, a first step was to identify them in the Regional Health Services databases (RHSD) of the ASL Mi1 which cover a wide area of the North-west of Milan hinterland (about 1 million of inhabitants).

Starting from the initial dataset which included for each subject, a personal code (assigned at the enrollment), full name, date and place of birth; a deterministic record linkage was performed between this dataset and the administrative databases of ASL Mi1 until 31 October 2012 (end of follow-up) using full name, sex and date of birth of subjects. From these information it has been possible to obtain fiscal and regional code for each subject using them as personal identifiers codes.

After this first step, of the 1693 subjects who were visited and interviewed at baseline, 1607 subjects with available data on the Regional Registry Office were identified. A further manual check was carried out in order to identify duplicated cases or discrepancies. During this phase two duplicated cases and one case with a discordant birth date were detected.

In the final dataset of 1604 subjects, information on life status (live, migrated, deleted, died), fiscal and regional code, personal identifier code, full name, date and place of birth were included.

Of the 1604 individuals, 30 migrated, 5 were deleted from the registries and 400 died before the end of the follow-up (figure 3).

Five data sources from the RHDS of ASL Mi1 were available:

- the Lombardy Registry Office consisting of the historical lists of residents assisted by the Health System in Lombardy and that stored as identifiers for each person the fiscal code and the regional code;
- the Regional Hospital Discharge databases (1997-2012), which stored all

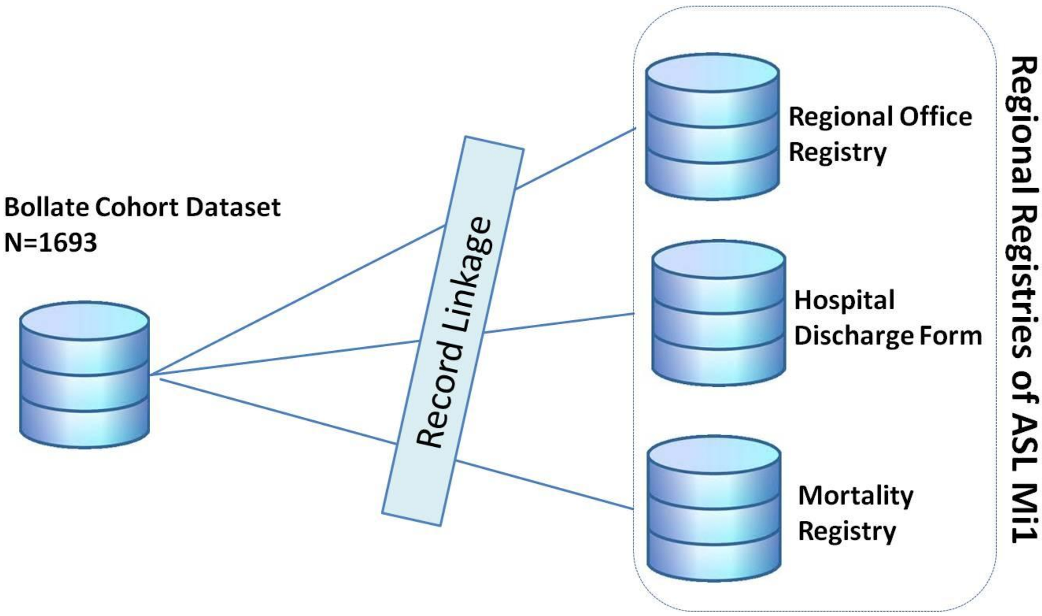
regional patient discharge records and reports clinical information about patient and their hospitalization, admission and discharge dates, the main and five secondary diagnoses (coded according to the ICD 9-CM) and date and type interventions;

- the Pharmacuetical Prescription Registry and the File F Registry (available for the period 2000-2012 and 2002-2012), which respectively contained all prescriptions of drugs administered directly in the outpatient setting and of selected novel high cost drugs administered in the inpatient setting or in Day Hospital (DH). Each record included a drug code, number of prescription, cost, drug description, class, dosage and date of administration, and Chemical therapeutic anatomic classification;
- the Mortality Registry (available for 1999-2012) which included the primary cause of death and date coded according to the ICD 9/10-CM.

The final dataset obtained through the Record Linkage procedure, was useful for three different purposes:

1. to determine the incidence of any diseases of interest and mortality in a cohort of subjects from the enrollment to the end of follow-up (20 years)
2. to study the association of clinical, biochemical and life-style factors with incident diseases and mortality
3. to evaluate this cohort of subjects over time using a large number of data from the Regional Registries

Figure 2. Record Linkage Procedures



3.6 Outcomes Ascertainment

In this work I considered CVD incident cases (fatal and non fatal events) because the high prevalence in this population, and all-causes death as two major endpoints, in order to illustrate the possible use of administrative data.

METHODOLOGICAL NOTE

In the attempt to measure the frequency of disease in a population, it is insufficient merely to record a number of people or the proportion of the population that is affected. It is also necessary to take into account the time elapses before disease occurrs, as well as the period of time during wich events are counted.

Incidence is a measure of disease that allows us to determine a person's probability of being diagnosed with a disease during a given period of time. Therefore, incidence is the number of newly diagnosed cases of a disease over a specific period of time.

An incidence rate is the number of new cases of a disease (incident cases) divided by the number of persons at risk for the disease (event), that is the period of time during which the event was a possibility and would have been counted as an event in the population, had it occurred. The lenght of time is called the person-time contribution the individual. The sum of these person-times over all population members is called the total person-time at risk or the population-time at risk.

Incidence is frequently given as an incidence rate that states the number of people estimated to be diagnosed with the disease for a predefined number of persons, usually 100,000 people. Incidence is often stated as an age-adjusted incidence rate. The number of people who fall into different age groups varies (for example, there are many more 30 to 40 year olds than 80 to 90 year olds). This is referred to as age distribution. Incidence rates can be adjusted to *account for these age distribution differences so that populations can be compared. When these statistics describe particular population groups, they are usually referred to as "specific." Example: The age-specific incidence rate for breast cancer in 20-year-olds to 24-year-olds is 1.5 (per 100,000 persons). Mortality is a term for death. Is used to describe the number of deaths during a specific time period. The mortality rate describes the number of deaths per a a predefined number of persons, usually 100,000 people during a specific time period, which depends by the frequency of cause-specific death, usually one year. As with incidence rates, mortality rates can also be given as age-adjusted mortality rates (Rothman 2008).*

The Regional Hospital discharge database and the Mortality Registry were used to identify cohort members who during follow-up experienced at least one hospitalization for any disease of the circulatory system according to ICD-9 codes 390-4599 recorded as main cause of hospitalization or others associated conditions, or died for CVD as main cause (ICD-9 codes 390-4599 and ICD-10 codes I00-I99). The earliest date of hospitalization or date of death for CVD was considered as that of outcome onset.

Because the multiple admission in hospital for a same patient, for a total of 8512 records, a grouping of hospitalizations for the same person was performed using the following criteria:

- the data was sorted by the fiscal code and date of admission
- aggregating to the first date of hospitalization for CVD.

Each subjects accumulated person-years of follow-up from the date of enrollment to death or emigration, to earliest CVD or 31 October 2012 whichever came first.

We considered prevalent cases (and thus excluded from the analysis) subjects:

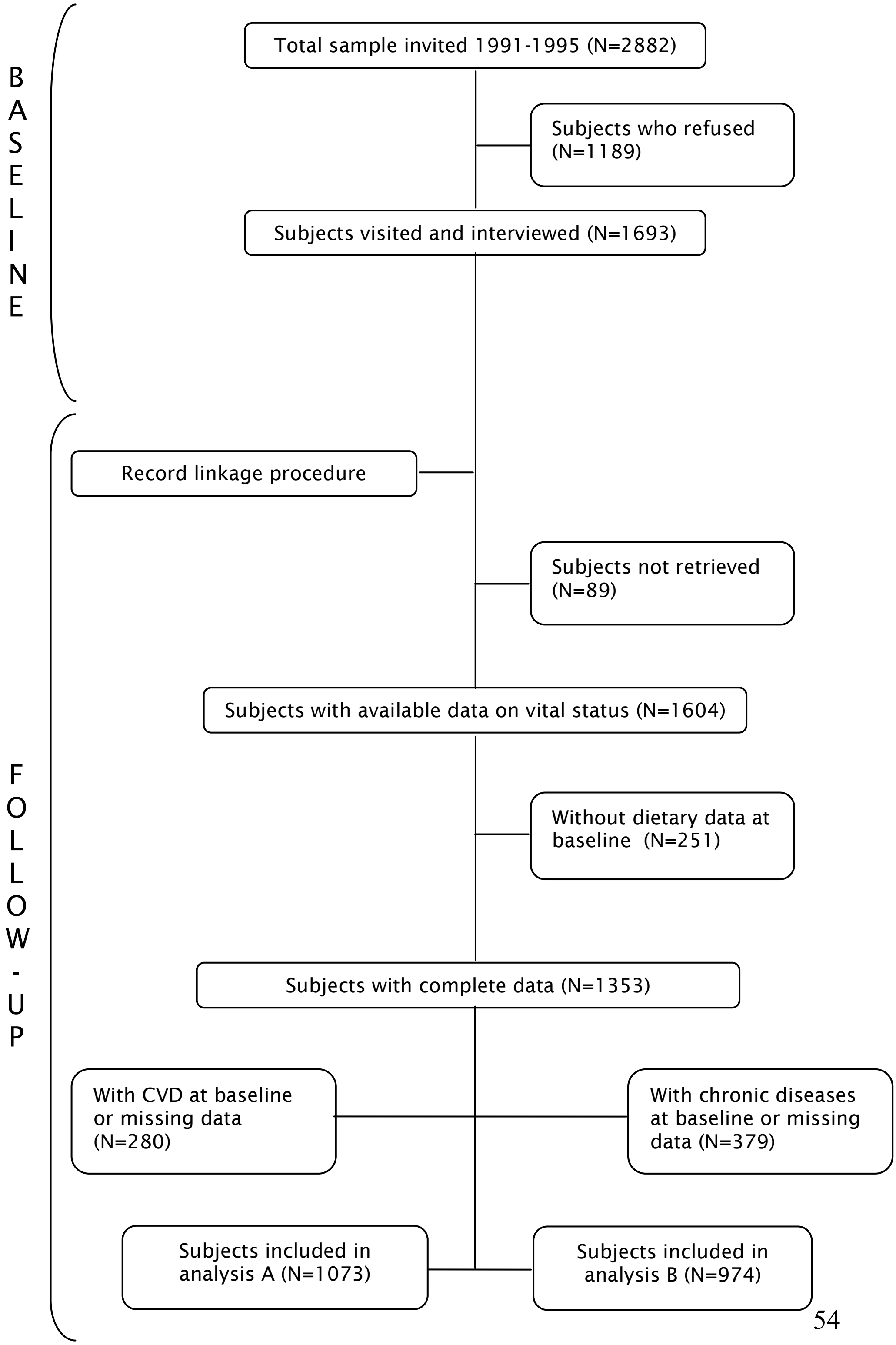
- with CVD follow-up years ≤ 0
- who self-reported to have a CVD or hypertensive subjects.

The final sample analyzed consisted of 1073 subjects (528 men and 545 women).

The cohort was also followed-up for all-causes death using Regional Registry of Mortality.
Causes of death were coded and classified according to the ICD-9/10 CM coding system. ICD-9 codes 390-4599 and ICD-10 codes I00-I99 were used for defining specific-cause death for CVD and ICD-9 codes 140-2399 and ICD-10 codes C00-D99 for cancer (that represented the two main causes of death in the cohort). Mortality from other causes was defined as total death minus the deaths from unknown causes, CVD, and cancer.
Follow-up was the date of enrollment to death or emigration or end of follow-up whichever came first.
Because CVD, diabetes and cancer increase the risk of mortality and to minimize the possibility that diet or lifestyle are changed in response to morbidity, we excluded subjects having at least one of these chronic diseases at baseline.
The final sample analyzed consisted of 974 subjects (483 men and 491 women).

A final record linkage of the cohort dataset with dietary and clinical data collected at baseline, with Hospital discharge and Mortality Registry was developed using fiscal and regional code as personal identifiers codes.

Figure 3. Flow-chart for Selecting Study Sample



3.7 Dietary Data Analysis

Three different approaches to characterize the dietary habits of the subjects were used. First, in terms of percentage of energy form macronutrient and by calculating dietary pattern through *a priori* and *a posteriori* approach.

METHODOLOGICAL NOTE

The analysis between diet and diseases have been traditionally carried out considering diet in terms of its chemical composition, for example nutrient content and alternatively, in terms of foods or foods group.

If the development of a disease is causally related to the intake of a dietary component, the examination of that component is likely to have the greatest power to identify its effect.

Conversely, focus on a specific compound that turns out to have no relation with disease may lead to the erroneous conclusion that diet has no effect. Epidemiological analysis based on foods, as opposed to nutrients, are generally most directly related to dietary recommendations because individuals and institutions ultimately manipulate nutrient intake largely by their choice of foods. Even if the intake of a specific nutrient is convincingly shown to be related to risk of disease, this is not sufficient information on which to make dietary recommendations. Because of the complexity of diet and the potential for interactions among dietary components, approaches that focus on individual foods or nutrients may have some limitation.

Thus, it is now suggested that a holistic dietary approach to disease prevention should be used. Due to their ability to capture the variation in overall food intake in a given population, dietary patterns have been frequently used to describe associations between diet and disease. A dietary pattern is a set of habits regarding consumption of foods and beverages. It is often influenced by environmental or cultural particularities, and by religion. In dietary pattern *analysis, the collinearity of nutrients and foods can be considered as an advantage, instead of a limitation, as the extracted patterns are characterised on the basis of usual consumption of foods and more closely describe real world conditions (Panagiotakos 2008). Dietary pattern are combinations of dietary components (food items, food groups, nutrients, or both foods and nutrients) intended to summarize the total diet or key fatcorsof the diet for a populatin under study. Three approaches have been proposed in the literature to define dietary pattern: hypothesis-oriented/a priori approach, exploratory/a posteriori approach and RRR, which combines characteristics of the exploratory and a priori approach (Edefonti et al. 2009).*

3.7.1 Percentage of Energy from Macronutrient

For each subject the mean daily consumption (grams) of each items was calculated as:

$$PF_j = P_i * Fw$$

PQ_i = usual portion by photographs of food portion

Fw = usual frequency weight of consumption individual food items during the past years.

For Fw it is simplest to assign a weight to 1.0 for once a day, and proportional weights to the other responses, that are "never"=0; "1-3 a month"=0.07; "1-2 a week"=0.21; "3-4 a week"=0.50; "2-3 times a day"=2.5; "4-5 a day"=4.50.

The nutrient foods composition was calculated using the food composition database compiled for epidemiological studies in Italy (Salvini 1998).

Given portion and frequency of each food *j* and nutrient content *i*, the amount of specific nutrient (grams) by single food items (N_{ij}) for each individual who consume the specific food item, was computed as:

$$N_{ij} = (PF_j * N_i)/100g$$

for every subjects K with $K = 1, \ldots\ldots K = j$

N_i = grams of nutrient *i* in 100 g of food *j*.

For each subject total daily nutrient intake was:

$$TN_i = \sum_j N_{ij}$$

Daily intake of carbohydrates (g/day) was multiplied by 3,75 (Kcal in 1 gram of carbohydrate), intake of proteins by 4, and fats intake by 9 to obtain the daily energy derived from each macronutrient. The proportion of total daily energy derived from total carbohydrates (% carbohydrates), fats (% fats), and proteins (% proteins), and ranked participants by tertiles of intake was computed.

3.7.2 *A priori* Dietary Pattern

In *a priori* approach, dietary patterns are defined as indexes or scores built on the basis on existing knowledge about the relationship between food and nutrients and diseases and generally include foods or nutrient supported by current nutrition guidelines, recommendations, and/or a specific dietary composition that is considered healthful (Edefonti, Randi et al. 2009). Several dietary scores or indexes are used to measure the degree in which an individual's diet conforms to specific dietary recommendations, for example, healthy eating index (HEI), Mediterranean Diet (MeDi) score, etc. The possible range of value for each score varies depending on its definition and on the number of individual food items available from each questionnaire. Each subject is directly associated with a score level that expressed the compliance of his/her diet to defined dietary pattern.

The most used hypothesis-oriented/*a priori* approach was the MeDi pattern. The concept of the Mediterranean diet (MeDi) was originally conceived by Ancel Keys, in the Seven Countries Study; the main finding of the prospective evaluation of the study participants was that all-causes and coronary heart disease death rates were lower in cohorts with olive oil as the main dietary fat compared to northern European cohorts (Keys et al. 1986). The identified dietary pattern in these cohorts was characterized by abundance of plant foods: fruits, mainly as the typical after-dinner dessert, vegetables, either as main or side dish, a lot of bread, other forms of cereals, legumes, nuts, and seeds. Olive oil was the principal source of fat. The so-called MeDi also included moderate amounts of dairy products (principally cheese and yogurt), low to moderate amounts of fish and poultry, red meat in low amounts and wine, consumed moderately, during meals (Willet 1995). Thereafter, the beneficial effects of the MeDi on several chronic diseases as cardiovascular disease (CVD), cognitive decline and Alzheimer's Disease, cancer and diabetes have been widely investigated and replicated in many observational studies and in some clinical trials (Vincent-Baudry et al. 2005, Estruch et al. 2006, Serra-Majem et al. 2006, Benetou et al. 2008, Martinez-Gonzalez et al. 2009, Salas-Salvado et al. 2011, Estruch et al. 2013, Lourida et al. 2013). Moreover, there is consistent evidence that MeDi reduces total mortality in Mediterranean and non-Mediterranean populations (Trichopoulou et al. 1995, Trichopoulou et al. 2003, Knoops et al. 2004, Trichopoulou et al. 2005, Lagiou et al. 2006, Mitrou et al. 2007, Trichopoulou et al. 2009, Buckland et al. 2011, Martinez-Gonzalez et al. 2012, McNaughton et al. 2012, Tognon et al. 2014) .

