World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:15, No:1, 2021

Building an Integrated Relational Database from Swiss Nutrition National Survey and Swiss Health Datasets for Data Mining Purposes

Ilona Mewes, Helena Jenzer, Farshideh Einsele

***Abstract*—**Objective: The objective of the study was to integrate two big databases from Swiss nutrition national survey (menuCH) and Swiss health national survey 2012 for data mining purposes. Each database has a demographic base data. An integrated Swiss database is built to later discover critical food consumption patterns linked with lifestyle diseases known to be strongly tied with food consumption. Design: Swiss nutrition national survey (menuCH) with approx. 2000 respondents from two different surveys, one by Phone and the other by questionnaire along with Swiss health national survey 2012 with 21500 respondents were pre-processed, cleaned and finally integrated to a unique relational database. Results: The result of this study is an integrated relational database from the Swiss nutritional and health databases.

***Keywords*—**Health informatics, data mining, nutritional and health databases, nutritional and chronical databases.

I. Introduction

LIFESTYLE diseases are diseases that increase in frequency as countries become more industrialized and people get more aged. Lifestyle diseases include obesity, hypertension, heart disease, type 2 diabetes, cancer, mental disorders and many others. They differ from infectious diseases, also called communicable diseases (CD) due to their due to their non-contagious, dispersive nature, often originating from nutritional behavior. Lifestyle diseases are therefore among the so-called non-communicable (NC) diseases. According to World Health Organization (WHO), the growing epidemic of chronic diseases afflicting both developed and developing countries are related to dietary and lifestyle changes. The rapidly increasing burden of chronic diseases is a key determinant of global public health. Already 79% of deaths attributable to chronic diseases are occurring in developing countries [1], [2].

In 2017, 11 million (95% uncertainty interval) deaths and 255 million DALYs were attributable to dietary risk factors. [3]-[5].

Study of nutritional patterns instead of individual food consumption has shown that, overweight and obesity, which is generally evaluated by various anthropometric measures including Body Mass Index and waist circumference, is increasing. According to [6] in 2017, 1.97 billion adults and over 378 million children and adolescents were categorized as overweight and obese over the globe. Moreover it has been assessed that greater body fatness can be cause of different cancers such as pancreas, oesophagus, liver, colorectum, breast and kidney [6]. Another study shows that nearly half of all US adults, i.e. 1.17 million individuals, have one or more preventable chronic diseases, which are related to poor quality eating patterns and physical inactivity. These chronic diseases are mostly cardiovascular disease, high blood pressure, type 2 diabetes, some sort of cancers, and poor bone health [7].

Schultze et al. [8] discuss current knowledge on the associations between dietary patterns and cancer, coronary heart disease, stroke, and type 2 diabetes, focusing on areas of uncertainty and future research directions. They report that nutrition support and dietary interventions are “key factors” to treat chronic conditions, although with some challenging aspects should be considered to improve dietary adherence.

Using data mining techniques has been proposed by various researchers. Several papers describe the use of pattern recognition and data mining to extract nutritional patterns. Hearty et al. [9] propose a coding system at the meal level that might be analyzed by using data mining techniques. These researchers used data from an existing conducted survey. Khan et al. [10] introduced a framework for mining market basket data to generate nutritional patterns (NPs) and a method for analyzing generated NPs using Fuzzy Association Rule Mining. The database used by Khan et al. was a synthetic grocery basket database from IBM Almaden [11]. Manikonda et al. [12] focused on an application of mining questionnaires of such kind to determine the current knowledge of participants and how this knowledge improved after the training session. Katsaras et al. carried out a study described in [13] using a nationwide survey of consumer preferences. Harris et al. [14] reported a study that aimed at quantifying food expenditures by age groups and contrast elderly expenditure patterns with other age groups, test for significant differences between elderly food-expenditures and younger age groups, and test for differences in food expenditures between two elderly age groups (age 65-74 versus age 75 and over).

