MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-682

Improving Clinical Decisions on T2DM Patients Integrating Clinical, Administrative and Environmental Data

Daniele Segagni[a], Lucia Sacchi[b], Arianna Dagliati[b], Valentina Tibollo[a], Paola Leporati[c], Pasqale De Cata[c], Luca Chiovato[c], Riccardo Bellazzi[b]

[a] *Laboratory of System Engineer for Clinical Research, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy*
[b] *Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy*
[c] *Unit of Internal Medicine and Endocrinology, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy*

Abstract

This work describes an integrated informatics system developed to collect and display clinically relevant data that can inform physicians and researchers about Type 2 Diabetes Mellitus (T2DM) patient clinical pathways and therapy adherence. The software we developed takes data coming from the electronic medical record (EMR) of the IRCCS Fondazione Maugeri (FSM) hospital of Pavia, Italy, and combines the data with administrative, pharmacy drugs (purchased from the local healthcare agency (ASL) of the Pavia area), and open environmental data of the same region. By using different use cases, we explain the importance of gathering and displaying the data types through a single informatics tool: the use of the tool as a calculator of risk factors and indicators to improve current detection of T2DM, a generator of clinical pathways and patients' behaviors from the point of view of the hospital care management, and a decision support tool for follow-up visits. The results of the performed data analysis report how the use of the dashboard displays meaningful clinical decisions in treating complex chronic diseases and might improve health outcomes.

Keywords:
Integrated Systems; Therapy Adherence; Clinical Pathways; Data Integration; Type 2 Diabetes.

Introduction

Treatment and management of chronic diseases often takes place outside clinical settings and impacts the daily life of patients. As a consequence, clinicians depend on patient reports of symptoms, side effects, functional status and treatment adherence. Patients typically report at clinical visits that are months apart, and recall accuracy can be highly fluctuating [1]. The possibility to gather and analyze accurate information at the right time from different sources, such as public healthcare systems, open data repositories, or hospital information systems is necessary to be able to perform correct clinical decisions [2]. The work presented in this paper aims to provide clinicians and researchers a software architecture based on data coming from heterogeneous data sources to support medical discovery and evidence-based practice about managing and controlling the evolution of chronic diseases. The system architecture relies on the open source i2b2 clinical data warehouse (CDW) [3] and the work done has been focused on expanding the basic functionalities of the framework by enhancing its visualization layer. A number of modules have been designed and developed to present the results through an intuitive and easy to use dashboard that shows results of process mining, temporal abstractions and similarity algorithms.

Materials and Methods

Type 2 Diabetes Mellitus (T2DM) is the most common form of diabetes. It accounts for at least 90% of all cases of diabetes. The World Health Organization (WHO) estimates that by 2030 there will be about 550 million people suffering from this disease [4]. This disease can remain undetected for many years because hyperglycemia (consequence of the insulin defects) develops gradually, and at earlier stages, the disease is not severe enough for the patient to notice any of the classic symptoms of diabetes. Despite the large number of models being developed and the increased interest and acknowledgement in the clinical field, only a small part of these models ends up being used in clinical practice [5]. One of the most important challenges of the MOSAIC European Union project is to combine the research activities related to the discovery of new risk factors, methods and models for diabetes onset, progression and evolution, with the development of software tools. The components and modules would incorporate innovations and make them usable by different end-users in a variety of settings. The biggest challenge of the project consists in translating innovations to achieve impact in the current clinical practice.

Data sources

To fill the gap resulting from infrequent clinical follow-ups of diabetic patients, the EMR used in FSM, a private hospital in Pavia, Italy, has been enhanced with data coming from the local public healthcare agency (ASL) of the Pavia area. This data contains administrative findings (e.g., prescription-based drug purchases), and data reporting environmental information (e.g., air temperature, air pollution, etc.) provided by the "Regione Lombardia" databases as open data.

Clinical usage

The main component of the software tool developed consists in a single page dashboard, where users can interact with a graphical presentation of clinical parameters. Dashboard users can analyze data though three different use cases:

- Use Case 1 (UC1): risk factors and indicators to improve current detection of T2DM

- Use Case 2 (UC2): hospital care management
- Use Case 3 (UC3): clinical decision support during follow-up visits

In this paper we will focus on use case 2 and 3, which are the ones that will be implemented for the management of already diagnosed T2DM patients at FSM. UC1 case is active in another research center participating in the MOSAIC project: the Hospital Universitari i Politècnic La Fe located in Valencia, Spain. UC1 will be integrated also in FSM by the end of the project, to allow a full evaluation of the system.