In this work the Mediterranean Dietary Score (MediDietScore) (Panagiotakos et al. 2007) was calculated. Assessment was based on the weekly consumption of 11 food groups: cereals, fruit, vegetables, legumes, potatoes, fish, meat and meat products, poultry, full fat dairy products (like cheese, yoghurt, milk), as well as olive oil and alcohol intake. For the consumption of items presumed to be close to this pattern (i.e., those suggested on a daily basis or more than 3 portions per week; i.e., non-refined cereals, fruits, vegetables, legumes, olive oil, fish and potatoes), when someone reported no consumption was assigned score 0 and scores 1 to 5 for rare to daily consumption. On the other hand, for the consumption of foods presumed to be away from this diet pattern (i.e., rare or monthly consumption; meat and meat products, poultry and full fat dairy products) were assigned scores on a reverse scale (i.e., from 5 when someone reported no consumption to 0 when they reported almost daily consumption). Especially for alcohol intake, for consumption of less than 300 ml of alcohol per day was assigned score 5, score 0 for no consumption or for consumption of >700 ml per day and scores 4 to 1 for consumption of 600–700, 500–600, 400–500 and 300–400 ml per day (considering that 100 ml have 12 g ethanol concentration). Higher values of this diet score indicate greater adherence to the MeDi.

3.7.3 A *posteriori* Dietary Pattern

In *a posteriori* approach, empirically derive dietary pattern are not defined *a priori*, and do not depend on how the authors define a healthful pattern. Multivariate statistical analysis are applied directly to the data under consideration using principal component analysis (PCA), cluster analysis (CA) and exploratory factor analysis (FA) (Edefonti, Randi et al. 2009). The aim of the PCA is to reduce the dimensionanality of data by transforming an original larger set of correlated foods or nutrients into a smaller and more easily interpretable set of uncorrelated variables known as posteriory dietary pattern. PCA uses simple matrix algebra to identify principal components based on the covariance/correlation matrix of the input variables (nutrients or foods). The resulting components are linear combinations of the original variables with suitable weights (loadings) that explain as much of the variation in the original variables as possible. The magnitude of each loading for a given component measures the importance of the corresponding food item/nutrient to that component, as each loading is proportional to the correlation coefficient between food item/nutrient and component. A continuous summary score is derived for each subject and for each factor indicating the degree to which a subject's diet conforms to each of the identified dietary patterns.

Dietary patterns identified by the exploratory techniques are meant to reflect effective dietary behaviour (Edefonti, Randi et al. 2009).

For the current analysis, in order to identify major dietary patterns based on selected 11 food groups for calculating the MedDietScore, exploratory PCA was performed.

Number of factors to retain were chosen combining factor eigenvalue>1 and interpretability. Varimax rotation to the factor-loading matrix was applied. Food groups with absolute rotated factors loading grater than 0.50 on a given factor were used to name the factors. The derived factors were labeled on the basis of the data interpretation. The factor-loading matrix and explained variance for the major dietary patterns identified by PCA, are shown in Table 5.

3.8 Statistical Analysis: Multivariate Survival Analysis Using Cox's Regression Model

Diseases develop and progress in time, thus the description of the course in time is an important issue in the characterization of diseases including their prognosis and the effects of therapies. Because a description of the course of disease may be complex, the problem has been dealt by analyzing for each individual, the time from a defined starting point of observation, e-g- the time of diagnosis or randomization in a controlled clinical trial or for observational studies, the beginning of exposure to some risk factors or when the subject is enrolled, to the occurrence of an *endpoint* (event of interest).

The survival analysis is useful whenever the researcher is interested not only in the frequency of occurrence of a certain type of event, but also in the time process underlying such occurrence. A distinctive characteristics of survival data is that the event of interest may not be observed on every experimental unit. Since investigations have a limited duration, some subjects may not yet have had the event, but are still "alive" at the end of the investigation. Other subjects, while alive, may be unwilling or unable, for various reasons to continue participating in the study and providing follow-up information. These subjects are called *lost to follow-up* or *drop-outs*. Such "incomplete" survival times from the starting point to the latest observation, is called *censored survival times* and hold the information that the event did not occur while the individual was being observed. While in clinical trials the beginning of observation commonly coincides with the beginning of exposure to treatment, this may not be the case in observational studies. Data where the observation beging some time after the time origin are said to be *left trunked*. An analougus problem could arise in studying the natural history of a disease form a hospital database in which patients entry is at the first contact with the hospital and not necessarily at first diagnosis of the disease.

In analyzing survival data, two functions that are dependent on time are of particular interest: the survival function and the hazard function. The survival function $S(t)$ is defined as the probability of surviving at least to time t . The hazard function $h(t)$ is the conditional probability of dying at time t having survived to that time. The graph of $S(t)$ against t is called *survival curve*. The Kaplan–Meier method can be used to estimate this curve from the observed survival times without the assumption of an underlying probability distribution. The method is based on the basic idea that the probability of surviving k or more periods (i.e. usually one day), from entering the study is a product of the k observed survival rates for each period (i.e. the cumulative proportion surviving), given by the following:

$$S(K) = p_1 * p_2 * p_3 * \ldots .* p_k$$

p_1 is the proportion surviving the first time period, p_2 is the proportion surviving beyond the second period conditional on having survived up to the second period, and so on. The proportion surviving period *i* having survived up to period *i* is given by:

$$p_1 = \frac{r_i - d_i}{r_i}$$

Where r_i is the number alive at the beginning of the period and d_i the number of deaths within the period (Bewick et al. 2004).

Comparison of two survival curves can be done using a statistical hypothesis test called *log rank test*. It is used to test the null hypothesis that there is no difference between the population survival curves (i.e. the probability of an event occurring at any time point is the same for each population) but it does not allow other explanatory variables to be taken into account.

Because the population is made heterogenous by risk factors in epidemiology or treatments and prognostic factors in clinical research, survival models are extended to include regression variables, called covariates, measured on the individuals in the study. These models enable to investigate the role of these covariates in modifying the endpoint and to take into account counfounding factors in the effect estimate of a treatment or an exposure.

The most commonly used multivariate approach for analysing survival time data include Cox's proportional hazards model. It is an analogous to a multiple regression model which describes the relation between the event incidence, as expressed by the hazard function and a set of explanatory variables. In this model, the dependent variable is the hazard function at a given time. Briefly, the hazard is the instantaneous probability of dying (or event of interest) given that patients have survived up to a given point in time, or the risk for death at that moment.

Mathematically, the Cox model can be written as:

$$\ln \mathrm{h}\,(\mathrm{t}) = \ln \mathrm{h}_0(\mathrm{t}) * \exp\,(\mathrm{b}_1\mathrm{x}_1 + \cdots + \mathrm{b}_\mathrm{p}\mathrm{x}_\mathrm{p})$$

Where $\mathrm{h}(\mathrm{t})$ is the hazard function at time *t* that is dependent on a set of *p* explanatory variables (x_1, x_2 … x_p) whose impact is measured by the size of the respective coefficients (b_1, b_2 … b_p). The regression coefficiens give the proportional change than can be expected in the hazard, related to changes in the explanatory variables. The term $\mathrm{h}_0(\mathrm{t})$ is called the *baseline* or *underlying hazard function* and corresponds to the probability of dying when all the explanatory variables are zero. The *t* in the function remainds that the hazard may vary over time. The Cox model is essentially a multiple linear regression of the logarithm of the hazard on the variable x_1, with the baseline hazard being an intercept term that varies with time. The covariates act multiplicatively on the hazard at any point in time, and this provides us with the key assumption of the proportional hazard model: the hazard of the event in any group is a constant multiple of the hazard in any other. This assumption implies that the hazard curves for the groups should be proportional and cannot cross. In other words, if an individual has a risk of death at some initial time point that is twice as high as that of another individual, then at all later times the risk of death remains twice as high. Proportionality implies that the quantities $\exp(b_1)$ are called *hazard ratios*. A hazard ratio above 1 indicates a explanatory variable that is positively associated with the event probability, and thus negatively associated with the length of survival.

In summary, survival analysis involves the consideration of the time between a fixed starting point and a terminating event. The statistical methods based on time-dependent analysis, are the appropriate tools for analyzing these kind of data because allow us to estimate not only the effects of explanatory variables but also the effect of time. For univariate analysis, both the Kaplan– Meier product limit method and the log rank test to investigate differences between groups, describe the survival with respect to the factor under investigation, but ignore the impact of any others. Cox model is a statistical technique for exploring the relationship between the survival and several explanatory variables providing an estimate of the exposure effect on survival after adjustment for other explanatory variables. This approach allows to estimate the hazard (or risk) of death of an individual, given the prognostic variables. A positive regression coefficient for an explanatory variable means that the hazard is higher, and thus the prognosis worse. Conversely, a negative regression coefficient implies a better prognosis for patients with higher values of that variable (Cox 1972, Christensen 1987, Bradburn et al. 2003, Bewick, Cheek et al. 2004, Marubini 2004, Freeman 2008).

3.8.1 Association between Dietary Intake and Endpoints

In order to assess the association of dietary intake with CVD and all-causes death risk, Cox Proportional Hazard Regression models were performed and the RRs were estimated as HRs with 95%CI.

For the first analysis, the % of energy derived from macronutrients was categorized in tertiles of intake and subjects in the lowest tertile were considered as the reference group. The analyses were performed by using 2 models: the first one was adjusted for age at baseline, sex, educational level (2 category, primary school or less as the reference category), BMI (continuous variable, Kg/m^2), smoking status (2 categories, never smokers as the reference category), physical activity (2 categories, physically active as the reference category), alcohol intake (3 categories, abstainers as the reference category), and total daily energy intake (continuous, Kcalories), for CVD only I adjusted also for hypertension (2 categories, no as reference category); the last one included carbohydrates, fats and proteins in the same model.

The MedDietScore was categorized in tertiles of intake considering subjects in the lowest tertiles of adherence as the reference group. The analyses were performed by using 2 models: age- and sex-adjusted model and fully adjusted model. The following variables were used for the adjustment: educational level, smoking status, physically active, BMI, total daily energy intake, for CVD only we adjused also for hypertension. The adjusted models were also performed considering adherence to MedDietScore as continuous variables.

Besides the relationship between the MedDietScore and endpoints, we investigated the effect of the single components of the diet score. The median of each component was used as cut-off point considering the value below the 50° percentile as reference category. Firstly, I evaluated the effect of every of the 11 components including age, sex, education, BMI, physical activity, smoking status, (hypertension for CVD only) and energy intake in the model.

The median of each factors obtained with the PCA was used as cut-off point considering the value below the 50° percentile as reference category. The HRs for each factors were calculated by using 2 models: an adjusted model and a mutually adjusted model. The mutually adjusted models included all factors simultaneously in the models.

In order to evaluate the combined effect of lifestyle factors, a score was calculated by adding the individual values for *a priori* MedDietScore, physical activity level and smoking status. Individuals scored 0 point if they belonged to the low tertile of Mediterranean diet and scored 1 point if they belonged to the medium or highest tertile. For smoking, individuals were considered to be at low risk (1 point) if they had never smoked and at high risk (0) if they were former or current smokers. Individuals physically inactive were considered the high-risk group (0 point) and subjects physically active were considered at low risk (1 point). The score ranged from 0 to 3. Subjects who scored 0 points on the lifestyle score were considered the reference group. The associations between healthy lifestyle score and incident CVD and all-causes death were evaluated in multivariate-adjusted Cox models controlling for age, sex, education, BMI and total daily energy intake (for CVD only, hypertension).

3.9 References

Benetou V, Trichopoulou A, Orfanos P, et al. (2008). Conformity to traditional Mediterranean diet and cancer incidence: the Greek EPIC cohort. Br J Cancer **99**(1): 191-195.

Bewick V, Cheek L & Ball J (2004). Statistics review 12: survival analysis. Crit Care **8**(5): 389-394.

Bradburn MJ, Clark TG, Love SB, et al. (2003). Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. Br J Cancer **89**(3): 431-436.

Buckland G, Agudo A, Travier N, et al. (2011). Adherence to the Mediterranean diet reduces mortality in the Spanish cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC-Spain). Br J Nutr **106**(10): 1581-1591.

Christensen E (1987). Multivariate survival analysis using Cox's regression model. Hepatology **7**: 1346-1358.

Cox D (1972). Regression models and life tables. J Roy Statist Soc. Series B (Methodological) **34**: 187–220.

Daniel J (2012). Sampling Essentials. Practical Guidelines for Making Sampling Choices. SAGE Publications, Inc

Edefonti V, Randi G, La Vecchia C, et al. (2009). Dietary patterns and breast cancer: a review with focus on methodological issues. Nutr Rev **67**(6): 297-314.

Estruch R, Martinez-Gonzalez MA, Corella D, et al. (2006). Effects of a Mediterranean-style diet on cardiovascular risk factors: a randomized trial. Ann Intern Med **145**(1): 1-11.

Estruch R, Ros E, Salas-Salvado J, et al. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med **368**(14): 1279-1290.

Freeman JW, SJ. Campbell, MJ. (2008). How to display data. Oxford: Blackwell BMJ Books.

Grimes DA & Schulz KF (2002). An overview of clinical research: the lay of the land. Lancet **359**(9300): 57-61.

Keys A, Menotti A, Karvonen MJ, et al. (1986). The diet and 15-year death rate in the seven countries study. Am J Epidemiol **124**(6): 903-915.

Knoops KT, de Groot LC, Kromhout D, et al. (2004). Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA **292**(12): 1433-1439.

Lagiou P, Trichopoulos D, Sandin S, et al. (2006). Mediterranean dietary pattern and mortality among young women: a cohort study in Sweden. Br J Nutr **96**(2): 384-392.

Leite ML & Nicolosi A (2006). Lifestyle correlates of anthropometric estimates of body adiposity in an Italian middle-aged and elderly population: a covariance analysis. Int J Obes (Lond) **30**(6): 926-934.

Lourida I, Soni M, Thompson-Coon J, et al. (2013). Mediterranean diet, cognitive function, and dementia: a systematic review. Epidemiology **24**(4): 479-489.

Martinez-Gonzalez MA, Bes-Rastrollo M, Serra-Majem L, et al. (2009). Mediterranean food pattern and the primary prevention of chronic disease: recent developments. Nutr Rev **67 Suppl 1**: S111-116.

Martinez-Gonzalez MA, Guillen-Grima F, De Irala J, et al. (2012). The Mediterranean diet is associated with a reduction in premature mortality among middle-aged adults. J Nutr **142**(9): 1672-1678.

Marubini EV, G. (2004). Analysing Survival Data from Clinical Trials and Observational Studies. Wiley-Interscience; 1 ed.

McNaughton SA, Bates CJ & Mishra GD (2012). Diet quality is associated with all-cause mortality in adults aged 65 years and older. J Nutr **142**(2): 320-325.

Mitrou PN, Kipnis V, Thiebaut AC, et al. (2007). Mediterranean dietary pattern and prediction of all-cause mortality in a US population: results from the NIH-AARP Diet and Health Study. Arch Intern Med **167**(22): 2461-2468.

Nicolosi A (1997). Le indagini di popolazione sulla prevalenza delle principali malattie oculari in Italia. In L'Epidemiologia oftalmica in Italia. Cerulli L, Miglior M and Ponte F. Roma.

Panagiotakos D (2008). α-priori versus α-posterior methods in dietary pattern analysis: a review in nutrition epidemiology. Nutrition Bulletin **33**: 311–315.

Panagiotakos DB, Pitsavos C, Arvaniti F, et al. (2007). Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore. Prev Med **44**(4): 335-340.

Rothman KG, S. Lash, TL. (2008). Modern Epidemiology. 3rd edn. Philadelphia: Lippincott Williams & Wilkins.

Salas-Salvado J, Bullo M, Babio N, et al. (2011). Reduction in the incidence of type 2 diabetes with the Mediterranean diet: results of the PREDIMED-Reus nutrition intervention randomized trial. Diabetes Care **34**(1): 14-19.

Salvini SP, M. Gnagnarella, P. (1998). Banca dati di composizione degli alimenti per studi epidemiologici in Italia. Milano: Istituto Europeo di Oncologia.

Sempos CT, Liu K & Ernst ND (1999). Food and nutrient exposures: what to consider when evaluating epidemiologic evidence. Am J Clin Nutr **69**(6): 1330S-1338S.

Serra-Majem L, Roman B & Estruch R (2006). Scientific evidence of interventions using the Mediterranean diet: a systematic review. Nutr Rev **64**(2 Pt 2): S27-47.

Tognon G, Lissner L, Saebye D, et al. (2014). The Mediterranean diet in relation to mortality and CVD: a Danish cohort study. Br J Nutr **111**(1): 151-159.

Trichopoulou A, Bamia C & Trichopoulos D (2009). Anatomy of health effects of Mediterranean diet: Greek EPIC prospective cohort study. BMJ **338**: b2337.

Trichopoulou A, Costacou T, Bamia C, et al. (2003). Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med **348**(26): 2599-2608.

Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, et al. (1995). Diet and overall survival in elderly people. BMJ **311**(7018): 1457-1460.

Trichopoulou A, Orfanos P, Norat T, et al. (2005). Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ **330**(7498): 991.