To gain understanding about the impact of using data mining techniques for the analysis of lifestyle diseases that can be influenced by nutrition, we have conducted a preliminary study using a big database gained from a grocery store chain over a certain period. To show the proof of our concept, a publicly on-line available grocery store dataset [15] served as

Ilona Mewes and Farshideh Einsele* are with the Section of Business Information, Bern University of Applied Sciences, Switzerland (*e-mail: eef1@bfh.ch).

Helena Jenzer is with the Hospital of Psychiatry, University of Zurich, Switzerland.

our data source. This consecutive research study reports of using a database gained from a nationwide Swiss survey about nutritional habits linked with a Swiss nationwide health database.

Fig. 1 Swiss Geography

II. Selected Datasets

A. Swiss Nutrition Database menuCH

The menuCH National Nutrition Survey [16] is the first to provide representative data on the food consumption and eating habits of the population living in Switzerland. National Nutrition Survey (menuCH) diet and exercise have a direct impact on health and quality of life.

From January 2014 to February 2015, around 2000 people from the Swiss resident population were interviewed. Men and women between the ages of 18 and 75 provided information about their food consumption and about their cooking, eating and exercise along with some demographical behavior.

The survey was conducted as a questionnaire in the first stage and orally by phone in the second stage. Three tables resulted from the survey:

- The table with the data from the questionnaire provides information on eating and drinking and cooking behavior, as well as intake of additives and salts, avoided foods and reasons for avoiding food. Additionally, the survey provides basic knowledge of healthy eating, activity patterns, body measurements, weight satisfaction, diet behavior, social structure of the interviewed persons
- The table with the data from the oral survey provides information on the interview and the interview context; age and body information; food consumed (preparation, category, nutritional values, amount and time of taking the food).
- The third table contains data on the demographic classification of the respondents: Telephone number, year of birth, age group, gender, relationship status, nationality, country of birth, household size, residence in the major Swiss regions.

B. Lifestyle-Diseases Database

Swiss Ministry of Health gathers every 5 years data from around 21500 Swiss citizens asking them from a dual approach of a questionnaire and a detailed phone interview categorized questions about their health issues [17]. For the sake of simplicity, only the data from the phone survey were retrieved. The issues with the highest link and priority to the health and nutrition were extracted and reduced to a table with 9 subject areas: Alcohol consumption, age problems, disability, cholesterol, chronic diseases, diabetes, drug use, nutrition, health status.

C. Demographic Database

As mentioned above, two demographic tables from the above databases were obtained. The first database from the database menuCH included appx. 2000 individuals and the second one 21500 individuals. Therefore, a third table as the profile table including attributes such as gender, age group, household size, marital status and language was built and linked to the two existing demographic tables to this table.

III. Selection, Cleaning and Integration of the Databases

A. Selection

The relevant data were identified and selected from the tables. The positive criteria were freedom from redundancy, completeness, relevance (for the question). In the database of

World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:15, No:1, 2021

nutritional data, 60 table columns were selected from the oral survey data, and 125 table columns were selected from the questionnaire data. In the database of the Swiss health survey, only the data on the blood pressure issue were not considered, since no descriptions were available for the indices and the data could therefore not be decoded.

B. Cleaning

1. Cleaning menuCH Database

The menuCH database was largely represented by codes. menCH database included occasionally the same content, which was mapped with two different codes. This inconsistency was corrected by defining one coding type and overwriting the second coding type with the one specified. If there were no data cells, it was checked whether the data record was still usable. If not (for example the absence of the Person-ID), the entire data record was deleted.

2. Cleaning Health Database

The data were broken down into separate lifestyle themes and divided into nine tables. Hence the tables include data about alcohol consumption, cholesterol, diabetes II, drug consumption, age problems, disability, nutrition, chronic diseases, health status. Redundant data in the tables have been removed as well as missing data.

C. Transformation

1. menuCH Database

The setup of the menuCH database was done in three stages. First, a database was created for the database of the questionnaire. Another database was then created for the database of the oral survey. A personal profile was created with the third database (data on demographic characteristics). The personal profile has the characteristics age group, gender, household size, marital status, language. This personal profile connects the other two menuCH databases**.** Finally, a relational database scheme was designed, and menuCH database was implemented into mySQL.