Architecture

We designed and developed a dashboard (Figure 1) to accomplish the goals described in UC2 and UC3. The dashboard communicates with the i2b2 DW where clinical, administrative and environmental data are integrated. A data mining module is responsible for running advanced temporal data mining algoritms on data queried from the CDW.

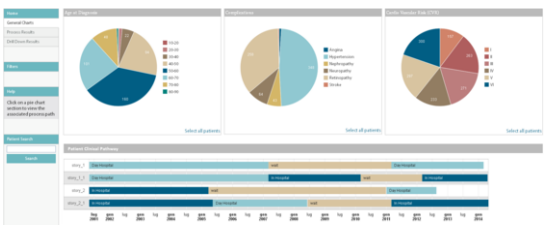

Figure 1 – starting page of the MOSAIC dashboard: charts show patients grouped by age at diagnosis, types of complications and cardio-vascular risk. The timeline displays four clinical pathways related to the selected patient set.

From this section, the user can either choose to select a subset of the population to use for further analyses or to search for an individual patient to examine his/her clinical charts. In the following paragraphs we will present some of the functionalities related to the two mentioned use cases.

Use Case 2.

UC2 is focused on the analysis of the clinical histories of a population of patients. This population could either be the whole patients' sample included in the data repository or a subset of patients conveniently selected by the user according to some clinical criteria available in the starting page of the dashboard. The models developed for data analysis rely on temporal and process mining techniques and are aimed at detecting the most frequent clinical patterns that are experienced by the patients' population [6, 7]. These models are focused on taking explicitly into account the temporal dimension, which strongly characterizes the evolution of chronic diseases. The ultimate goal of applying these techniques is to extract meaningful patients' stratifications that are based on the temporal evolution of some interesting clinical variables. For example, the most critical pathways in terms of severity that arise from complications and use of hospital resources could be highlighted. Temporal data mining is employed to understand which are the most interesting variables to use and to exploit the novel perspectives offered by temporal data integration.

Clinical histories are mined to extract behavioral patterns of prescription-related drug purchases, frequent clinical temporal patterns, cardio-vascular risk (CVR) profiles, level of complexity and evolution of complications. The extracted temporal patterns are presented to the user as timelines. The user can select a specific temporal pattern to drill down to the patients' population experiencing that behavior.

The target population of UC2 is of patients already diagnosed with T2DM. The data used to build and train UC2 models come from the FSM hospital EMR and are related to a cohort of T2DM patients followed up from 5 to 10 years. The EMR information related to these patients has been enhanced with public data coming from the Pavia ASL and with environmental information. This kind of data contains the history of all kind of therapies, hospitalizations, outpatient services, patient geo-localization, air temperature and degree of air pollution. To assess the complexity of care of the patients over time, a set of complexity stages, determined by the number of complications and hospitalizations patients undergo during the course of their disease, has been defined. The computation of these complexity stages is made possible by the availability of both clinical data related to the onset of complications and administrative information on hospitalizations and outpatient visits. Assuming that each patient starts with a stable state (it refers to pre-diabetic patients and type 2 diabetes patients who do not suffer any complications yet), when a complication arises the state shifts to “level 1.” If a second complication appears later than one year or if the second complication is a myocardial infarction, the state shifts to “level 2,” while if a second complication appears earlier than one year, the patient’s state shifts directly from “stable” to “level 2.” When an hospitalization, either due to the status of their diabetes-related complications or because of their metabolic instability occurs, the state change from “level 1” or “level 2” to "level 3.”

Use Case 3.