Vincent-Baudry S, Defoort C, Gerber M, et al. (2005). The Medi-RIVAGE study: reduction of cardiovascular disease risk factors after a 3-mo intervention with a Mediterranean-type diet or a low-fat diet. Am J Clin Nutr **82**(5): 964-971.

Willet W (1995). Mediterranean diet pyramid: a cultural model for healthy eating. Am J Clin Nutr **61,**: S1402S-S1406.

Willet WC (1990). Nutritional Epidemiology, 2nd ed. Oxford University Press, Inc.

Willett WC (1990). Nurses' Health Study Dietary questionnaire: Appendix 5-1. In Nutritional Epidemiology, 2nd ed. Oxford University Press, Inc.

4.RESULTS

4.1 Descriptive Analyses

Baseline characteristics of Italian subjects with complete data are presented in table 1.

The mean age of males and females was 57.5 and 56.2 years, respectively. The BMI values were almost similar in the two groups. Males were more educated, more frequently former or current smokers and heavy drinkers and slightly more physically active compared with females. The 17.7% of males and the 15.3 of females had self-reported a CVD at baseline. Males were less afflicted by cancer than females, but had diabetes and hypertension more frequently than women. The 70% of males were in the medium to highest tertile of the MedDietScore compared to the 51% of the women. Males assumed less energy from carbohydrates, proteins and fats than females. The 34% of males compared with the 46% of females had a lifestyle score-point ≥2.

Table 2 reports the baseline characteristics and number of deceased by incident CVD. A total of 1073 subjects without prevalent CVD, were included in the analysis, of them after a mean of 13.2 years, 530 subjects developed at least one CVD during the follow-up period and 543 were free from the disease.

As expected, the mean age of cases was higher compared with controls, 58.5 vs 53.8 years respectively. Cases had a slightly higher BMI values, were more frequently males, slightly less educated, former or current smokers and heavy drinkers, more physically inactive and with higher hypertension, compared with control. The 61.9% of cases and the 59.7% of controls were in the medium to highest tertile of the MedDietScore. There were no differences in the percentages of energy from macronutrients intake between the two groups, the 39% of cases compared with the 43% of controls had a score-point ≥2. The 32.6% of cases were died during the follow-up period in comparison with the 12.2% of controls.

Baseline characteristics by causes of death and number of deaths are showed in table 3.
A total of 193/974 deaths, during a mean of 17.4 years of follow-up, had occurred. The main cause of death was cancer, followed by CVD and other causes. Data are reported for the 129 subjects with available causes of death.
Among males 131 deaths occurred since baseline, the 9.3% for cancer, the 6.2% for CVD and the 3.3 for other causes. The mean age of total deaths was 62.4 years and the BMI was 27.4. Among persons less educated, the 21.9% died during follow-up. Subjects physically inactive represented the 78.5% of the total sample and of them the 21.6% died for all-causes. Among former or current smokers and heavy drinkers, about the 23% died for all-causes.
Considering subjects who were in the medium to highest tertile of the MedDietScore 113 died for all-causes. The mean energy % of carbohydrates, proteins and fats in the total sample was respectively about 50%, 15% and 27% without differences among specific causes of death.
Among people who had a score-point ≥2, 64 died.

In order to exclude that the subjects analyzed were selected with particular characteristics that could bias the quantification of the investigated association, I compared the subjects retrieved with the record linkage (n=1604) with un-retrieved people (n=89). The two populations differed only for smoking habits, the un-retrieved persons were more smokers than people retrieved.
Furthermore, I compared the subjects included in the analysis A (n=1073) with people excluded for the presence of prevalent CVD (n=280). Subjects excluded were significantly less educated and as expected, with higher comorbidity and mortality than persons included in the analyses. Finally, in the analysis B, subjects excluded for the presence of previous chronic diseases (n=379) were significantly older, with higher BMI, less educated, more physically inactive

and with higher mortality in comparison with subjects included in the analyses (n=974) (data not shown).

The food grouping used for both the *a priori* and *a posteriori* analysis, is showed in table 4.

In table 5 are reported the factor-loading matrix and explained variance for the four major dietary pattern emerged with the use of the PCA that overall explained the 53.5% of the total variability of the dietary intake of the population in analysis.

Each factor has been sorted according to the proportion of explained variance by each one.

Each principal component represents a dietary factor whose name may be assigned by assessing the weights, i.e. how each nutrient contributes to the composition of that dietary factor. In order to assign a name and facilitate the interpretation of the factors, the absolute values <0.20 have been omitted, and absolute values >0.50 have been highlighted in bold style.

The first factor named ‘Plant-foods' pattern was characterized by high intake of vegetables, potatoes and legumes; the second named ‘Animal foods’ pattern was rich in meat, fish and poultry; the ‘Dairy products and alcohol” pattern was characterized by high fats milk, yogurt and cheese and alcoholic beverages, and the ‘Olive oil’ pattern that included olive oil only.

4.2 Associations between Dietary Intake and CVD Incidence

Table 6 reports the association between the percentage of energy derived from macronutrients intake and CVD incidence. In the adjusted models the risk of CVD was near the one for the highest % carbohydrates tertile (HR 0.98, 95%CI 0.79-1.22), whereas a risk increased at medium (HR 1.45, 95%CI 1.18-1.79) and high % of protein (HR 1.21, 95%CI 0.97-1.52) was observed. The highest % fats tertile (HR 0.82, 95%CI 0.66-1.02) was associated with a borderline reduction of CVD risk. In mutually adjusted model which included % of carbohydrates, proteins and fats in the same model, the HR for CVD was non significantly reduced for highest % carbohydrates tertile (HR 0.83, 95%CI 0.64-1.09), an increase of risk was confirmed for medium and high intake of energy from proteins (respectively HR 1.48, 95%CI 1.20-1.83; HR 1.25, 95%CI 1.00-1.57) and the highest % fats tertile remained inversely associated with CVD, reaching the statistically significance (HR 0.72, 95%CI 0.55-0.94).

In table 7 are presented the proportional HRs with 95% CI for CVD incidence in relation to physical activity, smoking status, tertiles of MedDietScore, and healthy lifestyle score.

In age- and gender adjusted model, high adherence to MeDi (HR 0.88, 95% CI 0.85-1.05), engaging in physical activity (HR 0·75; 95% CI 0·60, 0·93) and being never smokers (HR 0·85; 95% CI 0·71, 1·03) were factors inversely associated with CVD risk, although not all the estimates were statistically significant.

Keeping the subjects with a zero healthy lifestyle score as the reference group, the age- and sex-adjusted HRs estimated in the 1-score, the 2-score and the 3-score groups were 0·98 (95% CI 0·75, 1·26), 0·83 (95% CI 0·63, 1·08), and 0·65 (95% CI 0·42, 1·02) respectively.

In multivariate models, adjusted for age, sex, educational level, BMI, hypertension and energy intake, a high adherence to the MeDi less reduced the risk of CVD and the HR was not statistically significant (HR 0·92; 95% CI 0·74-1·14), physically active subjects and never smokers had respectively a 24% (HR 0·76: 95% CI 0·70, 0·95) and a 15% (HR 0·85; 95% CI 0·70, 1·02) lower risk of CVD. Considering MedDietScore as continuous variable, a non-significant 2% reduction of CVD risk for each one-point increase was observed. In the fully adjusted HRs, the estimated in the 1-score, the 2-score and the 3-score groups were 0·96 (95% CI 0·74, 1·25), 0·86 (95% CI 0·65, 1·13), and 0·65 (95% CI 0·41, 0·99) respectively. The test for linear trend was statistically significant with $p < 0.035$. Cumulative survival curves by healthy lifestyle score are presented in Figure 4.

Table 8 reports associations of the 11 components of the MedDietScore with CVD incidence, contrasting high and low consumption and controlling for potential confounders. In the mutually adjusted model for all components of MedDietScore, high consumption of potatoes and olive oil was inversely associated with CVD incidence (respectively HR 0·85; 95% CI 0·70, 1·02 and HR 0·75; 95% CI 0·63, 0·89), high intake of alcoholic beverages increased the risk of CVD (HR 1.24, 95%CI 1.02-1.52). No significant association was seen between CVD risk and the other components.

CVD risks associated with the major 4 factors extracted from the PCA contrasting high and low intake, are showed in table 9. After control for age, sex, educational level, BMI, energy intake, smoking status, physical activity, alcohol intake, hypertension and energy intake, and including all pattern in the models. An inverse association between 'Plant-foods', 'Dairy products and alcohol’ and 'Olive oil' patterns, was observed (respectively, HR 0.85, 95%CI 0.71-1.02; HR 0.79, 95%CI 0.65-0.97 and HR 0.73, 95%CI 0.61-0.87).

4.3 Association between Dietary Intake and All-causes Death

Table 10 reports the association between the percentage of energy from macronutrients and all-causes death. In the adjusted models the risk of death was reduced for the highest % carbohydrates tertile (HR 0.77, 95%CI 0.53-1.12), in contrast the risk was increased at higher % of protein (HR 1.32, 95%CI 0.92-1.88) and % fats (HR 1.23, 95%CI 0.86-1.77) tertiles. In multivariable model including carbohydrates, fats and proteins in the same model, the results did not change significantly. None estimates reached the statistical significance.

In table 11 are presented the proportional HRs with 95% CI for all-causes death in relation to physical activity, smoking status, tertiles of MedDietScore and lifestyle score.

In age- and gender adjusted model, high adherence to MeDi (HR 0.64, 95% CI 0.45-0.91), engaging in physical activity (HR 0·56; 95% CI 0·38, 0·84) and being never smokers (HR 0·71; 95% CI 0·51, 0·97) were factors significantly inversely associated with all-causes death.

Keeping the subjects with a zero healthy lifestyle score as the reference group, the age- and sex-adjusted HRs estimated in the 1-score, the 2-score and the 3-score groups were 0·64 (95% CI 0·43, 0·89), 0·46 (95% CI 0·31, 0·70), and 0·30 (95% CI 0·13, 0·67) respectively.

In multivariate models, adjusted for age, sex, education, BMI, and energy intake, high adherence to the MeDi significantly reduced the risk of death (HR 0·62; 95% CI 0·43-0·89), physically active subjects and never smokers had respectively a significant 46% (HR 0·54: 95% CI 0·36, 0·82) and 29% (HR 0·71; 95% CI 0·51, 0·98) lower risk of death. Considering MedDietScore as continuous variable, a significant 5% reduction of all-causes death risk for each one-point increase was observed.

In the fully adjusted HRs, the estimated in the 1-score, the 2-score and the 3-score groups were almost similar with a reduction of 34% (HR 0·61, 95% CI 0·41, 0·89), 56% (HR 0·44, 95% CI 0·29, 0·67), and 73% (HR 0·27, 95% CI 0·12, 0·61) respectively. The test for linear trend was statistically significant with $p < 0{\cdot}001$. Cumulative survival curves by healthy lifestyle scores are presented in Figure 5.

Table 12 reports associations of the 11 components of the MedDietScore with all-causes death, contrasting high and low consumption and controlling for potential confounders. High consumption of fruits was inversely associated with all-causes death (HR 0·71; 95% CI 0·53, 0·96); the estimates did not change when all components were included simultaneously in the model (HR 0·70; 95% CI 0·51, 0·95). Death risk was not singularly associated with the other components.

All-causes death HR associated with the major 4 factors extracted from the PCA contrasting high and low intake, are showed in table 13. After control for age, sex, educational level, BMI, energy intake, smoking status, physical activity, and energy intake and including all patterns in the model, a borderline inverse association between 'Plant-foods' pattern and risk of death was observed (HR 0.81, 95%CI 0.60-1.09) whereas I did not find significant association with the other dietary patterns.

In order to minimize the potential bias due to undiagnosed chronic diseases at enrolment, a sensitivity analysis in the cohort of 974 subjects was carried-out excluding subjects who died in the first two years of follow-up (n=8) but I did not find differences in the estimated relative risks (data not shown).

5.DISCUSSION

The aims of the study were to reconstruct the history of a cohort of subjects who were visited and interviewed 20 years ago, to estimate the incidence of major diseases that occurred during this period and mortality for all causes, and to study the relationship between dietary habits and other lifestyles factors in middle age, and risk of CVD and death.

These objectives have been achieved through the use of administrative data that allowed to trace the personal history of each subject and put it in connection with the information collected prior to baseline.

The study involved the application of algorithms in order to identify first of all, the subjects of the cohort in the Regional Health Services Databases (RHSD) of the ASL Mi1. After the deterministic record linkage procedure, we have been able to retrieve the 95% of subjects recruited during the period 1991-1995. Furthermore all persons who experienced at least one hospitalization for CVD recorded as main cause of hospitalization (or others associated conditions), or who died for CVD as the primary cause, were identified until the end of October 2012 by means of a record linkage with data from the Regional Hospital discharge database and the Mortality registries.

Once defined the main endpoints of the study, CVD incidence and all-causes death, I classified and characterized the dietary habits during middle aged, with different approaches and then I evaluated the role of diet on these long-term outcomes.

Now, we see in detail the relevant conclusions that can be drawn about this work.

5.1 The use of Record Linkage in Observational Studies: Problems and Solutions

The practice of linking to administrative databases data from specific epidemiological studies is not so spread and, to my knowledge, very few studies have been carried out combining historical cohort and administrative data to evaluate the effects of life-styles in mid-life on long term health outcome.

The main evidence of the possible use of administrative data for these purposes focusing on the evaluation of the relationship between lifestyle in mid-life and outcomes, as mortality and rarely incidence of diseases, come mainly from Northern Europe (Osler et al. 2001, Brunner et al. 2008), Mursu et al. (2014) (Osler, Heitmann et al. 2001, Brunner, Mosdol et al. 2008, Mursu, Virtanen et al. 2014), USA (Marmot & Brunner 2005, Mitrou et al. 2007), Australia (Harriss et al. 2007) and from the cohort enrolled in the EPIC Study (Buckland et al. 2011). These studies are characterized by larges sample size but moderate follow-up duration.

At current date, in Italy the main findings on this issue, derive from Seven Country Study (Keys et al. 1986), (Huijbregts et al. 1997, Menotti et al. 2014), from the ILSA (Solfrizzi et al. 2005) and from the prospective cohort EPIC Study (Masala et al. 2007), and were carried out mainly for studying mortality. Similarly to the present study, the EPICOR Study performed on 29,689 women during a mean of 7.85 years, evaluated the effect of consumption of leafy vegetables and olive oil on the incidence of fatal and nonfatal coronary heart diseases identified by a passive follow-up, the results showed an inverse association between the increase consumption of these foods and the risk of CHD (Bendinelli et al. 2011).

The use of administrative data is becoming more and more frequent in an overview of the scientific literature and challenges in conducting observational studies based on administrative databases are currently of great interest.

In this study the procedure of matching wide information collected from a cohort enrolled in 1991-1995 with outcomes of interest from Regional registries after twenty years of passive follow-up achieved the remarkable result of identifying 95% of the original subjects. This goal was reached applying a deterministic record linkage that used the couple fiscal code/regional code as key, and that allowed to obtain the most certain as possible subject identification and to minimize the false positive rate.

The use of administrative data is increasingly growing and, despite the fact that they have been developed with administrative aims (costs, healthcare performance), they represent a wide, even if raw, source of information also for other aims.

This large amount of available information, resulting from the potentiality and the diffusion of the ICT, however needs to be adequately integrated to produce quality knowledge in many fields. This issue clearly can be extended also to the field of scientific knowledge, with the integration of clinical, genetic, socio-demographic, laboratory data from different sources.

In this study it was possible to almost completely merge epidemiological data collected twenty years ago with the ones coming from a quality (in terms of reliability, accuracy, completeness, formal correctness) source as the informative systems nowadays are.

In general I think that the approach of designing epidemiological studies in which expositions of interest are accurately collected with a view to access administrative data for long-term health outcomes represent a great opportunity for the future, both for quality results and for lower costs or resources effort.

In datasets deriving from such a linking (as well as when using administrative data in general), attention must be paid to well-known concerns.

Adopted algorithms to query administrative data should be validated and refined using data coming from a real clinical setting or other studies.

The correct identification of the subjects is an important issue to be managed (Corrao 2013). Other important concerns are about i) controlling for confounding, classification errors specifically ii) exposure and iii) disease/outcome misclassification. In studies like the present one, in which expositions and confounders are accurately and directly collected at baseline, issues i) and ii) can be overcome. In other studies that lack this data collection, analytic strategies can be adopted to account for confounding and for the effects of possible exposure misclassification in biasing the results (Greenland 1996, Corrao 2013)

Errors that can occur in measuring outcome lead to classification bias, that is (i) identifying subjects as having experienced the disease outcome when they have not (or being exposed to a drug when they are not), or (ii) not having experienced the outcome when they experienced it (or not being exposed when they are). Classification bias is further categorized as differential or non-differential .

Differential misclassification is present when the likelihood of outcome misclassification differs between exposed and unexposed (or exposure misclassification differs between subjects who experience outcome and those who do not). Non-differential misclassification occurs when the likelihood of misclassification is the same across either the exposed or outcome groups. Errors in diagnosis coding or due to mistyping are typical examples of non-differential misclassification (Greenland 1996, Corrao 2013).