2. Health Database

The database for health data could be created in one step. A person profile was created according to the same demographical characteristics as in the menuCH database (age group, gender, household size, marital status, language). All data on the health-related fields were linked by the person profile. Finally, a relational database scheme was designed, and health database was implemented into mySQL.

3. Linking Table for the Integration of Databases

For the integration of the nutrition and health databases, a third person profile table had to be created, which connects the person profile tables of the nutrition database and the health database. Six attributes were selected which were available in both databases for the personal description:

a) Gender (m/f)
b) age group (15-29/30-39/40-49/50-64/65+)
c) Household size (1/2/3/4/5/6+)
d) Marital status (single/married or registered/widowed/ divorced/other)
e) Language (de/fr/it)

The selected attributes and their categories resulted in 720 different categories of people. The PersonIDs in the menuCH database and the PersonIDs in the Health database were each assigned to a person category in the PersonProfil table.

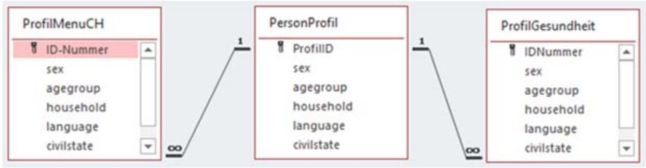

Fig. 2 Integrated database design via linking table

IV. Integration

The integrated database consists of 43 tables. The database is structured as follows:

- 14 tables contain the data from the menuCH «Questionnaire» database. 7 tables contain the data from the menuCH «Nutritional values» database. 18 tables contain the data from the "Health" database. The tables "Alcohol, age problems, disabilities, health status, cholesterol, chronic diseases, diabetes, drug consumption, nutrition" and the associated intermediate tables create a connection to the table "Profile health", the personal profile of the database health.
- Central to the integrated database is the three person profiles of the data base nutrition the data base health and the person profile which connects the two person profiles, as shown in Fig. 2. Fig. 3 illustrates the integrated menuCH and health database from Switzerland.

V. Conclusion and Future Work

The present study aimed at linking data sources of a nutritional database to demographical and health statistics to address the influence of food consumption patterns on lifestyle diseases such as obesity, hypertension, cardiovascular diseases, cancer, type 2 diabetes and mental disorder. According to WHO [15], "lifestyle diseases are among the main causes of premature death and disability in industrialized countries and in most developing countries. Developing countries are increasingly at risk, as are the poorer populations in industrialized countries".

For our future work, we intend to use data mining techniques to discover patterns. Our cooperation with the health institutes in Switzerland and in the European countries will be essential to receive accurate demographical and health data which should help us derive interesting and groundbreaking hidden patterns. Our goal is to find valid rules in order to be able to predict and prevent lifestyle diseases by detecting critical food consumption patterns. Data mining is an enormously mighty technique that allows us to help reach our goal without the common limitations of the previous research efforts, which used the classical statistical hypothesis-bound methods.