This use case refers to supporting clinicians to better manage T2DM chronic patients. Starting from patient’s EMR record, the system is able to project in a graphic way clinical information and behavioral patterns of prescription-related drug purchases. In this use case, clinicians are interested in having a better picture of what had happened to a single patient and what needs to be improved (e.g., changes in therapy, prescriptions of special examinations). Integrating information coming from the ASL and from the FSM EMR, analyzing and making them available in a unique tool, is a first step to achieve this goal. The part of the dashboard related to UC3 depicts data related to an individual patient, with a specific focus in the integration of data coming from different sources. Data integration is very important, as the clinical information related to the period that goes from the onset to the first follow-up at FSM is missing. The longer this period lasts, the more information becomes unobservable. This can be a limitation, especially if the objective of the analysis is to build predictive models (e.g. for complications prediction), as the events that occur in this not known time window might be the early determinant of a complication to arise. The data integration performed helps to mitigate this problem. In fact, it offers some additional information which comes from the public healthcare data. With this data, the information related to the not known time window becomes partially observable, given that the patient has been diagnosed after 2003.

The dashboard functionalities related to UC3 involve clinical and administrative data representation. Regarding clinical information, the dashboard reports temporal data related to (1) HbA1c values from laboratory exams performed within FSM hospital, (2) the evolution of the patient’s level of complexity and (3) the time course of the cardiovascular risk (Figure 2).

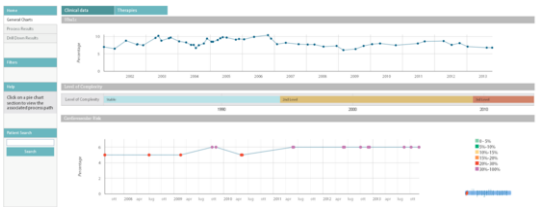

Figure 2 – MOSAIC dashboard example representing clinical valuesevolution over time.

One of the largest parts of information contained in the ASL data warehouse is related to prescription-based drug purchases by citizens at the pharmacy. Since this information is not directly available to the physician during clinical practice, it is very important to visualize it in a meaningful way in the dashboard. In particular, we have focused on drug purchase representation and on evaluation of the adherence to treatment behavior of the patient. Each drug purchase has been inserted in the i2b2 CDW using its Defined Daily Dose (DDD) that allows the expected number of therapy days related to that purchase to be computed. As the information about the actual drug dosages a patient should take is not available in our datasets, we use drug purchase as a proxy for estimating drug intake. Under this assumption, the DDD information represents the days supply for a specific purchase and we use DDDs to evaluate patients' behavioral patterns related to drug treatments. For each patient we retrieved the first prescription of a specific drug and divided the whole prescription period in semesters. For each semester we compute the sum of DDDs, which is an estimate of the total drug intake of the patient in that semester. We chose to use six months as an observation chunk, as in the current practice at FSM the average time between two consecutive follow-ups for chronic patients is one semester. Setting this temporal constraint gives the possibility of defining the granularity to observe modifications, adherence and continuity in drug treatments. As result a profile for each patient was constructed including the date of each purchase and the time period covered by the dispensed drugs, so to define the purchase pattern over time. In addition to the temporal profiles describing drug purchases, specific attention has been devoted to investigating the behavior of the patient in terms of treatment adherence. In recent years, several works studied the problem of harmonizing the integration of different healthcare information coming from heterogeneous sources while processing data coming from EHR or administrative databases [8]. One of the most common measures is the medication possession ratio (MPR) that indicates the percentage of received days supply divided by a period of time [9]. In our analyses, we have computed adherence to treatment using a validated index that corresponds to the MPR index for drug acquisition. The index is the Continuous, Single Interval Measure of Medication Acquisition (CSA). CSA index is calculated as the ratio: sum of DDDs over an observation period/length of the observation period (in our case one semester, 182 days). The CSA index indicates the percentage of days covered by the prescription. Besides DDDs, the information on CSA has also been included in the CDW. According to the CSA value, patients have been grouped into four clusters: no adherence (range of adherence between 0% and 40%), poor adherence (range between 40% and 80%), adherence (range between 80% and 100%), and over adherence (more than 100%). Patients belonging to the last group purchase, in a specific time window, more drugs than their prescription. The graphical representation of DDD values and therapy adherence percentage are shown in Figure 3.