The direction of misclassification regarding measures of association will depend on the type of misclassification (Greenland 1996, van Walraven et al. 2011); when it is differential, association measures would be biased either toward or away from the null (Chyou 2007). The effect of non-differential misclassification varies by the factor that is misclassified. When outcome

variables suffer of non-differential misclassification, association measures are typically biased toward the null (Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol 1977; 105: 488-95). When non-differential misclassification affects variables used to define the cohort (e.g., by excluding prevalent disease at baseline), association measures could be biased away from the null, with the degree of bias varying by disease incidence and prevalence (Pekkanen et al. 2006). Many authors who discussed the qualitative loss that may occur in disease coding underlined the role of specificity in classification, pointing out that when it’s close to 100% the relative risks are unbiased (Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. 2nd edition. New York: Oxford University Press; 1996).
A high specificity of diagnostic coding data can be expected because if a diagnosis is coded and recorded, this diagnosis was most likely made, particularly in hospital discharge summaries (Romano & Mark 1994). Since chronic diseases usually require multiple contacts with the health system, a single diagnostic code may not suffice to accurately identify cases, it is necessary to use of diagnostic algorithms in identifying patients who experience a given outcome. Furthermore, the use of hospital charts to identify cases limits the possibility of detecting all the outcomes (e.g., those that do not require hospital admission), thus introducing a bias due to the selective inclusion of more severe outcomes (Delgado-Rodriguez & Llorca 2004).

Besides the possible misclassification errors related to the outcome, also misclassification errors due to record linkage procedures could occur.
Record-linkage helps the identification of the same patient in various files containing different types of information. The record-linkage allowed to count people with a given diagnosis in a given period; to identify incident cases of a

disease and to characterize the patients included in the cohort, to study mortality and to register a more detailed reconstruction of the patient's medical history and a more complete identification of comorbidities.

Although RL is an important tool in observational research, it may be associated with various types of error. When linking two databases, there is a proportion of records that matches and a proportion that remains unmatched. It is possible that if data sources do not consistently capture the same cases, several errors can happen. For example, records that correspond to the same person fail to link due to missing or inaccurate data (false negatives), or unrelated records are mistakenly linked (false positives) (Bohensky et al. 2010). A good record linkage procedure, generally defined in terms of mismatches reduction, can depend on a number of factors, including the quality of the information used in the linkage process and how this information is uniquely identified. The probabilistic record linkage, that require enormous computational resources, reduces the number of false negatives, and increasing the matching sensitivity, implying a reduction of matching specificity (i.e., an increased probability of matching unrelated records). In this case we would be able to capture all patients who present that disease, even if some false positives are necessarily included among cases captured by probabilistic RL. If specificity must be privileged, we should be able to ensure that patients classified as experiencing the outcome are true positives, in this case the deterministic RL ensures this objective, since it is unlikely that said procedure may introduce false matching (Corrao 2013).

In this study two outcomes were investigated: mortality for all causes and incidence of CVD. Mortality for all causes was a strong outcome and it was assessed through the Mortality Registry, so that it seems very unlike that a misclassification for this event occurred. The only possible criticism in classifying the life status may have been the expected delay in registering the death of a subject, but this issue was overcome in the analyses by anticipating the end of the follow-up period.

Incident cardiovascular diseases were obtained from the Hospital Discharge forms across a long period of time (1997-2012). Data collected in these databases have considerably grown in quality during this time period so that, considering also the length of follow-up and the multiple contacts of citizens with the health system, it seems reasonable that very few CVD were missed and only in particular situations (like for example an omitted CVD diagnosis in short-term survivors in whom these diseases were not causes of death) as well as that the rate of false positives was low (Romano & Mark 1994).

However, should a disease misclassification have occurred, even due to coding errors or mistyping, it would have been non-differential. This consideration, together with the high specificity level achieved, supports the conclusion that in these analyses biases potentially due to disease misclassification are towards the null.

5.2 Effects of Diet and other Lifestyle Factors on CVD and All-causes Death Risk

CVD are the leading cause of morbidity and mortality in the world, and the clustering of other pathologies (obesity, type 2 diabetes, hyperlipidemia, and hypertension) defined as metabolic syndrome, results in a higher risk of CVD occurrence and mortality (Grosso et al. 2014).

Robust scientific evidence consistently showed that the adherence to MeDi reduces CVD morbidity and mortality in several longitudinal studies (Keys, Menotti et al. 1986, Fung et al. 2009, Kastorini et al. 2011, Nordmann et al. 2011) and RCTs (de Lorgeril et al. 1999, Llorente-Cortes et al. 2010, Estruch et al. 2013).

Although, we observed a weakly and non significant inverse association between the global MedDietScore and CVD risk, when we evaluated the role of the single components of the MeDi, we observed a significant 25% reduction of risk for a high consumption of olive oil.

High olive oil intake is one of the most constant features of the MeDi and is the main source of fat calories, thus a great deal of research has been dedicated to characterize its role in the development of different diseases (Lopez-Miranda et al. 2010). Some observational studies, reviewed by Ros E. (Ros 2012), investigated the role of olive oil exposure and incident CVD. Globally these data suggest that a high intake of olive oil was inversely associated with the CVD risk. Bendinelli et al. (Bendinelli, Masala et al. 2011) reported a strong reduction in CHD risk by 44% among women with the highest quartile of consumption of olive oil. Olive oil contains 80% monounsaturated fats (MUFA) in the form of oleic acid, 20% polyunsaturated fats, and several antioxidant components, including phenolic compounds found in virgin olive oil. Olive oil has several cardioprotective properties. The demonstrated vascular beneficial effects of olive oil include blood pressure reduction, improvement of blood lipid profile, reduction of low-density lipoprotein susceptibility to oxidation, and improvement of oxidative vascular damage and endothelial function (Lopez-Miranda, Perez-Jimenez et al. 2010) possibly through a modulation of key genes implied in vascular inflammation, foam cell formation, and thrombosis. All of these effects were primarily attributed to oleic acid (Mensink et al. 2003) and more recently, also to phenolic compounds found in virgin olive oil (Beauchamp et al. 2005).

The same inverse relationship, when considered the pattern extracted from the PCA, was observed. Indeed, a high adherence of ‘Olive oil’ pattern was associated with a 27% CVD risk reduction. Finally, the study of the relationship between macronutrients intake and CVD incidence showed that people in the medium to high tertile of proteins, had an increased risk of CVD whereas a high intake of energy form fats, reduced the risk by 28%. A possible interpretation could derive from the food sources of macronutrients in this cohort that were red meat for proteins and olive oil for fats.

Among the MeDi components, I found also that a high intake of potatoes was associated with a borderline reduction of CVD risk. The potato is the highest source of potassium of all foods, which could be, among all the components that constitute potatoes, those with the main protective effect against CVD. A meta-analysis of 11 prospective studies of potassium intake, stroke, and cardiovascular disease involving about 250,000 individuals showed the strongest association with reduction of risk of stroke by 21% for every 1.64-g/d (423-mmol/d) increase in potassium intake and a trend toward a lower risk of cardiovascular disease (D'Elia et al. 2011). High dietary potassium is associated with a decrease in blood pressure which is the major risk factor for the development of stroke, coronary heart disease and heart failure (Weaver 2013).

The food group ‘potatoes’ contributed, together with vegetables and legumes, to the definition of the ‘Plant-foods’ factor obtained with the PCA. A high adherence on this dietary pattern was associated with a borderline reduction by 15% of CVD risk as previously reported in literature (Widmer et al. 2014). Besides potassium, the presence of folic acid, antioxidant vitamins might contribute to reduce blood homocysteine concentration (Joshipura et al. 2001) and protects against lipid peroxidation (Knekt et al. 2004).

Considering the median value as the cut-off point, I observed that an alcohol intake over the 50° percentile, was associated with an increase risk of CVD independently from the other components of the MeDi. Indeed, several lines of evidence in literature suggest that, light to moderate alcohol consumption is associated with a reduced risk of multiple cardiovascular outcomes conversely, chronic heavy alcohol consumption and binge drinking are associated with increased risk of cardiovascular events (Ronksley et al. 2011).

Although I did not find a protective effect of dairy products when considered them as single component of the MedDietScore, the factor derived from the PCA, which included high fats dairy products and alcoholic beverages, was protective against CVD with a significant 21% risk reduction.

An interesting interpretation could be the "French Paradox". This term was coined from the epidemiological observation that some French populations suffered a relatively low incidence of coronary heart disease (CHD), despite a relatively high dietary intake of saturated fatty acids, the protective effect was potentially attributable to the consumption of red wine (Lippi et al. 2010).

It has been proposed a positive effect of dairy products ingestion on lipoprotein turnover and plasma lipid profile, hemorheological parameters and inflammatory status. Furthermore, several peptides capable of inhibiting the angiotensin-converting enzyme which controls systemic blood pressure have been isolated (Petyaev & Bashmakov 2012). Other data from the literature seem to suggest that the consumption of milk and dairy products is associated with a reduction of vascular diseases (Elwood et al. 2010).

People in the highest tertile of adherence to the MedDietScore had a 38% reduction of death risk compared with subjects in the lowest tertile and a 5% reduction of death risk for each one-point increase in the MedDietScore.

These results are consistent with other European and US prospective cohort studies that showed inverse associations between adherence to MeDi and total mortality (Trichopoulou et al. 1995, Trichopoulou et al. 2003, Knoops et al. 2004, Trichopoulou et al. 2005, Lagiou et al. 2006, Mitrou, Kipnis et al. 2007, Trichopoulou et al. 2009, Buckland, Agudo et al. 2011, Martinez-Gonzalez et al. 2012, McNaughton et al. 2012, Tognon et al. 2014). Two recent meta-analyses summarized the prospective cohort studies that evaluated the association between the Mediterranean Diet Score (MDS) (Trichopoulou, Kouris-Blazos et al. 1995, Trichopoulou, Costacou et al. 2003) and overall mortality reporting an 8% reduction for a 2-point increase in adherence to the Mediterranean Diet (Sofi et al. 2010, Sofi et al. 2013).

The biological explanations for the protective role of Mediterranean Diet rich in plant-based foods might include both its antioxidant and anti-inflammatory effects (Dai et al. 2008, Kastorini, Milionis et al. 2011).

Evaluating the single components of MeDi I found that subjects with a high intake of fruits had a 30% reduction of risk for all-cause death; however no strong associations were evident for the other score components. I speculate that the high interaction between the components of the score and the fact that single foods may have small effects that may appear only when all the components are included in a pattern (Hu 2002). Still, by analyzing the effect of specific components one might miss associations between diet and disease because the effects of the individual components are examined against the background of average risk associated with other nutrients or foods (Jacques & Tucker 2001).

Participants who were engaged in at least one sport during the adulthood, had a significant reduction of CVD risk and death, respectively by 25% and 46%. Overall, the results of a recent systematic review of longitudinal studies showed a positive long-term effect where people who were physically active in middle aged, had a lower risk of suffering from a coronary heart disease later in their life (Reiner et al. 2013). Another review on prospective studies, reported that the amount of time being sedentary, i.e. watching television, watching videos and using a computer, is associated with increased risk of fatal and non-fatal CVD (Ford & Caspersen 2012).

A recent systematic review and meta-analysis on cohort studies reported that increasing levels of total and domain-specific physical activity reduced by 35% the risk of all-cause death, with stronger associations for exercise and sports, leisure-time activities and activities of daily living, rather than for occupational and transport-related activity (Samitz et al. 2011). In another meta-analysis on longitudinal studies the authors found that 2·5 h/week (equivalent to 30 min daily of moderate intensity activity on 5 days/week) compared with no activity was associated with a reduction in mortality risk by 19% while 7 h/week of moderate activity compared with no activity reduced the mortality risk by 24%(Woodcock et al. 2011). The mechanisms underlying this protective effect could involve the improvement of cardio-metabolic profile, resulting in reduced risks of vascular diseases associated with high mortality risk (Gill & Malkova 2006, Ahmed et al. 2012). In addition, regular physical activity decreases the susceptibility of cells and tissues to oxidative stress, increases vascularization, and enhances energy metabolism (Radak et al. 2010).

Never smokers had respectively a 15% and a 29% lower risk of CVD and all-causes death, than ever smokers. Cigarette smoking is probably the most complex and the least understood among the risk factors for CVDs, it is responsible for a preventable 10% of CVDs (http://www.who.int/mediacentre/factsheets/fs339/en/).

Tobacco smoke is a mixture of several toxic compounds involved in the disease development by different mechanisms. In particular, smoking plays a strong role not only in CVD initiation but also significantly contributes to and causes disease progression and fatal cardiovascular outcomes. The key processes in smoking induced atherogenesis initiation are endothelial dysfunction and damage, increase in and oxidation of proatherogenic lipids, as well as decrease of high-density lipoprotein, induction of inflammation, and the shift toward a procoagulant state in the circulation (Messner & Bernhard 2014). Several observational studies reported that non-smokers have a significant reduction of death risk (Zheng et al. 2014). Smoking remains the leading preventable cause of disease and death and approximately 6 million deaths worldwide every year have been attributed to smoking (WHO. The world health report 2002: Reducing risks, promoting healthy life. Geneva: WHO; 2002.

Considering the combined effects of diet, physical activity and smoking status, the highest number of healthy lifestyle factors was associated to a 35% CVD risk reduction.

These findings were supported by previous studies which evaluated the association between different combinations of lifestyle factors, such as alcohol consumption, smoking habits, diet, BMI, physical exercise, and CVD fatal and non fatal event (Larsson et al. 2014),(Kurth et al. 2006, Eguchi et al. 2012). Although only few studies included the MeDi in the score (Chiuve et al. 2011, Booth et al. 2014), (Ahmed et al. 2013).

Compared with subjects in the lowest healthy lifestyle score group, subjects scoring with 1, 2 and 3 had respectively a significant 39%, 56%, and 73%

reduction of risk of death during follow-up. The authors of a recent meta-analysis (Loef & Walach 2012) reported a reduction in mortality by 66% (95% CI 58%, 73%) for a combination of different healthy lifestyle behaviors as smoking, BMI, physical activity, alcohol habits and diet, although only two studies considered MeDi food pattern, physical activity and smoking habits. In the SENECA Study, 1281 subjects aged 70-75 followed up for 10 years with all three unhealthy behaviors, having a low adherence to MeDi, smoking, and being physically inactive had three-to four folds increased risk in mortality risk (Haveman-Nies et al. 2002). In the Italian Rural Area Seven Country Study the combination of the 3 unhealthy risk factors as smoking, sedentary activity and a diet far from the Mediterranean style, was associated with 4·8-year life loss in the 20-year follow-up and 10·7-year life loss in the 40-year follow-up (Menotti, Puddu et al. 2014).

5.3 Weaknesses and Strengths of the Project

The first limitation concerns the use of administrative data that were not created exclusively for epidemiological purposes. This approach may be insensitive, rather than the standardized clinical assessments administered periodically to all cohort members. It is also likely that we missed some cases for subjects who died or who were hospitalized prior the start of the electronic registration.

Another weakness is the single dietary intake assessment in adult life that might not reflect the long-term dietary intake of participants thus we cannot assert that persons may have followed similar diet in the years before and after the assessment. Furthermore physical activity was self reported and assessed also on a single time point throughout adulthood. We considered subjects to be physically active when engaged in at least one sport and we did not take into account the free-living physical activity due to standing/walking, sitting or moving around; therefore, we might underestimate levels of physical activity of the participants.

Another aspect concerns the dietary patterns approach. The major limitation of the *a priori* dietary patterns is that they vary across studies and include different dietary components, different weighting components and cut off point resulting indexes that potentially measure different definitions of healthful behavior. This approach is limited by current knowledge and an understanding of the diet-disease relationship, and can be fraught with uncertainties in selecting individual components of the score and subjectivity in defining cut-off points. Related to the exploratory approach, it generates patterns based on available empirical data without *a priori* hypothesis, that do not necessarily represent optimal patterns. The PCA involves several arbitrary but important decisions, including the consolidation of food items into food groups, the number of factors to extract, the method of rotation, and even the labelling of the components. In addition, one should evaluate whether the patterns generated

into the commonly recognized eating habits in the population, because these patterns are generated simply on the basis of eating behaviors.
Finally, we cannot also rule out the possibility of residual confounding due to other factors that were not considered in this study, even if we adjusted for the main known risk factors.

The major strength of this study is the longitudinal design. Because the time sequence in which the nutritional exposure precedes disease development can be established, prospective studies are considered to be the best of the observational study design. Indeed, prospective cohort avoid most of the potential sources of methodological bias associated with case-control studies, because the dietary and other life-style related factors are collected before the diagnosis of disease, illness cannot affect the recall of diet, distributions of dietary factors in the study population may be affected by selective participation in the cohort; low participation rates at enrollment, however, will not distort the relationships between dietary factors and disease and lastly, they provide the opportunity to examine the effects of diet on a wide variety of disease, including total mortality, simultaneously.
This study is also characterized by a long period of follow-up, a large sample size and by the small number of participants lost at follow-up, limiting the possibility of selection bias.
Another important point of this work is the use of administrative data. The primary advantages of passive case-finding compared to active case-finding are the lower cost of research and the lower burden on the population studied. Furthermore, the integration of databases, within this context, allows monitoring cohort studies, creation of health history, enabling improvement of information quality and consistency, preparation of records for studying

diseases, and following cohorts to determine life status of the individual stories and genealogical study or historical.

In this work we investigated a range of potential confounding factors as age, sex, educational level, body mass index and energy intake, allowing us to control for their potential confounding effect in the analysis.

A further potential strength is the dietary assessment tool. Although standardized methods exist for collecting food records and 24-hour dietary recalls, FFQ tends to be more variable with respect to purpose (measuring total diet or selected foods/ nutrients), number of foods asked, whether or not (or how) portion size is queried, wording of questions, and food/nutrient database. The advantage of the FFQ is its ease of administration, assessment of intake over an extended period of time, and low cost.

