Technological Applications for the Automation of Food Questionnaires in Medical Studies: a state-of-art-review and future prospective

Zoe Falomir\textsuperscript{a,b}, Francisco Madueño\textsuperscript{a,b}, María Arregui\textsuperscript{c,a}, Dolores Corella\textsuperscript{d,e}, Óscar Coltell\textsuperscript{a,b}

\textsuperscript{a}Department of Languages and Computer Systems, University Jaume I
\textsuperscript{b}Cooperative Research Thematic Network in Computational Biomedicine (RETIC COMBIOMED), Carlos III Health Institute
\textsuperscript{c}DIfE, German Institute of Human Nutrition Potsdam-Rehbrücke
\textsuperscript{d}Department of Preventive Medicine and Public Health, Faculty of Medicine, Universitat de València
\textsuperscript{e}CIBER Physiopathology of Obesity and Nutrition.

Abstract

Objective: In this paper we aim to provide a general view of reported software tools and Web-based applications developed for automating the extraction of data from the most widely used dietary interviews in the scope of nutritional research: Food Frequency Questionnaires (FFQs) and 24h dietary Recalls (24HDRs).

Methods: A comprehensive search of electronic databases was carried out for publications between January 1990 to December 2011. Findings from the search were commented by a group of experts in nutrition and computer sciences.

Results: The papers obtained were classified from the computational science point of view into two study types: (1) Computerized Questionnaires and (2) Web-based or Internet-based Questionnaires. Those works were classified from a nutritional point of view into: (i) Food Frequency Questionnaires (FFQs); (ii) 24h Dietary Recalls (24HDRs) and combinations of them. And a last classification was made regarding the type of administration to participants: (a) interviewer-administered or used by the researcher for interviewing the participants, and (b) self-administered or completed by the participants themselves. Then, a discussion is given regarding the classification
made and the works reported. Finally, some works that apply innovative technologies are outlined and the future trends for automating questionnaires in nutrition are identified and described as conclusions.

**Keywords:** Food Frequency Questionnaire (FFQ), 24h Dietary Recall (24HDR), Food Composition Table (FCT), Review, State-of-art, Epidemiological studies, Engineering Technologies, World Wide Web (WWW)

---

1. **Introduction**

In the study of epidemiological or complex multifactorial diseases is essential to take into account the gen*diet interaction because of the importance of nutrition in human beings and because dietary factors are linked to 4 of the 10 leading noncommunicable causes of death: cardiovascular disease, some cancers, stroke, and type 2 diabetes [1]. The assessment of a person’s diet is a difficult and tedious task which should consist of an analysis of their daily intake throughout a year. However, in epidemiological studies of diet-disease association, this assessment is not feasible and, in practice, only a portion of the intake information is evaluated and then the habitual participants’ intake is extrapolated. In order to obtain enough statistical power to avoid possible measurement errors and changes in diet, it is necessary to obtain repeated measures of dietary information from a large number of participants over time. For extracting information regarding the diet of the participants, nutritionist use Food Frequency Questionnaires (FFQ) or Food Propensity Questionnaires (FPQ), 24 hour dietary recalls (24HDRs), dietary records or dietary histories. These surveys collect consumed food or dishes, which can then be transformed into energy and nutrient intake using food composition tables.

Food Frequency Questionnaires (FFQ) or Food Propensity Questionnaires (FPQ), and 24 Hour Dietary Recalls (24HDRs) are the most widely used tools to extract information regarding diet in the scope of epidemiological studies. Both techniques assume that participants have some regularity in their diet and are able to quantify it.

A food frequency questionnaire (FFQ) consists of a list of foods or food types that are combined with a set of choices of frequency estimates. They are used to establish
usual or habitual consumption of foods or nutrients. For example, FFQs present general questions such as ‘Do you eat olive oil?’ and the participants of the studies have to respond ‘yes’ or ‘no’, and if confirmed, the FFQs ask the frequency of consumption (i.e. ‘How often do you eat olive oil? Units per day, units per week, units during the last 30 days, etc.’). These FFQs are also referred as food propensity questionnaires (FPQ) when they do not collect information about the portion size. In those cases, they are usually combined with data from multiple 24HDRs to complete the dietary information. The FFQs that include portion-sizes of the food items and groups are also called as semiquantitative FFQs.