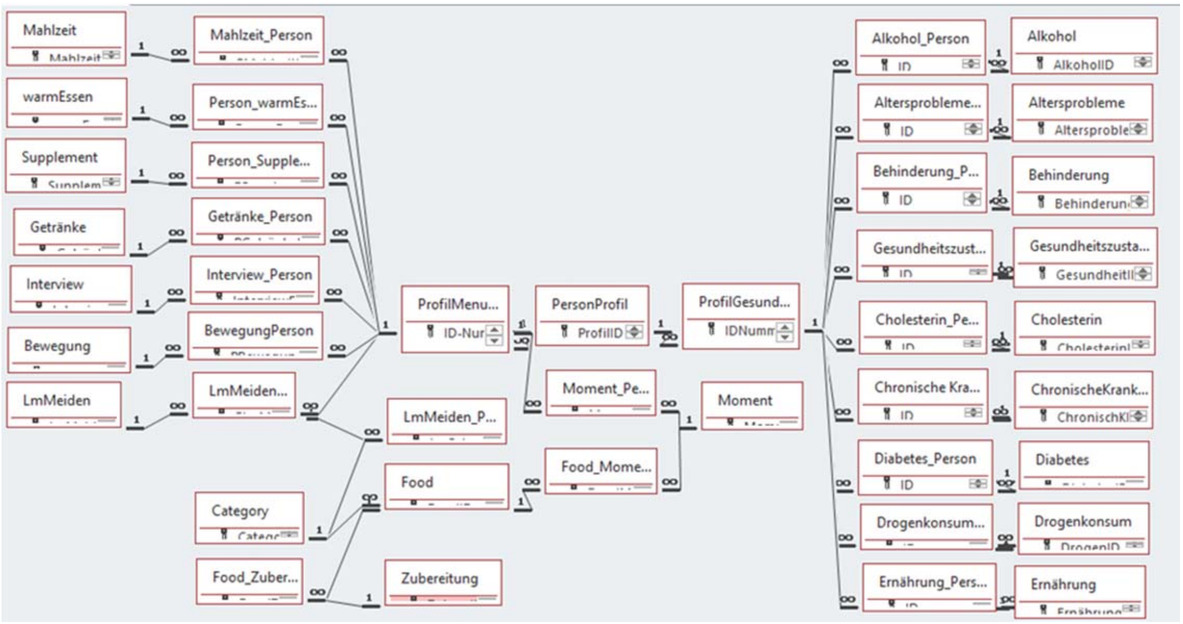

Fig. 2 Integrated Relational Database scheme

References

[1] Diet, Nutrition and the Prevention of Chronic Diseases, Report of a Joint WHO/FAO Expert Consultation, World Health Organization Geneva 2003.

[2] Diet, physical activity and health. Geneva, World Health Organization, 2002 (documents A55/16 and A55/16 Corr.1).

[3] Lim SS Vos T Flaxman AD et al., A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012; 380: 2224-2260

[4] Forouzanfar MH Alexander L et al., GBD 2013 Risk Factors Collaborators, Global, regional, and national comparative risk assessment of 79 behavioral, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 386: 2287-2323

[5] Global, regional, and national comparative risk assessment of 79 behavioral, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015.Lancet. 2016; 388: 1659-1724

[6] World Cancer Research Fund International, Recommendations and public health and policy implications 2018

[7] https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines/guidelines/introduction/nutrition-and-health-are-closely-related/

[8] M. B. Schulze, M. A. Martínez-González, T. T Fung, A. H. Lichtenstein, F Nita G Forouhi, Food based dietary patterns and chronic disease prevention, BMJ 2018; 361 doi: https://doi.org/10.1136/bmj.k2396 (Published 13 June 2018)

[9] I. P. Hearty and M. J. Gibney, Analysis of meal patterns with the use of supervised data mining techniques artificial neural networks and decision trees, 2008;88:1632–42. American Society for Nutrition

[10] M. Sulaiman Khan, M. Muyeba, F. Coenen, On Extraction of Nutritional Patterns (NPS) Using Fuzzy Association Rule Mining, healthinf 2008

[11] Agrawal and R. Srikant, Quest Synthetic Data Generator. IBM Almaden Research Center, http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/data/assoc.gen.tar.Z

[12] L. Manikonda, R. Mall, V.Pudi and R. Rao, Mining Nutrition Survey Data,

[13] J.D. Kinsey, P. Wolfson, N. Katsaras, B. Senauer, Data mining, A segmentation analysis of US grocery shoppers, Working paper (University of Minnesota. Retail Food Industry Center), 01-01

[14] J. Harris and N. Blisard, Food-consumption patterns among elderly age groups, Journal of Food Distribution Research, 2002

[15] https://sites.google.com/site/dataminingcourse2009/spring2016/annoucement2016/ assignment3 /D11-02.ZIP, date: 7/17/2020

[16] https://www.blv.admin.ch/blv/de/home/lebensmittel-und-ernaehrung/ernaehrung/ menuch.html, date: 7/17/2020

[17] https://www.bag.admin.ch/bag/de/home/zahlen-und-statistiken.html, date: 7/17/2020