Finally, the main strong point has been the use of dietary patterns that overcame the methodological issues related to the use of single nutrient or food. The first consideration is that people do not eat isolated nutrients, they consume meals consisting of a variety of foods with complex combinations of micro- and macronutrients. Furthermore, there is a good chance of a high level of collinearity of the food variables. For example, fruit intake may be highly and positively correlated with intake of vegetables because of factors such as culture or beliefs. In addition, multicollinearity is a similar phenomenon that occurs where more than two independent variables are highly correlated. In this situation, the effect sizes may change erratically in response to small changes in the model or the data. Additionally, multicollinearity does not actually bias the results, but produces large standard errors in the related independent variables. In other words, large standard errors increase the uncertainty of the findings. In this work, adherence to the Mediterranean diet was evaluated by means of a score in agreement with the principles of the Mediterranean pattern and not specific of the dietary consumption of the population studied. Advantages of the

MedDietScore are the weighting of the selected food groups, depending on the frequency of consumption (thresholds are chosen based on *a priori* hypothesis) and regardless of the consumptions of the sample studied.

5.4 References

Ahmed HM, Blaha MJ, Nasir K, et al. (2013). Low-risk lifestyle, coronary calcium, cardiovascular events, and mortality: results from MESA. Am J Epidemiol **178**(1): 12-21.

Ahmed HM, Blaha MJ, Nasir K, et al. (2012). Effects of physical activity on cardiovascular disease. Am J Cardiol **109**(2): 288-295.

Beauchamp GK, Keast RS, Morel D, et al. (2005). Phytochemistry: ibuprofen-like activity in extra-virgin olive oil. Nature **437**(7055): 45-46.

Bendinelli B, Masala G, Saieva C, et al. (2011). Fruit, vegetables, and olive oil and risk of coronary heart disease in Italian women: the EPICOR Study. Am J Clin Nutr **93**(2): 275-283.

Bohensky MA, Jolley D, Sundararajan V, et al. (2010). Data linkage: a powerful research tool with potential problems. BMC Health Serv Res **10**: 346.

Booth JN, 3rd, Levitan EB, Brown TM, et al. (2014). Effect of sustaining lifestyle modifications (nonsmoking, weight reduction, physical activity, and mediterranean diet) after healing of myocardial infarction, percutaneous intervention, or coronary bypass (from the REasons for Geographic and Racial Differences in Stroke Study). Am J Cardiol **113**(12): 1933-1940.

Brunner EJ, Mosdol A, Witte DR, et al. (2008). Dietary patterns and 15-y risks of major coronary events, diabetes, and mortality. Am J Clin Nutr **87**(5): 1414-1421.

Buckland G, Agudo A, Travier N, et al. (2011). Adherence to the Mediterranean diet reduces mortality in the Spanish cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC-Spain). Br J Nutr **106**(10): 1581-1591.

Chiuve SE, Fung TT, Rexrode KM, et al. (2011). Adherence to a low-risk, healthy lifestyle and risk of sudden cardiac death among women. JAMA **306**(1): 62-69.

Chyou PH (2007). Patterns of bias due to differential misclassification by case-control status in a case-control study. Eur J Epidemiol **22**(1): 7-17.

Corrao G (2013). Building reliable evidence from real-world data: methods, cautiousness and recommendations. Epidemiology Biostatistics and Public Health **10**(3): e8981 8981-8940.

D'Elia L, Barba G, Cappuccio FP, et al. (2011). Potassium intake, stroke, and cardiovascular disease a meta-analysis of prospective studies. J Am Coll Cardiol **57**(10): 1210-1219.

Dai J, Jones DP, Goldberg J, et al. (2008). Association between adherence to the Mediterranean diet and oxidative stress. Am J Clin Nutr **88**(5): 1364-1370.

de Lorgeril M, Salen P, Martin JL, et al. (1999). Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. Circulation **99**(6): 779-785.

Delgado-Rodriguez M & Llorca J (2004). Bias. J Epidemiol Community Health **58**(8): 635-641.

Eguchi E, Iso H, Tanabe N, et al. (2012). Healthy lifestyle behaviours and cardiovascular mortality among Japanese men and women: the Japan collaborative cohort study. Eur Heart J **33**(4): 467-477.

Elwood PC, Pickering JE, Givens DI, et al. (2010). The consumption of milk and dairy foods and the incidence of vascular disease and diabetes: an overview of the evidence. Lipids **45**(10): 925-939.

Estruch R, Ros E, Salas-Salvado J, et al. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med **368**(14): 1279-1290.

Ford ES & Caspersen CJ (2012). Sedentary behaviour and cardiovascular disease: a review of prospective studies. Int J Epidemiol **41**(5): 1338-1353.

Fung TT, Rexrode KM, Mantzoros CS, et al. (2009). Mediterranean diet and incidence of and mortality from coronary heart disease and stroke in women. Circulation **119**(8): 1093-1100.

Gill JM & Malkova D (2006). Physical activity, fitness and cardiovascular disease risk in adults: interactions with insulin resistance and obesity. Clin Sci (Lond) **110**(4): 409-425.

Greenland S (1996). Basic methods for sensitivity analysis of biases. Int J Epidemiol **25**(6): 1107-1116.

Grosso G, Mistretta A, Frigiola A, et al. (2014). Mediterranean diet and cardiovascular risk factors: a systematic review. Crit Rev Food Sci Nutr **54**(5): 593-610.

Harriss LR, English DR, Powles J, et al. (2007). Dietary patterns and cardiovascular mortality in the Melbourne Collaborative Cohort Study. Am J Clin Nutr **86**(1): 221-229.

Haveman-Nies A, de Groot L, Burema J, et al. (2002). Dietary quality and lifestyle factors in relation to 10-year mortality in older Europeans: the SENECA study. Am J Epidemiol **156**(10): 962-968.

Hu FB (2002). Dietary pattern analysis: a new direction in nutritional epidemiology. Curr Opin Lipidol **13**(1): 3-9.

Huijbregts P, Feskens E, Rasanen L, et al. (1997). Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. BMJ **315**(7099): 13-17.

Jacques PF & Tucker KL (2001). Are dietary patterns useful for understanding the role of diet in chronic disease? Am J Clin Nutr **73**(1): 1-2.

Joshipura KJ, Hu FB, Manson JE, et al. (2001). The effect of fruit and vegetable intake on risk for coronary heart disease. Ann Intern Med **134**(12): 1106-1114.

Kastorini CM, Milionis HJ, Esposito K, et al. (2011). The effect of Mediterranean diet on metabolic syndrome and its components: a meta-analysis of 50 studies and 534,906 individuals. J Am Coll Cardiol **57**(11): 1299-1313.

Keys A, Menotti A, Karvonen MJ, et al. (1986). The diet and 15-year death rate in the seven countries study. Am J Epidemiol **124**(6): 903-915.

Knekt P, Ritz J, Pereira MA, et al. (2004). Antioxidant vitamins and coronary heart disease risk: a pooled analysis of 9 cohorts. Am J Clin Nutr **80**(6): 1508-1520.

Knoops KT, de Groot LC, Kromhout D, et al. (2004). Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA **292**(12): 1433-1439.

Kurth T, Moore SC, Gaziano JM, et al. (2006). Healthy lifestyle and the risk of stroke in women. Arch Intern Med **166**(13): 1403-1409.

Lagiou P, Trichopoulos D, Sandin S, et al. (2006). Mediterranean dietary pattern and mortality among young women: a cohort study in Sweden. Br J Nutr **96**(2): 384-392.

Larsson SC, Akesson A & Wolk A (2014). Healthy diet and lifestyle and risk of stroke in a prospective cohort of women. Neurology **83**(19): 1699-1704.

Lippi G, Franchini M, Favaloro EJ, et al. (2010). Moderate red wine consumption and cardiovascular disease risk: beyond the "French paradox". Semin Thromb Hemost **36**(1): 59-70.

Llorente-Cortes V, Estruch R, Mena MP, et al. (2010). Effect of Mediterranean diet on the expression of pro-atherogenic genes in a population at high cardiovascular risk. Atherosclerosis **208**(2): 442-450.

Loef M & Walach H (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med **55**(3): 163-170.

Lopez-Miranda J, Perez-Jimenez F, Ros E, et al. (2010). Olive oil and health: summary of the II international conference on olive oil and health consensus report, Jaen and Cordoba (Spain) 2008. Nutr Metab Cardiovasc Dis **20**(4): 284-294.

Marmot M & Brunner E (2005). Cohort Profile: the Whitehall II study. Int J Epidemiol **34**(2): 251-256.

Martinez-Gonzalez MA, Guillen-Grima F, De Irala J, et al. (2012). The Mediterranean diet is associated with a reduction in premature mortality among middle-aged adults. J Nutr **142**(9): 1672-1678.

Masala G, Ceroti M, Pala V, et al. (2007). A dietary pattern rich in olive oil and raw vegetables is associated with lower mortality in Italian elderly subjects. Br J Nutr **98**(2): 406-415.

McNaughton SA, Bates CJ & Mishra GD (2012). Diet quality is associated with all-cause mortality in adults aged 65 years and older. J Nutr **142**(2): 320-325.

Menotti A, Puddu PE, Lanti M, et al. (2014). Lifestyle habits and mortality from all and specific causes of death: 40-year follow-up in the Italian Rural Areas of the Seven Countries Study. J Nutr Health Aging **18**(3): 314-321.

Mensink RP, Zock PL, Kester AD, et al. (2003). Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. Am J Clin Nutr **77**(5): 1146-1155.

Messner B & Bernhard D (2014). Smoking and cardiovascular disease: mechanisms of endothelial dysfunction and early atherogenesis. Arterioscler Thromb Vasc Biol **34**(3): 509-515.

Mitrou PN, Kipnis V, Thiebaut AC, et al. (2007). Mediterranean dietary pattern and prediction of all-cause mortality in a US population: results from the NIH-AARP Diet and Health Study. Arch Intern Med **167**(22): 2461-2468.

Mursu J, Virtanen JK, Tuomainen TP, et al. (2014). Intake of fruit, berries, and vegetables and risk of type 2 diabetes in Finnish men: the Kuopio Ischaemic Heart Disease Risk Factor Study. Am J Clin Nutr **99**(2): 328-333.

Nordmann AJ, Suter-Zimmermann K, Bucher HC, et al. (2011). Meta-analysis comparing Mediterranean to low-fat diets for modification of cardiovascular risk factors. Am J Med **124**(9): 841-851 e842.

Osler M, Heitmann BL, Gerdes LU, et al. (2001). Dietary patterns and mortality in Danish men and women: a prospective observational study. Br J Nutr **85**(2): 219-225.

Pekkanen J, Sunyer J & Chinn S (2006). Nondifferential disease misclassification may bias incidence risk ratios away from the null. J Clin Epidemiol **59**(3): 281-289.

Petyaev IM & Bashmakov YK (2012). Could cheese be the missing piece in the French paradox puzzle? Med Hypotheses **79**(6): 746-749.

Radak Z, Hart N, Sarga L, et al. (2010). Exercise plays a preventive role against Alzheimer's disease. J Alzheimers Dis **20**(3): 777-783.

Reiner M, Niermann C, Jekauc D, et al. (2013). Long-term health benefits of physical activity--a systematic review of longitudinal studies. BMC Public Health **13**: 813.

Romano PS & Mark DH (1994). Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care **32**(1): 81-90.

Ronksley PE, Brien SE, Turner BJ, et al. (2011). Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. BMJ **342**: d671.

Ros E (2012). Olive oil and CVD: accruing evidence of a protective effect. Br J Nutr **108**(11): 1931-1933.

Samitz G, Egger M & Zwahlen M (2011). Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies. Int J Epidemiol **40**(5): 1382-1400.

Sofi F, Abbate R, Gensini GF, et al. (2010). Accruing evidence on benefits of adherence to the Mediterranean diet on health: an updated systematic review and meta-analysis. Am J Clin Nutr **92**(5): 1189-1196.

Sofi F, Macchi C, Abbate R, et al. (2013). Mediterranean diet and health status: an updated meta-analysis and a proposal for a literature-based adherence score. Public Health Nutr: 1-14.

Solfrizzi V, D'Introno A, Colacicco AM, et al. (2005). Unsaturated fatty acids intake and all-causes mortality: a 8.5-year follow-up of the Italian Longitudinal Study on Aging. Exp Gerontol **40**(4): 335-343.

Tognon G, Lissner L, Saebye D, et al. (2014). The Mediterranean diet in relation to mortality and CVD: a Danish cohort study. Br J Nutr **111**(1): 151-159.

Trichopoulou A, Bamia C & Trichopoulos D (2009). Anatomy of health effects of Mediterranean diet: Greek EPIC prospective cohort study. BMJ **338**: b2337.

Trichopoulou A, Costacou T, Bamia C, et al. (2003). Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med **348**(26): 2599-2608.

Trichopoulou A, Kouris-Blazos A, Wahlqvist ML, et al. (1995). Diet and overall survival in elderly people. BMJ **311**(7018): 1457-1460.

Trichopoulou A, Orfanos P, Norat T, et al. (2005). Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. BMJ **330**(7498): 991.

van Walraven C, Bennett C & Forster AJ (2011). Administrative database research infrequently used validated diagnostic or procedural codes. J Clin Epidemiol **64**(10): 1054-1059.

Weaver CM (2013). Potassium and health. Adv Nutr **4**(3): 368S-377S.

Widmer RJ, Flammer AJ, Lerman LO, et al. (2014). "The Mediterranean Diet, its Components, and Cardiovascular Disease". Am J Med.

Woodcock J, Franco OH, Orsini N, et al. (2011). Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. Int J Epidemiol **40**(1): 121-138.

Zheng W, McLerran DF, Rolland BA, et al. (2014). Burden of total and cause-specific mortality related to tobacco smoking among adults aged >/= 45 years in Asia: a pooled analysis of 21 cohorts. PLoS Med **11**(4): e1001631.

6.CONCLUSIONS

In summary, this thesis project shows a possible application of the use of administrative data, and provides a detailed description of strengths and limitations of administrative databases, rarely used for epidemiological analytical study in Italy. Furthermore this work stresses their promising role for carrying out observational cohort studies with a high amount of data collected at baseline, guaranteeing a reduction in times and costs compared to traditional studies, ensuring a minimal number of persons lost to follow-up and resulting in a high number of subjects to analyze.

In this project, I applied the use of administrative data for exploring the role of diet and other lifestyle factors collected at baseline on the incidence of cardiovascular diseases and mortality during a long period of follow-up.

My results support the evidence of a positive effect of a healthy lifestyle, i.e. adherence to a Mediterranean dietary pattern and the intake of specific components of this pattern, abstinence from smoking and engaging in regular physical activity, on the risk of CVD and all-causes death, especially when these modifiable risk factors were considered in a healthy lifestyle score. These results seem to be clinically relevant in terms of public health, particularly for policy makers aiming to reduce the risk CVD and of all-causes death in the healthy general population. Primary prevention strategies with respect to the prevention of chronic disease, as CVD, in old age or early mortality, encouraging a modification in lifestyle during adult life (e.g. changing dietary habits, increasing the amount of daily physical activity, quitting smoking) will be necessary in order to improve individual and community health.

These strategies could be better defined starting from scientific evidence deriving from studies in which lifestyles parameters should be accurately collected during mid-life in sufficiently large cohorts of people and mortality or incidence of long latency diseases (not only CVD but also neurodegenerative or oncological ones) can be evaluated if a long follow-up period, active or passive, is achieved.

Thus, the use of administrative data in this direction could be a suitable tool to meet these needs.

For the future perspective, it will be necessary to proceed towards standardization and integration of information systems, and to develop methodologies that guarantee the quality of administrative data to support epidemiological studies and to make important contributions to the public health area.