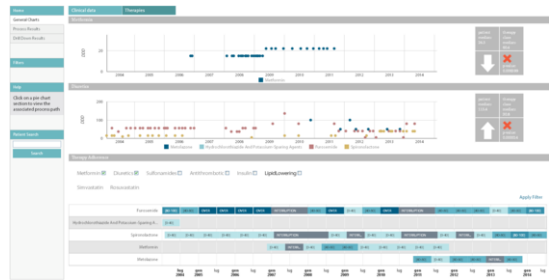

Figure 3– MOSAIC dashboard example representing therapies DDD and adherence

Results

In this section we will present the results we obtained on a first sample of 433 patients using the methodologies described in the dashboard. Retrieving data from FSM EMR we had to deal with the fact that diabetic patients do not start to be followed at the hospital immediately after diabetes diagnosis. In Italy, this situation happens because this type of chronic patient is initially managed by the general practitioner (GP) who manages the patient until the GP believes it is more suitable for the patient to start being followed at the hospital outpatient service. Despite the wide variability of the period of time elapsed between onset of the first encounter, the majority of the patients start to be followed by FSM at a stable stage (73%). Analyzing the pattern of development of the first complication for our patients' sample and computing the time between the detection of the first complication and the first encounter at FSM, we observed how the diagnosis of the first complication occurs close to the first visit at the hospital. This is reasonable, if we consider that it is often the worsening of a patient's conditions that triggers a GP's decision to send a patient to the hospital outpatient service. As soon as the patient is managed by the hospital, the process that allows diagnosing possible complications is started, resulting in the recording of the complication in the EMR. Table 1 shows the distribution of the first complication for the patients who developed it after the first encounter at FSM.

Table 1 – Distribution of patients' first complication developed after first visit at FSM

Type of complication	% of development
Occlusion and stenosis of carotid artery	32.5 %
Fatty liver disease	25.8%
Nephropathy	12.8%
Retinopathy	9.7%
Neuropathy	8.2%
Peripheral vascular	6.9%
Chronic ischemic heart disease	4.1%

The administrative data related to drug purchases of about 433 patients considered in the analysis process consist in 244.190 records and a number of 21.055 distinct ATC codes. Classification of the purchased drugs on the basis of clinical relevance was the first type of analysis that we performed. In this way it was possible to reduce the variety in the data set to build simpler and meaningful scenarios based on treatment behaviors. To this end, the following ATC classes have been selected: A10 (drugs used in diabetes), B01 (antithrombotic agents), C02 (anti-hypertensives), C03 (diuretics), C10 (lipid modifying agents). The resulting codes have been further

grouped in 17 classes defined on the basis of medical knowledge and of the ATC-WHO system [10]. Table 2 details the number of patients and the number of box purchases per drug. Drugs that are not specific for the treatment of diabetes have been considered at a higher level of the ATC taxonomy, whereas drugs used to treat diabetes have been analyzed at a deeper level.

Table 2 – Knowledge-based classification of the ATC codes considered in our analysis (drugs not directly related to diabetes treatment)*

Drug	Patients	N° of Purchases
Alpha glucosidase inhibitors	85	331
Antihypertensive *	73	554
Antithrombotic *	338	3598
Phenformin_Sulfonamides	7	17
Diuretics *	157	1542
Gliptin	80	220
Incretin	23	79
Insulin	103	1381
LipidLowering *	303	3484
Metformin	352	3318
Metformin_Incretin	20	66
Metformin_Sulfonamides	73	709
Metformin_Thiazolidinediones	16	56
Repaglinide	98	638
Sulfonamides	150	1108
Sulfonamides_Thiazolidinediones	1	4
Thiazolidinediones	36	137

After computing the CSA values, we fixed a threshold (80%) to define if in a period, the patient purchased the appropriate dosage of the treatment, according to our assumptions. If the CSA is lower than 80% we associate a label to the semester indicating that there is Adherence = NO, otherwise we state Adherence = YES. We also represent the semesters during which there are no prescription as Adherence = INTERRUPTION and semesters where the purchases exceed the number of days in the time windows as Adherence = OVER. Table 3 summarizes the count of semesters on the basis of the CSA value.