Many FFQs have been developed and used in a variety of ways, ranging from capturing usual intake among large, population-based samples [2] to tailoring the questionnaire to measure intake of a particular nutrient, food, or food group in small, specialized samples, such as: iron [3], omega-3 fatty acids [4], calcium [5], phytosterols [6], etc.

FFQs are widely used for estimating nutritional intake, particularly in epidemiology, because of their advantages in the ease of administration and translation into nutrients that enables their use in large population studies and also because they can cover seasonal variations and occasional consumption of foods in the participants’ intake. However, the work associated with automation of nutrients calculation is intensive and requires considerable computing and nutritional expertise [7].

Twenty-four-hour dietary recalls (24HDR) ask open-answer questions to the participants, such as ‘List all the beverages you drank and all the foods you ate yesterday between midnight and midnight’ (unstructured recall) or ‘List all the beverages you drank and all the foods you ate yesterday for breakfast/lunch/dinner/snacks’ or ‘What you ate when you woke up?’(meal based recall). 24HDRs are used to collect high-quality dietary data because: (i) they are based on short-term memory and are less likely to be biased by social desirability and (ii) they do not consist of a closed list of foods and provide quantitative information rather than consumption ranges, therefore they do not require adaptation to specific populations such as FFQs. A single 24HDR is not considered representative of an individual’s usual diet, therefore multiple 24HDRs are desired for many studies of nutrient intake. Moreover, as they require highly trained interviewers and collection of more than 1 day to assess usual intake, 24HDRs are not
considered economical or practical in research settings with large samples and FFQs are frequently used. If 24HDRs could be self-administered using computer technologies to substitute the interviewer, they could be more feasible for large-scale studies [8].

Estimating the validity of a food survey is difficult because there is no ideal method of reference or gold standard. In practice, relative validations of a questionnaire (i.e. FFQ) are performed with respect to another questionnaire (i.e. 24HDR) taking into account that the sources of error between the reference survey and that evaluated must be as independent as possible. By comparing both surveys correlation coefficients are obtained which indicate the validity of the instrument and the calibration coefficients to be applied for correcting executions in the future. In epidemiological studies, generally FFQs are validated with reference to 24HDRs and vice versa.

Traditionally, FFQs and 24HDRs have been administered in paper (some examples are the free-access Harvard paper FFQs ¹). However, as the information and communication technologies gained importance in the last years, a great effort has been done in applying engineering technologies for automatizing questionnaires involved in epidemiological and nutritional studies in order to save costs.

At the beginning, computer tools/programs were designed and developed for helping the researchers to administer questionnaires to the participants and to accelerate the extraction of the important data, from FFQs [9, 10, 11, 7, 4, 12] and from 24HDRs [13, 14, 15]. Moreover, other SW applications that automated self-administered FFQs [16, 17, 3, 18, 4, 5] and 24h dietary recalls [19, 20, 21] were developed. Those tools accelerated the data extraction and processing, but they needed a specific computer system for working.

In the recent years, as the World Wide Web has been widespread, dietary Web-based questionnaires have substituted computerized questionnaires for improving accessibility and for obtaining a multi-platform functionality. Some examples are online FFQs [22, 6, 23, 24, 25, 26, 27, 28, 29, 30], online 24 hour dietary recalls [8, 31, 32, 33, 34] and combinations of both FFQs and 24HDRs [35, 36, 37, 38].

---

¹Harvard paper FFQs: https://regepi.bwh.harvard.edu/health/nutrition.html
In literature there are numerous papers concerning the validation, repeatability of FFQs and 24HDRs in epidemiological studies. However, there are less papers dealing with the development of computer technologies and nutritional databases used to support these studies. Those engineering developments are very important for the evolution and acceleration of the extraction of results in epidemiological studies. Our contribution here is: (i) providing a broad state-of-art-review from a technological point of view of those works that have automated a questionnaire involved in a nutritional study, (ii) reporting some guidelines for effective further automation of questionnaires and (iii) giving some outlines of future prospectives.