7.TABLES: DESCRIPTIVES ANALYSIS

Table1. Baseline characteristics by sex in the Italian population (N=1353)

	Sex	
	Males (N=672)	**Females (N=681)**
Age at baseline (mean, range)	57,5 (42-74)	56,2 (42-74)
Body Mass Index (mean±SD)	26,9±3,4	27,1±4,7
Educational level		
Primary school or less	331 (49,3)	445 (65,3)
Medium school or graduate	341 (50,7)	236 (34,7)
Prevalent diseases		
Cardiovascular diseases	119 (17,7)	104 (15,3)
Cancer	19 (2,8)	33 (4,8)
Diabetes	53 (7,9)	36 (5,3)
Hypertension	415 (61,8)	387 (56,8)
Smoking status		
Never	197 (29,3)	501 (73,6)
Ever	475 (70,7)	180 (26,4)
Alcohol intake (g/day)		
Abstainer	89 (13,4)	318 (47)
≤30g/day	175 (26,4)	119 (17,6)
>30g/day	400 (60,2)	240 (35,5)
Physical activity		
Inactive	527 (78,4)	561 (82,4)
Active	145 (21,6)	120 (17,6)
MedDietScore		
Low	202 (30,1)	334 (49)
Medium	231 (34,4)	191 (28)
High	239 (35,6)	156 (22,9)
% of Energy from		
Carbohydrates	49,2±9,7	52,8±8,2
Proteins	14,9±3,2	16,0±3,5
Fats	26,3±7,2	27,9±6,7
Lifestyle Score		
0	118 (17,6)	77 (11,3)
1	326 (48,5)	289 (42,4)
2	198 (29,5)	266 (39,1)
3	30 (4,5)	49 (7,2)

Data were reported as number and column percentages

Table 2. Baseline characteristics and number of deceased by incident cardiovascular diseases in the Italian population (N=1073)

	Cardiovascular Disease	
	Yes (N=530)	**No (N=543)**
Age at baseline (mean, range)	58,5 (42-74)	53,8 (42-74)
Body Mass Index (mean±SD)	27,3±4,2	26,5±4,0
Male sex	285 (53,8)	243 (44,8)
Educational level		
Primary school or less	304 (57,4)	287 (52,9)
Medium school or graduate	226 (42,6)	256 (47,1)
Smoking status		
Never	267 (50,4)	289 (53,2)
Ever	263 (49,6)	254 (46,8)
Alcohol intake (g/day)		
Abstainer	156 (29,7)	171 (31,8)
≤30g/day	105 (20)	125 (23,3)
>30g/day	264 (50,3)	241 (44,9)
Physical activity		
Inactive	436 (82,3)	416 (76,6)
Active	94 (17,7)	127 (23,4)
Hypertension	361 (68,1)	249 (45,9)
MedDietScore		
Low	202 (38,1)	219 (40,3)
Medium	175 (33)	160 (29,5)
High	153 (28,9)	164 (30,2)
% of Energy from		
Carbohydrates	51,1±9,2	50,9 ±9,1
Proteins	15,4±3,3	15,5±3,5
Fats	26,5±6,9	27,8±7,0
Lifestyle score		
0	77 (14,5)	76 (14)
1	244 (46)	233 (42,9)
2	182 (34,3)	195 (35,9)
3	27 (5,1)	39 (7,2)
Deceased	173 (32,6)	66 (12,2)

Data were reported as number and column percentages
Subjects with antecedent CVD and subjects without complete data at baseline were excluded from the analysis (N=280)

Table 3. Baseline characteristics by causes of death in the Italian population (N=974)

	CVD (N=44)	Cancer (N=64)	Other causes (N=21)	N. of deaths (N=193)	Total subjects* (N=974)
Male sex	30 (6,2)	45 (9,3)	16 (3,3)	131 (27,1)	483 (49,6)
Age at baseline (mean, range)	64,7 (48-73)	61,0 (43-73)	65,1 (48-73)	62,4 (43-74)	55,7 (42-74)
Body Mass Index (mean±SD)	26,7±4,3	28,2±4,9	26,4±4,3	27,4±4,6	26.8±4,1
Educational level					
Elementary or less	26 (5)	37 (7,1)	14 (2,7)	114 (21,9)	521 (53,5)
Medium school or graduate	18 (4)	27 (6)	7 (1,5)	79 (17,4)	453 (46,5)
Physical activity					
Inactive	36 (4,7)	57 (7,5)	17 (2,2)	165 (21,6)	765 (78,5)
Active	8 (3,8)	7 (3,3)	4 (1,9)	28 (13,4)	209 (21,5)
Smoking status					
Never smoked	19 (3,8)	27 (5,5)	6 (1,2)	84 (17)	495 (50,8)
Ever	25 (5,2)	37 (7,7)	15 (3,1)	109 (22,8)	479 (49,2)
Alcohol intake (g/day)					
Abstainers	9 (3,1)	17 (5,9)	2 (0,7)	52 (17,9)	290 (30,1)
<=30g/d	6 (2,9)	10 (4,9)	1 (0,5)	25 (12,3)	204 (21,2)
>30g/d	29 (6,2)	35 (7,4)	18 (3,8)	113 (24)	470 (48,8)
MedDietScore					
Low	16 (4,2)	23 (6,1)	8 (2,1)	80 (21,2)	378 (38,8)
Medium	15 (4,9)	20 (6,5)	8 (2,6)	61 (19,9)	306 (31,4)
High	13 (4,5)	21 (7,2)	5 (1,7)	52 (17,9)	290 (29,8)

% of Energy from					
Carbohydrates	51,6±8,2	48,4±10,4	48,1±9,3	50,1±9,5	51,1±9,2
Proteins	15,0±3,2	15,2±340	15,5±2,7	15,3±3,4	15,3±3,4
Fats	25, 4±7,7	27,3±8,7	26,2±7,1	26,6±7,5	27,2±7,2
Lifestyle Score					
0	8 (5,7)	12 (8,6)	3 (2,1)	39 (27,9)	140 (14,4)
1	18 (4,2)	31 (7,2)	14 (3,2)	90 (20,9)	431 (44,3)
2	17 (5)	19 (5,6)	3 (0,9)	57 (16,8)	340 (34,9)
3	1 (1,6)	2 (3,2)	1 (1,6)	7 (11,1)	63 (6,5)

Data reported as number and row percentages
*Column percentages
Subjects with cardiovascular diseases, cancer and diabetes at baseline were excluded from the analysis (N=379)
°Causes of death were available for 129/193 subjects. Data of 64 subjects with unknown causes of death, have not been reported

7.TABLES: DIETARY DATA

Table 4. Food grouping used in the *a priori* and *a posteriori* analyses

Food Groups	Food items
Fruits	Apricot, orange, banana, cherry, watermelon, fig, lemon, strawberry, clementines and tangerines, pear, apple, peach, grapefruit, kiwi, plum, grape, fruit juice
Vegetables	Green salad, tomatoes, fennel raw and cooked, celery, asparagus, cucumber, carrots raw and cooked, zucchini, pepper, swiss chard, artichoke, cauliflower and broccoli, onion, mushrooms, aubergine, spinach, cabbage
Potatoes	Potatoes and pureè, french fries, soft pasta with potatoes
Red Meat and processed meat	Beef canned, beef lean and fatty cuts, horse, cutlet, pork chop, stew with potatoes or peas, stew beef, meet with tuna sauce, lamb, entrails, game bird, cured ham, sausages, coppa, bacon, ham, bresaola, speck, salami
Legumes	Green beans, beans, peas
Fish	Dover sole, mackerel, anchovies, trout, hake, eel, crustaceans, tuna canned in oil
Poultry	Turkey, rabbitt, chicken leg and breasts,
Dairy	Whole milk, whole yogurt, processed cheese, crescenza, parmesan, pecorino, ricotta, emmenthal, mozzarella, provola sweet and hot, mascarpone
Cereals	Rice, various kind of pasta (maccaroni, orecchiette), egg pasta, semolina, pasta prepared with broth, bean and pasta soup, polenta
Ethanol	Spirits, dessert wine, cherry, beer, wine (red, white and rose)
Olive oil	Olive oil

Table 5. Factor-loading matrix and explained variance for the four major dietary patterns identified by PCA

Food Groups (servings/week)	Plant-foods (Fact.1)	Animal foods (Fact.2)	Dairy products and alcohol (Fatc.3)	Olive oil (Fact.4)
Fruits	,468			,307
Vegetables	**,827**	,200		
Potatoes	**,532**	,234	,375	-,245
Meat		**,682**		
Legumes	**,821**			
Fish		**,597**		
Poultry		**,742**	-,321	
Dairy			**,615**	
Cereals		,426	,399	
Alcoholic beverages (g/die)			**,584**	
Olive oil				**,898**
Prop. expl. variance (%)	17, 858	15,357	11,950	9,354
Cumul. expl. variance (%)	**17,858**	**33,216**	**44,166**	**53,520**

Estimated from principal component analysis performed on 11 food groups. Loading greater than 0.50 (in absolute value) are shown in bold style. Factors are sorted by % of explained variance. Loadings lesser than 0.20 (in absolute value) are excluded.

7.TABLES: STATISTICAL ANALYSIS

CARDIOVASCULAR DISEASES

Table 6. Association between energy derived from macronutrients intake and cardiovascular diseases

% of Energy from	No. of subjects	Cases	Person-year	Adjusted° HR [95% CI]	Mutually adjusted* HR
Carbohydrates					
Low	344	177	4384.21	Reference	Reference
Medium	373	167	5023.42	0.91 [0.73-1.13]	0.83 [0.66-1.05]
High	356	186	4763.82	0.98 [0.79-1.22]	0.83 [0.64-1.09]
Proteins					
Low	371	166	5117.16	Reference	Reference
Medium	362	199	4574.38	1.45 [1.18-1.79]	1.48 [1.20-1.83]
High	340	165	4479.91	1.21 [0.97-1.52]	1.25 [1.00-1.57]
Fats					
Low	359	194	4588.34	Reference	Reference
Medium	356	182	4670.23	0.98 [0.79-1.20]	0.91 [0.73-1.14]
High	358	154	4912.88	0.82 [0.66-1.02]	0.72 [0.55-0.94]

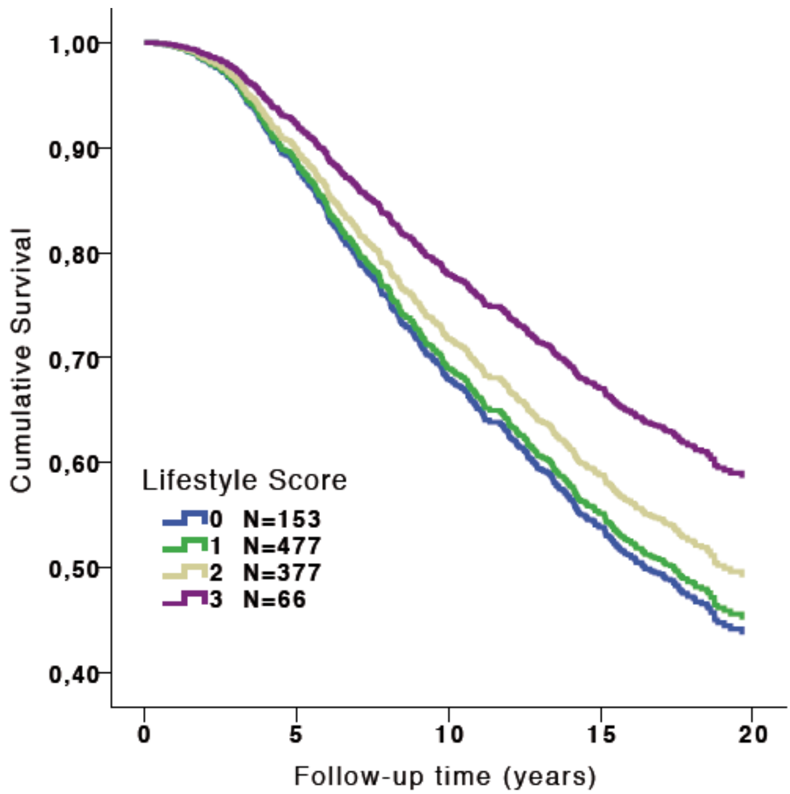
°Adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity, alcohol intake, hypertension (yes/no). *Including all macronutrients in the models

Table 7. Association between smoking status, physical activity, adherence (increasing tertiles) to MedDietScore and to healthy lifestyle score and incidence of Cardiovascular diseases in the Italian cohort

	No. of subjects	Cases	Person-year	Age- and sex adjusted HR [95% CI]	Adjusted§ HR [95% CI]
MedDietScore					
Low	421	202	5533.65	Reference	Reference
Medium	335	175	4242.73	1.10 [0.89-1.35]	1.14 [0.93-1.40]
High	317	153	4395.07	0.88 [0.85-1.05]	0.92 [0.74-1.14]
Physical					
Inactive	852	436	11054.19	Reference	Reference
Active	221	94	3117.26	0.75 [0.60-0.93]	0.76 [0.61-0.95]
Smoking					
Ever	517	263	6675.16	Reference	Reference
Never	556	267	7496.29	0.85 [0.71-1.03]	0.85 [0.70-1.02]
Lifestyle					
0	153	77	1959.54	Reference	Reference
1	477	244	6121.04	0.98 [0.75-1.26]	0.96 [0.74-1.25]
2	377	182	5142.31	0.83 [0.63-1.08]	0.86 [0.65-1.13]
3	66	27	948.56	0.65 [0.42-1.02]	0.65 [0.41-0.99]

§Model additionally adjusted for educational level, BMI, energy intake, and hypertension

Figure 4. Adjusted° cumulative survival curve of cardiovascular diseases by healthy lifestyle score (range 0-3) including the factors: MeDi, physical activity, and smoking habits



°Adjusted for age, sex, educational level, BMI, energy intake and hypertension

Table 8. Cardiovascular diseases adjusted Hazard Ratios (95%CI) associated with intake of components of MedDietScore

MedDietScore components	HR [95%CI] [1]	HR [95%CI] [2]
Vegetables (serving/day)		
≤Median	Reference	Reference
>Median	1.03 [0.86-1.23]	1.04 [0.84-1.23]
Legumes (serving/day)		
≤Median	Reference	Reference
>Median	1.00 [0.84-1.19]	1.00 [0.84-1.21]
Fruits (serving/day)		
≤Median	Reference	Reference
>Median	1.01 [0.84-1.21]	1.05 [0.87-1.27]
Cereals (serving/day)		
≤Median	Reference	Reference
>Median	1.16 [0.96-1.40]	1.15 [0.95-1.40]
Potatoes (serving/day)		
≤Median	Reference	Reference
>Median	0.85 [0.71-1.02]	0.85 [0.70-1.02]
Fish and sea foods (serving/day)		
≤Median	Reference	Reference
>Median	1.13 [0.95-1.35]	1.11 [0.92-1.34]
Dairy Products (serving/day)		
≤Median	Reference	Reference
>Median	0.95 [0.79-1.15]	0.97 [0.80-1.17]
Red meat and products		
≤Median	Reference	Reference
>Median	0.99 [0.82-1.19]	0.97 [0.81-1.17]
Poultry (serving/day)		
≤Median	Reference	Reference
>Median	1.11 [0.93-1.32]	1.10 [0.92-1.31]
Olive oil (serving/day)		
≤Median	Reference	Reference
>Median	0.77 [0.65-0.92]	0.75 [0.63-0.89]
Ethanol intake (serving/day)		
≤Median	Reference	Reference
>Median	1.24 [1.02-1.52]	1.24 [1.02-1.52]

[1] Model adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity, hypertension (yes/no); [2] Mutually adjusted for all components of the MedDietScore

Table 9. Cardiovascular diseases adjusted hazard ratios (95%CI) associated with intake of *a posteriori* dietary pattern

Dietary pattern Components	**HR [95%CI]** [1]	**HR [95%CI]** [2]
Plant-foods		
≤Median	Reference	Reference
>Median	0.86 [0.72-1.03]	0.85 [0.71-1.02]
Animal foods		
≤Median	Reference	Reference
>Median	1.06 [0.88-1.28]	1.01 [0.84-1.23]
Dairy products and alcohol		
≤Median	Reference	Reference
>Median	0.83 [0.68-1.01]	0.79 [0.65-0.97]
Olive oil		
≤Median	Reference	Reference
>Median	0.74 [0.62-0.88]	0.73 [0.61-0.87]

[1] Model adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity, hypertension
[2] Mutually adjusted for all dietary patterns

ALL-CAUSES DEATHS

Table 10. Association between energy derived from macronutrients intake and all-causes death

% of Energy from	No. of subjects	Cases	Person-year	Adjusted° HR [95% CI]	Mutually adjusted* HR
Carbohydrates					
Low	309	66	5244.70	Reference	Reference
Medium	340	66	5943.61	0.92 [0.64-1.31]	0.94 [0.65-1.36]
High	325	61	5737.60	0.77 [0.53-1.12]	0.84 [0.54-1.33]
Proteins					
Low	345	67	5975.39	Reference	Reference
Medium	327	60	5715.06	1.13 [0.80-1.62]	1.12 [0.79-1.61]
High	302	66	5235.46	1.32 [0.92-1.88]	1.26 [0.87-1.83]
Fats					
Low	329	70	5696.94	Reference	Reference
Medium	317	65	5515.95	1.07 [0.75-1.52]	0.97 [0.67-1.42]
High	328	58	5713.02	1.23 [0.86-1.77]	1.08 [0.69-1.66]

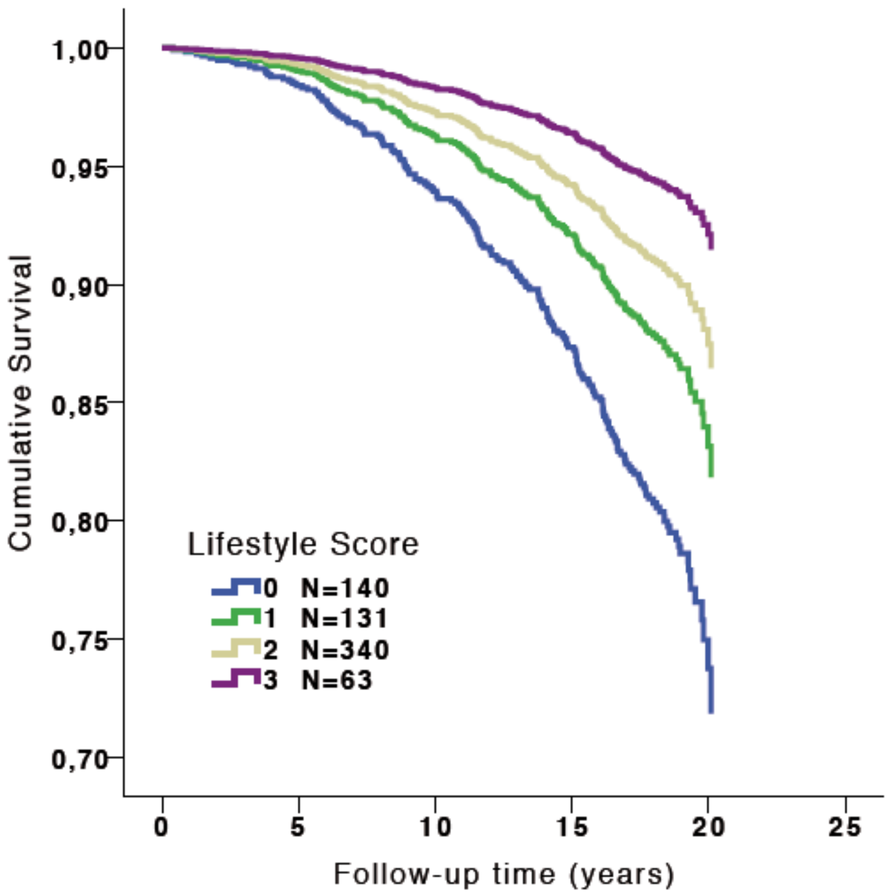
°Adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity, alcohol intake, hypertension (yes/no). *Including all macronutrients in the models