Table 3 - Overall number of semesters classified as CSA >=80% (YES), CSA<80% (NO), CSA>100% (OVER), no prescription (INTERRUPTION)

Drug	YES	OVER	NO	INTERRUPTION
Alpha glucosidase inhibitors	39	34	253	5
Antihypertensive *	61	133	296	28
Antithrombotic *	875	555	1515	291
Biguanides_Sulfonamides	7	7	3	
Diuretics *	126	147	883	151
Gliptin			74	5
Incretin	58	66	92	6
Insulin	109	286	429	58
LipidLowering *	582	714	1587	228
Metformin	490	694	1932	202
Metformin_Incretin	21	23	21	1
Metformin_Sulfonamides	108	359	188	54
Metformin_Thiazolidinediones	18	10	28	1
Repaglinide	61	161	375	41
Sulfonamides	158	382	488	74
Sulfonamides_Thiazolidinediones	1	1	1	
Thiazolidinediones	32	23	80	2

Discussion

Despite the different values of CSA medians for different ATC groups, we analyzed the results in order to extract reliable and accurate patient's characterization on the basis of the general purchase behaviors. Our analysis has been focused on detecting main patterns of drug purchases. We identified the following subject classes:

- patients who purchase less drug boxes than the population taking the same drug class (CSAPatient<CSAPopulation)
- patients who purchase more drug boxes than the population taking the same drug and for whom the median CSA values stay below the 100%, not exceeding the recommended dosage (CSAPopulation <CSAPatient<100%)
- patients who purchase more drug boxes than the population taking the same drug but whose median CSA values stay above the 100%, exceeding the recommended dosage (100%<CSAPatient< CSAPopulation)

By identifying these groups it is possible to detect those patients who behave differently from the population of patients treated with the same drugs. A different purchase pattern could identify critical patients, independently from the CSA value. The developed procedure addresses the stratification challenge thanks to three consecutive steps, which are devoted to transform raw data into meaningful features and then using these features to in-depth tailoring specific patients' traits.

Step 1: Detection of subjects with statistically significant different purchase patterns from the population.

For each patient and each drug calculated, the median values of adherence for the whole prescription period were compared to the median value of the patient with the median value of the population has been performed (Wilcoxon Rank-Sum Test to asses is significant different, $p < 0.05$). If $p<0.05$, the patient was been assigned to one of two groups

Step2: Profiling of the $CSA_{Patient}$>$CSA_{Population}$ group.

For each patient belonging to this group and each drug check, if median values are <100% then classify as $CSA_{Population}$ <$CSA_{Patient}$<100% (ADHERENT) else if median values are >100% then classify as 100%<$CSA_{Patient}$< $CSA_{Population}$ (OVER ADHERENT)

Step3: Detect an overall purchase behavior

Each patient has the count of the number of drugs where the patient is UNDER - ADHERENT – OVER. If the behavior is detected for >50% over the total drug purchased then classify as: Under Adherent (LESS than POPULATION), Adherent (MORE than the population but less than 100%), Over Adherent (MORE than the population and more than 100%).

After we applied the process and retrieved the adherence stratification, we compared laboratory tests values and life style data in the different mined groups to understand if behavioral patterns of prescription-based drug purchases can be used as

biomarkers of specific clinical conditions of the patients who verify those patterns. The clinical variables (BMI, cholesterol and HbA1c), show significant differences ($p<0.01$, calculated with Kruskal-Wallis test) in the three groups. In particular, the patients identified as Adherent have better clinical values that the other groups. For BMI and Cholesterol, this difference becomes even more apparent ($p <0.01$, calculated with Wilcoxon test) when the Adherent group is highlighted and compared with the two Under or Over adherent groups considered together. This comparison emphasizes the clinical conviction that patients that not have a good adherence to therapy might suffer from metabolic disorders. Furthermore, considering other observations about patients diet, it is possible to detect better habits in adherent patients. During the whole observation period, adherent subjects have 118 observations of good diet habits, while they show bad habits only in 44 visits. The dashboard has been used to represent this information for UC3. Four data items are indicated: the median value of patient DDDs, the median value of group DDDs, the p-value resulted from Kruskal-Wallis test made to verify if there is difference between the patient and the group DDD median value, and an arrow used as visual indicator to display a positive or a negative difference between medians. Furthermore, we integrated from the ASL data warehouse, the cardiovascular risk index calculated as defined by the Italian Ministry of Health in the CUORE project [11] that takes into account values of gender, age, smoke habits, systolic pressure, cholesterol, HDL cholesterol, the diagnosis of diabetes and use of anti-of hypertensive drugs to produce the risk value.

Conclusion

In this paper, we presented an informatics approach that addresses the challenge of providing adequate clinical information such as therapy adherence, evolution of clinical pathways, and levels of complexity during inpatients encounters. The aim of our implementation is to improve the access to medical information through integrated informatics techniques necessary to gather data from different and heterogeneous sources such as hospital EMRs and public health or administrative records. By providing information in advance to physicians and not following the traditional approach based on questions to patients, we want to minimize the rise of bias and misunderstanding that can occur while patients report about medications and therapies they are following. With the introduction of the developed system into work practice and incorporating it into the hospital EMR, it will be easier to continually maintain the assessment of T2DM patients during their follow-up visits. Moreover, using the developed dashboard, clinicians are able to compare patient data they are visiting with a similar group on the basis of the variable they are interested. Analyzing data at the individual level and comparing them to similar group might have more impact in predicting how the clinical situation could evolve and might be useful for clinicians to perform correct clinical decisions.

Acknowledgement

This work is part of the MOSAIC EU project, funded by the 7th Framework Programme (http://www.mosaicproject.eu/).

References

[1] Kim C, Williamson DF, Herman WH, et al. Referral management and the care of patients with diabetes: the Translating Research Into Action for Diabetes (TRIAD) study. Am J Manag Care 2004;10(2 Pt 2):137

[2] Estrin D, Sim I. Health care delivery. Open mHealth architecture: an engine for health care innovation. Science. 2010 Nov 5;330(6005):759-60.

[3] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010, 17(2):124-130.

[4] Centers for Disease Control and Prevention. Diabetes report card 2012. In: US Department of Health and Human Services, ed. Atlanta, GA: Centers for Disease Control and Prevention, 2012:1–4. http://www.cdc.gov/Diabetes/pubs/reportcard.htm. Last access: December 22, 2014.

[5] Kengne AP, Beulens JW, Peelen LM, Moons KG, van der Schouw YT, Schulze MB, Spijkerman AM, Griffin SJ, Grobbee DE, Palla L, Tormo MJ, Arriola L, Barengo NC, Barricarte A, Boeing H, Bonet C, Clavel-Chapelon F, Dartois L, Fagherazzi G, Franks PW, Huerta JM, Kaaks R, Key TJ, Khaw KT, Li K, Mühlenbruch K, Nilsson PM, Overvad K, Overvad TF, Palli D, Panico S, Quirós JR, Rolandsson O, Roswall N, Sacerdote C, Sánchez MJ, Slimani N, Tagliabue G, Tjønneland A, Tumino R, van der A DL, Forouhi NG, Sharp SJ, Langenberg C, Riboli E, Wareham NJ. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. Lancet Diabetes Endocrinol. 2014 Jan;2(1):19-29.

[6] Wright AP, Wright AT, McCoy AB, Sittig DF. The use of sequential pattern mining to predict next prescribed medications. J Biomed Inform. 2014 Sep 16.

[7] Sacchi L, Dagliati A, Bellazzi R. Analyzing complex patients' temporal histories: new frontiers in temporal data mining. Methods Mol Biol. 2015;1246:89-105.

[8] Raebel MA, Schmittdiel J, Karter AJ, Konieczny JL, Steiner JF. Standardizingterminology and definitions of medication adherence and persistence in researchemploying electronic databases. Med Care. 2013 Aug;51(8 Suppl 3):S11-21.

[9] Kozma CM, Dickson M, Phillips AL, Meletiche DM. Medication possession ratio:implications of using fixed and variable observation periods in assessing adherence with disease-modifying drugs in patients with multiple sclerosis. Patient Prefer Adherence. 2013 Jun 12;7:509-16.

[10] The ATC/DDD Index 2015: http://www.whocc.no/atc_ddd_index. Last access: December 22, 2014.

[11] Donfrancesco C, Palmieri L, Cooney MT, Vanuzzo D, Panico S, Cesana G, FerrarioM, Pilotto L, Graham IM, Giampaoli S. Italian cardiovascular mortality charts of the CUORE project: are they comparable with the SCORE charts? Eur J CardiovascPrev Rehabil. 2010 Aug;17(4):403-9. 1995;2:316 –322

Address for correspondence

Daniele Segagni: daniele.segagni@fsm.it