Table 11. Association between smoking status, physical activity and adherence (increasing tertiles) to MedDietScore and to healthy lifestyle score and 20-year all-causes death in the Italian cohort

	No. of subjects	Cases	Person-year	Age- and sex adjusted HR	Adjusted§ HR 95%CI
MedDietScore					
Low	378	80	6596.65	Reference	Reference
Medium	306	61	5220.10	0.80 [0.57, 1.12]	0.79 [0.56, 1.12]
High	290	52	5109.17	0.64 [0.45, 0.91]	0.62 [0.43, 0.89]
Physical					
Inactive	765	165	13155.86	Reference	Reference
Active	209	28	3770.05	0.56 [0.38, 0.84]	0.54 [0.36, 0.82]
Smoking					
Ever	479	109	8125.07	Reference	
Never	495	84	8800.84	0.71 [0.51, 0.97]	0.71 [0.51, 0.98]
Lifestyle					
0	140	39	2294.73	Reference	Reference
1	431	90	7488.82	0.64 [0.43, 0.93]	0.61 [0.41, 0.89]
2	340	57	6015.75	0.46 [0.31, 0.70]	0.44 [0.29, 0.67]
3	63	7	1126.61	0.30 [0.13, 0.67]	0.27 [0.12, 0.61]

§Model additionally adjusted for educational level, BMI, energy intake, and hypertension

Figure 5. Adjusted° Cumulative survival curve of all-causes death by healthy lifestyle score (range 0-3) including the factors: MeDi, physical activity, and smoking habits



° Adjusted for sex, age, educational level, BMI, and energy intake

Table 12. All-causes death adjusted Hazard Ratios (95%CI) associated with intake of components of MedDietScore

MedDietScore components	HR 95% CI [1]	HR 95% CI [2]
Vegetables (serving/day)		
≤Median	Reference	Reference
>Median	0.92 [0.68, 1.24]	0.93 [0.68, 1.27]
Legumes (serving/day)		
≤Median	Reference	Reference
>Median	1.07[0.80, 1.44]	1·08 [0.79, 1.48]
Fruits (serving/day)		
≤Median	Reference	Reference
>Median	0.71 [0.53, 0.96]	0.70[0.51, 0.95]
Cereals (serving/day)		
≤Median	Reference	Reference
>Median	0.93 [0.68, 1.27]	0.91 [0.66, 1.26]
Potatoes (serving/day)		
≤Median	Reference	Reference
>Median	0.99 [0.73, 1.35]	0.99 [0.73, 1.36]
Fish and sea foods (serving/day)		
≤Median	Reference	Reference
>Median	0.99 [0.74, 1.33]	0.98 [0.72, 1.35]
Dairy Products (serving/day)		
≤Median	Reference	Reference
>Median	1.26 [0.93, 1.72]	1.27 [0.93, 1.73]
Red meat and products		
≤Median	Reference	Reference
>Median	1.21 [0.89, 1.63]	1·23 [0.91, 1.68]
Poultry (serving/day)		
≤Median	Reference	Reference
>Median	1.19 [0.89, 1.59]	1.22 [0.91, 1.65]
Olive oil (serving/day)		
≤Median	Reference	Reference
>Median	0.98 [0.73, 1.31]	0.99 [0.74, 1.34]
Ethanol intake (serving/day)		
≤Median	Reference	Reference
>Median	1.01 [0.72, 1.42]	1.01 [0.71, 1.42]

[1] Model adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity; [2] Mutually adjusted for all components of the MedDietScore

Table 13. All-causes death adjusted hazard ratios (95%CI) associated with *a posteriori* dietary patterns

Dietary pattern Components	HR 95%CI [1]	HR 95%CI [2]
Plant-foods		
≤Median	Reference	Reference
>Median	0.81 [0.60-1.09]	0.81 [0.60-1.09]
Animal foods		
≤Median	Reference	Reference
>Median	0.97 [0.71-1.32]	0.96 [0.70-1.31]
Dairy products and alcohol		
≤Median	Reference	Reference
>Median	0.99 [0.72-1.37]	0.96 [0.69-1.34]
Olive oil		
≤Median	Reference	Reference
>Median	0.93 [0.70-1.25]	0.93 [0.69-1.25]

[1] Model adjusted for sex, age, educational level, BMI, energy intake, smoking status, physical activity
[2] Mutually adjusted for all dietary patterns

APPENDIX 1: PUBLICATIONS, PROCEEDINGS AND SELECTED ABSTRACTS

Journal of Alzheimer's Disease 35 (2013) 349–361
DOI 10.3233/JAD-122140
IOS Press

A Meta-Analysis of the Efficacy of Donepezil, Rivastigmine, Galantamine, and Memantine in Relation to Severity of Alzheimer's Disease

Simona Gabriella Di Santo[a,*], Federica Prinelli[c], Fulvio Adorni[c], Carlo Caltagirone[a,b] and Massimo Musicco[a,c]
[a]*Fondazione IRCCS "Santa Lucia", Roma, Italy*
[b]*Università degli Studi di Roma "Tor Vergata", Roma, Italy*
[c]*Istituto di Tecnologie Biomediche – Consiglio Nazionale delle Ricerche, Segrate (MI), Italy*

Accepted 23 January 2013

Abstract.
Background: Randomized clinical trials have evaluated the efficacy of acetylcholinesterase inhibitors (AChE-Is) and memantine across a wide range of Alzheimer's disease (AD) severity. However, these drugs are prescribed and reimbursed according to precise upper and lower cut off scores of cognitive tests.
Objectives: To verify whether the efficacy of pharmacological treatment had any dependence on the severity of dementia in AD patients.
Methods: Published English-language randomized, placebo-controlled trials evaluating the efficacy of AChE-Is or memantine at any dose, over any length of time, in patients with any severity of dementia due to AD were included. Cognitive, behavioral, and functional outcomes were extracted from each study and multiple outcomes from the same trial were pooled to obtain a unique indicator of efficacy for cognition, functional impairment, and behavioral and psychological disturbances. The existence of a relationship between size of the treatment effect and severity of dementia, measured with the Mini-Mental State Examination, was determined using parametric and non-parametric correlation analyses.
Results: Both AChE-Is and memantine had significant effects on cognition. Functional and psycho-behavioral outcomes were reported less frequently but also showed significant efficacy of treatment. High heterogeneity among studies was found within and between the different drugs. The efficacy of all drugs except memantine was independent from dementia severity in all domains. Memantine effect on functional impairment was better in more severe patients.
Conclusions: The modest beneficial effects of anti-dementia drugs on cognition are independent from dementia severity. Memantine is more effective on functional incompetence only in severe patients.

Keywords: Activities of daily living, Alzheimer's disease, behavioral and psychological symptoms of dementia, cholinesterase inhibitors, cognition disorders, memantine, meta-analysis

Early Human Development 89S1 (2013) S51–S57

Contents lists available at ScienceDirect

Early Human Development

journal homepage: www.elsevier.com/locate/earlhumdev

Maternal milk protects infants against bronchiolitis during the first year of life. Results from an Italian cohort of newborns

Marcello Lanari [a], Federica Prinelli [b,c], Fulvio Adorni [b], Simona Di Santo [d], Giacomo Faldella [e], Michela Silvestri [f], Massimo Musicco [b,d,*], and Collaborators[1], on behalf of the 'Italian Neonatology Study Group on RSV Infections'

[a] *Pediatrics and Neonatology Unit, Imola Hospital, Imola, Italy*
[b] *Institute of Biomedical Technologies, National Research Council Milan, Italy*
[c] *Department of Food, Environmental and Nutritional Sciences, University of Milan, Italy*
[d] *Foundation IRCCS Santa Lucia, Rome, Italy*
[e] *Neonatology and Neonatal Intensive Care Unit, S. Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy*
[f] *Pediatric Pulmonology and Allergy Unit, Istituto Giannina Gaslini, Genoa, Italy*

ARTICLE INFO

Keywords:
Lower respiratory tract infection
Bronchiolitis
Respiratory syncytial virus
Breastfeeding
Risk factor

SUMMARY

Objective: Bronchiolitis is one of the primary causes of hospitalization in infancy. We evaluated the effect of breastfeeding on the occurrence of hospitalization for bronchiolitis in the first year of life.

Methods: In a prospective cohort study, 1,814 newborns of ≥33 weeks of gestational age (wGA) were enrolled in 30 Italian Neonatology Units and followed-up for 1 year to assess hospitalizations for bronchiolitis. Children were grouped as 'never breastfed' and 'ever breastfed'; these latter were further divided into those 'exclusively breastfed' and 'breastfed associated with milk formula'. The risk of hospitalization for bronchiolitis was evaluated with survival analysis, and hazard ratios (HR) with 95% confidence interval [95% CI] were calculated.

Results: Among enrolled newborns 22.9% were 'never breastfed'; in the breastfed group, 65% were 'exclusively breastfed' and 35% were 'breastfed with associated milk formula'. At 12 months of age, the risk of hospitalization for bronchiolitis was significantly higher in the 'never breastfed' group (HR: 1.57; 95% CI: 1.00–2.48). 'Breastfed associated with formula milk' and 'exclusively breastfed' groups were at similar risk of hospitalization for bronchiolitis. This observed protective effect of maternal milk was not explained by the higher prevalence of conditions able to increase the risk of bronchiolitis among 'never breastfed newborns'.

Conclusions: Breastfeeding, even in association with formula milk, reduces the risk of hospitalization for bronchiolitis during the first year of life. Encouraging breastfeeding might be an effective/inexpensive measure of prevention of lower respiratory tract infections in infancy.

Published Ahead of Print on July 10, 2013 as 10.1212/WNL.0b013e31829c5ec1

Inverse occurrence of cancer and Alzheimer disease

A population-based incidence study

Massimo Musicco, MD
Fulvio Adorni, PhD
Simona Di Santo, MSc
Federica Prinelli, MSc
Carla Pettenati, MD
Carlo Caltagirone, MD
Katie Palmer, PhD
Antonio Russo, MD

Correspondence to
Dr. Musicco:
massimo.musicco@itb.cnr.it

ABSTRACT

Objective: To evaluate the incidence of cancer in persons with Alzheimer disease (AD) and the incidence of AD dementia in persons with cancer.

Methods: This was a cohort study in Northern Italy on more than 1 million residents. Cancer incidence was derived from the local health authority (ASL-Mi1) tumor registry and AD dementia incidence from registries of drug prescriptions, hospitalizations, and payment exemptions. Expected cases of AD dementia were calculated by applying the age-, sex-, and calendar year–specific incidence rates observed in the whole population to the subgroup constituted of persons with newly diagnosed cancers during the observation period (2004–2009). The same calculations were carried out for cancers in patients with AD dementia. Separate analyses were carried out for the time period preceding or following the index diagnosis for survivors and nonsurvivors until the end of 2009 and for different types and sites of cancer.

Results: The risk of cancer in patients with AD dementia was halved, and the risk of AD dementia in patients with cancer was 35% reduced. This relationship was observed in almost all subgroup analyses, suggesting that some anticipated potential confounding factors did not significantly influence the results.

Conclusions: The occurrence of both cancer and AD dementia increases exponentially with age, but with an inverse relationship; older persons with cancer have a reduced risk of AD dementia and vice versa. As AD dementia and cancer are negative hallmarks of aging and senescence, we suggest that AD dementia, cancer, and senescence could be manifestations of a unique phenomenon related to human aging. *Neurology*® 2013;81:1–7

J Endocrinol Invest 36:1062-1068, 2013

Effect of mechanical and metabolic factors on motor function and fatigue in obese men and women: a cross-sectional study

Claudio L. Lafortuna[1], Federica Prinelli[2], Fulvio Adorni[2], Fiorenza Agosti[3], Alessandra De Col[3], Alessandro Sartorio[3,4]

[1]Istituto di Bioimmagini e Fisiologia Molecolare and [2]Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Segrate (MI), Italy.

[3]Laboratorio Sperimentale di Ricerche Auxo-endocrinologiche and [4]Divisione Malattie Metaboliche, Istituto Auxologico Italiano, IRCCS, Milano and Piancavallo (VB) Italy; 5 Dipartimento di Scienze per gli Alimenti, la Nutrizione e l'Ambiente, Università degli Studi di Milano, Italy.

Abstract

Background - Mechanical overload and poor quality of contractile elements related to metabolic abnormalities concur to motor disability of obesity. The independent contribution of these factors to motor dysfunction in obese individuals is scarcely defined.

Aim - Goal of the study is to test the hypothesis that metabolic factors may independently affect motor function in obesity.

Methods - Leg maximum power output per unit body mass (&W Mb), per unit fat-free mas (&WFFM) and fatigue in daily functioning were assessed in 635 obese (BMI≥35 kg/m2) individuals (286 men, 349 women) aged 19-78 yr. The independent effects of age, BMI, insulin resistance and the five components of the metabolic syndrome on &W Mb, &W FFM and fatigue were evaluated by multivariate analysis.

Results - A multiple regression analysis revealed that in both genders &W Mb (denoting the individual's performance capability during anaerobic tasks) was independently reduced by age ($p<0.001$), BMI ($p<0.05$-0.001) and abnormalities of glucose metabolism ($p<0.06$- 0.01), while &W FFM (representing the muscle intrinsic anaerobic capability) was affected only by age ($p<0.001$) and glucose metabolism impairment ($p<0.06$-0.01). In both genders fatigue was increased by age ($p<0.001$) and BMI ($p<0.05$-0.01), but augmented by low levels of HDL-cholesterol in men only ($p<0.05$).

Conclusions - Beside the dependence from mechanical overload and age, low muscle power output in obese individuals was independently associated also with metabolic abnormalities related to impaired glucose homeostasis. Fatigue and performance, although similarly influenced by age and body mass excess, are affected by different metabolic factors.

Early Human Development 89S4 (2013) S88–S90

Contents lists available at SciVerse ScienceDirect

Early Human Development

journal homepage: www.elsevier.com/locate/earlhumdev

Risk factors of hospitalization for lower respiratory tract infections in infants with 33 weeks of gestational age or more: a prospective Italian cohort study on 2210 newborns

Marcello Lanari[a], Federica Prinelli[b,c], Fulvio Adorni[b], Simona Di Santo[d], Massimo Musicco[b,d,*] and the "Italian Study Group on Risk Factors for RSV Hospitalization"[☆]

[a]Pediatrics and Neonatology Unit, Imola Hospital, Imola, Italy
[b]Institute of Biomedical Technologies, National Research Council Milan, Italy
[c]Department of Food, Environmental and Nutritional Sciences, University of Milan, Italy
[d]Foundation IRCCS Santa Lucia, Rome, Italy

In: British Journal of Nutrition (accepted)

Mediterranean Diet and other lifestyle factors in relation to 20-year all-causes death: a cohort study in an Italian population.

Federica Prinelli[1,2], Mary Yannakoulia[3], Costas A. Anastasiou[3], Fulvio Adorni[2], Simona G. Di Santo[4], Massimo Musicco[2,4], Nikolaos Scarmeas[5,6] and Maria L. Correa Leite[2]
[1] Department of Environmental, Food and Nutritional Sciences, University of Milan, Via Celoria 2, 20101, Milan, Italy; [2] Epidemiology and Biostatistics Unit, Institute of Biomedical Technology-National Research Council, Via Fratelli Cervi 93, Segrate (MI), Italy; [3] Department of Nutrition & Dietetics, Harokopio University, 70, El. Venizelou Ave. 76 71 Athens, Greece; [4] Department of Neuroscience, Santa Lucia Foundation IRCCS, Via Ardeatina 306, 00179 Rome, Italy; [5] Taub Institute for Research in Alzheimer's Disease and the Aging Brain, Gertrude H. Sergievsky Center, Department of Neurology, Columbia University, 622 W 168th St., P&S #16, New York, NY 10032, USA; [6] Department of Social Medicine, Psychiatry and Neurology, National and Kapodistrian, University of Athens, Eginition Hospital 72 Vasilisis Sofias Avenue, Athens, Greece, 11528

Abstract
Aim of the present analysis was to evaluate the association of Mediterranean Diet (MeDi), smoking habits and physical activity with all-causes death in an Italian population during 20-years of follow-up. A total of 1693 subjects aged 42-74, enrolled in 1991-1995 were asked about dietary and other lifestyle information at baseline. Adherence to the MeDi was evaluated by the Mediterranean Dietary Score (MedDietScore). A healthy lifestyle score was computed assigning 1 point for the medium or high adherence to the MedDietScore, nonsmoking and physical activity. Cox models were used to assess the associations between lifestyle factors and healthy score and all-causes death adjusting for potential confounders. Final sample included 974 subjects with complete data and without chronic disease at baseline. During a median of 17.4 years of follow-up, 193 people died. Subjects with high adherence to the MedDietScore [hazard ratio (HR) 0.62; 95% CI 0.43, 0.89), nonsmokers (HR 0.71; 95% CI 0.51, 0.98) and physically active (HR 0.55; 95% CI 0.36, 0.82) were at low risk of death. Each point increase in the MedDietScore was associated with a significant 5% reduction of the death risk. Subjects with 1, 2 and 3 healthy lifestyle behaviors had respectively a significantly 39%, 56%, and 73% reduced risk of death. A high adherence to MeDi, nonsmoking and physical activity were factors strongly associated with a

reduced risk of all-causes death in healthy subjects with long-term follow-up. This reduction was even stronger as a result of their combined effect.

Under submission

Memory impairment is not invariably the specific, distinctive and prominent cognitive marker of Alzheimer disease. A study on a cohort of 194 newly diagnosed patients.

Simona Gabriella Di Santo[1 2 6], Paolo Caffarra[3], Carlo Caltagirone[1 2], Carla Pettenati[4], Carlo Serrati[5], Fulvio Adorni[6], Federica Prinelli[6], Massimo Musicco[1 6] and the Italian Dementia Study of the Italian Society for the Study of dementias (SINDEM).

[1] Fondazione IRCCS "Santa Lucia" Rome; [2] University "Tor Vergata". Department of Neurosciences Rome; [3] University of Parma. Department of Neurosciences. Parma; [4] Alzheimer Unit. Passirana di Rho AO Casati Garbagnate Milan; 5 AO IRCCS San Martino Hospital. Genua; [6] Institute of Biomedical Technologies. Italian National Research Council. Segrate (Milan).

Abstract

Background/aims: The unique pathological process of Alzheimer Disease (AD) is expected to determine almost homogeneous disease phenotypes and memory loss is considered specific to AD. However, "variant" presentations exist, with prominent impairment in language, visuo-spatial and executive functions. We explored whether, in patients with typical AD, patterns of cognitive compromission similar to those observed in atypical AD could be identified.

Methods: The neuropsychological assessments of 194 newly diagnosed AD patients according to NINCDS-ADRDA criteria were considered. Four tests' corrected scores were considered representative of memory, language, visuo-spatial and frontal domains and standardized with reference to the mean values and standard deviations of the entire group of patients, in order to obtain comparable indicators of severity of impairment in the four domains.

Results: Four groups of patients of comparable numerosity, with the same mean severity of global cognitive impairment and mean disease duration, could be identified, according to the severity of the domain-specific cognitive impairment. Within each group, the standardized mean score of the more severely affected cognitive function was significantly lower than those of the other three tests, and the number of patients obtaining abnormal scores at the test for the most compromised domain was significantly higher than the number of patients with abnormal scores at other tests. Across the groups, the test scores of the most compromised domain and the number of patients with abnormal scores in the test for the most compromised domain were significantly lower than the scores obtained by and the number of abnormal patients at the corresponding test observed in the other three groups.

Conclusion: Our findings emphasize the variability of AD presentation and suggest that at least 4 cognitive profiles, recalling cognitive presentation of typical and AD variants might be recognized, thus questioning the characterization of AD as a disease of memory.

Oral communication and poster presented at the: 3rd International Conference on Clinical Neonatology, Torino, May 2012.
Early Human Development 88S2 (2012): S115

Breastfeeding protects infants against lower respiratory tract infections caused by respiratory syncytial virus. Results from the Italian Cohort Study Group on RSV.

F. Prinelli[1], M. Musicco[1], F. Adorni[1], on behalf of the Italian Neonatology Society[2] Study Group of RSV infections.
[1]Institute of Biomedical Technologies-National Research Council, Segrate (Mi), Italy; [2]Societa' Italiana di Neonatologia, Italy

Background: Newborns are susceptible to infections as a consequence of their immunological immaturity. Respiratory Syncytial Virus (RSV) is the most frequent pathogen responsible of infant lower respiratory tract infections (LRTI). The advantages of breastfeeding in reducing the occurrence of infectious diseases are widely documented but the role of breastfeeding in RSV-related LTRI is controversial.
Aims: To evaluate whether breastfeeding influence the occurrence of LRTI caused by RSV during the first year of life.
Subjects and Methods: These are preliminary data from a multicenter, prospective study that involved 30 Italian neonatology Units. A number of 1814 newborns of >33 weeks of GA were enrolled and followed-up during the first year of age. Conditions occurring during pregnancy potentially influencing the outcomes of the newborn, disturbances (respiratory and non-respiratory) of the newborn after delivery, and environmental risk conditions for RSV infection were recorded. After hospital discharge newborns were followed up until the end of the RSV epidemic season and at completion of the first year of life. Parents/Guardians were interviewed about hospitalisations and breastfeeding. Causes of hospitalisation were ascertained by the enrolling physician. With reference to mother's milk assumption during the first year of life, infants were categorized as never or ever breast fed (BF). The risk of occurrence of RSV-LRTI was evaluated with survival analysis; the relative risks (RR) were calculated considering as reference category the infants ever BF.
Results: In the entire cohort, on a total of 18,432 person-months of follow-up, we recorded 85 LTRI episodes with proven or suspected RSV infection etiology. About 7% of the LTRI episodes occurred in infants never BF (415) and 3.9% in infants ever BF. The crude RR for the never BF group was 1.5 (95% confidence interval [CI] 1.0–2.4) and 1.7 (95% CI 1.1–2.6) when adjusting for negative conditions of pregnancy, presence of disturbances after delivery and negative environmental exposure during the first year of life. The risk increase was independent from the characteristics of breast feeding

(exclusive or partial) and also considering as reference category the weaned infants not assuming breast milk.
Conclusions: According to these preliminary data, breastfeeding might confer protection from RSV induced LRTI in both exclusively or partially breast fed infants and even after weaning. These results are consistent with a positive role of breastfeeding in maturation of newborn immune system.

Poster presented at the: 19th European Congress of Obesity, Lione, May 2012.
Obes Facts 2012; 5 (suppl 1):86

Effect of mechanical and metabolic factors on motor function and fatigue in obese men and women.

A. Sartorio[1], F. Agosti[1], A. De Col[1], F. Adorni[2], F. Prinelli[2], C. Lafortuna[3]
[1] Istituto Auxologico Italiano, Laboratorio Sperimentale di Ricerche Auxo-endocrinologiche, Milano, Italy; [2] Istituto di Tecnologie Biomediche del CNR; [3] Istituto di Bioimmagini e Fisiologia molecolare del CNR

Background: Obesity is a major cause of motor disability deriving from a quantitative mechanical overload of contractile elements unbalanced to body mass excess and from a poor quality of skeletal muscles due to adipose tissue infiltration in relation to metabolic disturbances.
Methods: In 635 obese (BMI _35 kg/m2) individuals (286 men, 349 women) aged 19-78 yr., maximum power output per unit body mass (WMb) and the global sense of fatigue (GSF) were assessed with Margaria stair climbing test and with Fatigue Severity Scale questionnaire. The independent effects of age, BMI, insulin resistance and components of the metabolic syndrome on WMb and GSF were evaluated by a multivariate linear regression analysis.
Results: WMb and GSF resulted moderately correlated in women (Pearson's R: -0.20, $p<0.01$) and uncorrelated in men. A multiple regression analysis with stepwise selection procedures revealed that WMb was independently reduced in both genders by advancing age ($p<0.001$) and increasing BMI ($p<0.05$ in men, $p<0.001$ in women), while different metabolic negative determinants of motor performance were detected in men (fasting plasma glucose, $p<0.01$) and women (fasting plasma insulin, $p<0.05$), in line with a sex-dependent regulation of glucose metabolism. Gender-related differences were also detected in trends of GSF, which was increased by age ($p<0.001$) and BMI (men: $p<0.01$; women: $p<0.05$), but augmented by low levels of HDL-cholesterol in men only ($p<0.05$).
Conclusions: Metabolic abnormalities related to impaired glucose homeostasis significantly contribute to reduce muscle power output of obese individuals in addition to limitations deriving from mechanical overload and age. Global sense of fatigue and motor capabilities are affected by different metabolic factors, reflecting the multidimensionality of fatigue.

Oral communication and poster presented at the: 4rd International Conference on Clinical Neonatology Torino, May 2013.
Early Human Development 89S1 (2013): S90

Exposure to traffic pollution and risk of hospitalization for bronchiolitis in the first year of life. Results from a cohort of 1814 Italian newborns.

F. Prinelli[1,] F. Adorni[1], S. Di Santo[2], M. Silvestri[3], M. Lanari[4], M. Musicco[1].
[1]Epidemiology and Biostatistics Unit, Institute of Biomedical Technologies (ITB), National Research Council (CNR), Milan, Italy;
[2]Clinical Epidemiology Unit, Foundation IRCCS Santa Lucia, Rome, Italy;
[3]Pediatrics Pulmunology and Allergy Unit, Giannina Gaslini Hospital, Genoa, Italy; [4]Pediatrics and Neonatology Unit, Imola Hospital, Imola, Italy

Objective: Bronchiolitis is the leading cause of hospitalization in the first year of life. Several studies investigated the association between exposure to traffic pollutants and the risk of respiratory infections; however, few epidemiological data on hospitalization for bronchiolitis are available. Aim of this study was to evaluate the effect of traffic pollution exposure on the occurrence of bronchiolitis in an Italian cohort of newborns.
Methods: In a multicenter prospective cohort study, newborns of >33 weeks of gestational age were enrolled in 30 Neonatology Units and their prenatal, neonatal and postnatal data were recorded. After hospital discharge, newborns' parents were interviewed at the end of the bronchiolitis epidemic season and at one year about hospitalizations and exposure to traffic pollution, defined as living in urban area with intense road traffic. Causes of hospitalization were ascertained by the enrolling physician. The risk of bronchiolitis was evaluated with survival analysis and hazard ratios (HR) with 95% confidence intervals (95%CI) were calculated.
Results: Out of 2,154 newborns recruited, 1,814 had complete data for the analysis. Eighty-five bronchiolitis were observed, 16/221 in infants exposed to traffic pollution. The HR of bronchiolitis for exposed infants was 2.0 (95% CI 1.1–3.4) adjusted for prenatal, neonatal and other postnatal conditions as presence of siblings, parents' smoking habits and no breastfeeding.
Conclusions: Our results showed that exposure to traffic pollution significantly increases the risk of children's hospitalization for bronchiolitis, independently from other known neonatal and postnatal risk conditions. Enforcing the strategies on relieving traffic congestion might consistently reduce the burden of this disease.

Oral communication and poster presented at the: Sindem4junior 2nd winter seminar on Dementia, Bressanone, January 2014 and VIII Congresso Sindem, Firenze, March 2013.

Occurrence of cancer and Alzheimer'disease are inversely associated in elderly persons: a population‐based study.

Federica Prinelli[1], Massimo Musicco[1,2], Dr Carlo Caltagirone[2]
[1] ITB-National Research Council, Epidemiology and Biostatistics Unit; [2] Santa Lucia Foundation , Clinical Neurology Laboratory

Background and Objectives: A lower incidence of Alzheimer' disease (AD) in people who have cancer, and less cancer in persons with AD has been reported. The study evaluated the risk of co-occurrence of AD and cancer independently from the temporal order of appearance of each disease in a large general population.
Subjects and Methods: This is a cohort study in Northern Italy of about one million of residents. Cancer incidence in the whole population was derived from the Local Health Authority (ASL-Mi1) tumor registry and AD incidences from registries of drug prescriptions, hospitalizations and payment exemptions. The risk of AD in persons with cancer and vice-versa, relative to general population, was estimated as observed timed expected cases for from 2004 to 2009. Expected cases of AD were calculated applying the age-, sex-, and calendar year specific incidence rates observed in the whole population of ASL-Mi1 to the subgroup constituted of person with newly diagnosed cancers during the period of observation. The same calculations were carried-out for cancers in AD persons. Separate analyses were carried-out for the time period preceding or following the index diagnosis, for survivors and non-survivors until the end of 2009 and for different types of cancer.
Results: The risk of cancer in person with AD was halved and the risk of AD in persons with cancer was 35% reduced. This inverse relationship of occurrence was observed in almost all the subgroup analyses suggesting that potential confounding factors did not influenced the results.
Discussion: The occurrence of both cancer and AD increases exponentially with age, but with an inverse relationship; older person with cancer have a reduced risk of having AD and vice versa. Since AD and cancer are negative hallmarks of aging and senescence we suggest that AD, cancer and senescence could be manifestation of a unique phenomenon related to human aging.

Oral communication at the: IX Congresso Sindem, Firenze, March 2014
Journal of Alzheimer's Disease (2014) 41; S41-S42

Is there a link between modifiable risk factors and the inverse occurrence of cancer and neurodegenerative diseases?

Federica Prinelli.
Epidemiology and Biostatistics Unit, Institute of Biomedical Technologies (ITB), National Research Council (CNR), Milan, Italy.

Background and Objectives: Many different factors related to genetics and environment could be responsible for decreased co-occurrence of neurodegenerative diseases (ND) and cancer. Aim of this study was to evaluate the role of dietary habits, lifestyle and clinical factors in determining the risk of cancer in persons with ND and vice-versa.
Subjects and Methods: This is a prospective cohort study on a population of residents in a district of Northern Italy aged 40-74 during the period from 1991 to 1995. For 1693 subjects, socio-demographic, anthropometric, clinical, dietary and lifestyle variables were collected at baseline. Incident cancer cases and incident ND cases, including Alzheimer's disease and Parkinson's disease, were ascertained using a passive case-finding procedure. Cox proportional hazards regression was used to evaluate associations between exposure and incident cancer in persons with ND and vice-versa. The multivariate model included gender, baseline age, educational level, body mass index (BMI), fasting glucose, HDL-cholesterol, blood pressure, smoking status, alcohol intake, physical activity, Mediterranean Diet (MeDi) score adherence and energy intake.
Results: In total, 54 incident ND cases and 347 cancer cases had occurred (9 co-occurrence). People with incident cancer were younger, more likely to be male, more educated, more smokers, less active physically, reported slightly less adherence to MeDi and were more deceased during follow-up than those with ND. The two groups did not differ by BMI, glucose, HDL, blood pressure, alcohol intake. The risk of cancer was 64% reduced in persons with ND (95% Confidence Intervals 0,15-0,87) and in subjects with cancer the ND risk reduction was
similar. These results were independent of socio-demographic, anthropometric, clinical, dietary and life-style factors.
Conclusions: This study confirms that subjects with cancer have a reduced risk to develop neurodegenerative diseases compared with individuals without cancer and vice-versa independently of environmental exposure, supporting the hypothesis of a role of genetics in determining this inverse relationship.

Oral communication at the: IX Congresso Neuroepidemiologia, Milano, Novembre 2014.
Neuroepidemiology (2014) 43; 169

Administrative data and inverse occurrence of cancer and Alzheimer's disease in elderly people I: a methodological approach to control for biases due to under diagnosis.

Federica Prinelli, Fulvio Adorni, Massimo Musicco
Istituto di Tecnologie Biomediche – Consiglio Nazionale delle Ricerche di Milano

Introduction: Lower incidence of cancer in persons with Alzheimer disease and vice-versa has been previously reported, however this might be due to an under-diagnosis of the second disease after the occurrence of the first. Aim of this study was to estimate the unbiased co-occurrence of the two diseases in a population of about one million persons, using administrative health data (AHD).
Subjects and Methods: In AHD we identified incident Cancer and AD cases over a period of five years. To control for under-diagnosis of the second disease three analyses were carried-out: for times preceding or following the index diagnosis, for surviving or not the follow-up period and for different types and sites of cancer. The relative risks (RRs) of AD in persons with cancer and vice-versa, relative to general population, were estimated as observed timed expected cases.
Results: The RRs of cancer in persons with AD and of AD in persons with cancer were both reduced. The inverse relationship of occurrence was independent from the temporal order of appearance of the diseases, from life-expectancy reduction and from different cancer types.
Discussion: Administrative data represent a valid source for the conduction of cohort studies on the co-occurrence of diseases.

Poster presented at the: Congresso Neuroepidemiologia, Milano, November 2014

Neuroepidemiology (2014) 43; 156

Administrative data and inverse occurrence of cancer and Alzheimer's disease in elderly people II: a misclassification problem for the diagnosis of Alzheimer's disease.

Fulvio Adorni , Federica Prinelli, Massimo Musicco
Istituto di Tecnologie Biomediche – Consiglio Nazionale delle Ricerche di Milano

Introduction: In a cohort of 115,524 person-years we observed a 40% risk reduction for the occurrence of Alzheimer's disease (AD) in persons with cancer and vice-versa. A relevant concern about these findings is about the potential effect of misclassification in biasing the results.
Subjects and Methods: All cancer cases were identified from a population Tumor Registry. AD cases were traced from registries of drug prescriptions, hospitalizations and payment exemptions by a classification algorithm that had high specificity but lacked in sensitivity. In the co-occurrence study misclassification of "exposure" (cancer occurrence in the AD cohort) and of disease (AD in cancer cohort), might be alternatively differential (DM) or non-differential (NDM). A sensitivity analysis was performed in this cohort.
Results: An exposure or disease NDM had the only effect to overestimate the relative risks (RR), weakening the strength of the inverse association. The effect of a DM, unlikely to have occurred, was dependent from levels of sensitivity/specificity in the subgroups of exposure and of disease, and led to biased RR estimates only when classification performances were highly differential.
Discussion: Large population-based studies on administrative data allow to obtain epidemiological results accounting for possible misclassifications.

Acknowledgments

First, I would like to thank my thesis supervisor Prof. Amelia Fiorilli, for the help and guidance she has provided during these three years.

Thank to my tutor, Prof. Ambrogina Pagani.

I would like to thank Prof. Monica Ferraroni of the Department of Clinical Sciences and Community Health of the University of Milan, who has assisted me with her precious advices.

In particular, I thank persons of the Epidemiology and Biostatistics Unit of the ITB-CNR. Thank to Dr. Massimo Musicco for the opportunities he has given me, contributing to my professional growing during the thesis period.
Especially, I would like to thank Fulvio for his continuous and unconditional support to my research work, thank to Lea for her previous work on this project, they always listened me and provided encouraging words.

Finally, thank to Mary and Costas from the Harokopio University in Athens, for the pleasant collaboration during my work period in Greece and for their interest for my project.