A few similar reviews or states-of-the-art are found in the literature[39, 40, 1, 41]. A very brief review on technologies applied to FFQ was presented by García-Segovia et al.[39]. A classification of Web tools and other computer applications used in nutrigenomic research was done by Stumbo et al.[40]. They presented the most used tools in US and Europe (without a wide review of related works) providing a description from the point of view of a user or researcher in nutrition. Long et al. [1] presented a review of the evidence on the effectiveness of technology-based methods for dietary assessment by reporting six technology-based methods. Ngo et al.[41] give a short review of computerized nutrition questionnaires and outlines some innovative methods for automating questionnaires, such as smart cards, personal digital assistants (PDAs) and mobile phones. In this paper, a wider period of time is taken into consideration, a broader description is presented regarding computerized and Web-based FFQs and 24HDRs, a comparative discussion is given regarding the automation and administration of questionnaires and also innovative technologies are reported. Moreover, an engineering perspective of all the reviewed works is given and useful recommendations for automatizing questionnaires in nutrition are outlined.

The rest of the paper is organized as follows. Section 2 explains the methodology used for carrying out the state-of-art-review and presents a classification of all the works found in the literature. Section 3 describes the computerized systems appearing in the literature that implemented food frequency questionnaires (C-FFQs) and 24h dietary recalls (C-24HDRs). Section 4 explains the Web-based approaches that automatize FFQs (Web-FFQ) and 24HDRs (Web-24HDRs). In Section 5 a discussion
is presented and Section 6 outlines a future prospective regarding the automation of questionnaires in nutrition. Finally, in Section 7 conclusions are given.

2. Methodology

A review was undertaken of nutritional studies conducted or published since 1980 in which the use of automated (computerized or Web-based) nutrition questionnaires were described. In order to identify relevant studies that describe the design, development and/or use of food questionnaires, a comprehensive search procedure was developed. Electronic databases including *PubMed*, *Embase* and *Web of Science* were searched from 1980 to December 2011. Hand searches of published conference proceedings, key nutrition journals and reference lists of retrieved articles were also undertaken. Search terms used were based on the following titles/topics and keywords/abstracts searches:

- computer* food* questionnaire*, computer* 24h* diet*, computer* diet* questionnaire*,
- Web* food* questionnaire*, Web* diet questionnaire, Web* diet* questionnaire*, Web* 24h* questionnaire*, Web* 24h* recall*,
- Internet* food* questionnaire, Internet* diet* questionnaire, Internet* 24h* recall*, Internet* 24h* questionnaire, Internet* diet* recall*, Internet* diet* questionnaire*.

All the references were downloaded into a BibTeX³ database which facilitated handling of the publications obtained and MIKTeX⁴, an open-source LaTeX⁵ compiler, was used for organizing and generating the bibliography and the document.

For the purposes of the research, some experts in nutrition and computer science were asked and the papers obtained were classified from the computational science

---

²Note that the symbol * means any combination of following characters for the searcher.
³BibTeX: http://www.bibtex.org/
⁴MIKTeX: http://miktex.org/
⁵LaTeX: http://www.latex-project.org/
point of view into two study types: (1) Computerized Questionnaires and (2) Web-based or Internet-based Questionnaires. Then those works were classified from a nutritional point of view into: (i) Food Frequency Questionnaires (FFQs); (ii) 24h Dietary Recalls (24HDRs) and combinations of them. And finally, a last classification was made regarding the type of administration to participants: (a) interviewer-administered or used by the researcher for interviewing the participants, and (b) self-administered or completed by the participants themselves. Table 1 and Table 2 shows the results of this classification.

3. Computerized Questionnaires

The advantages that computerized questionnaires offered with respect to paper questionnaires caused their progressive introduction in epidemiological studies. Some of them are the following:

- the higher speed of data collection because answers are automatically stored on databases saving time of data entry;
- the higher quality of data extracted since there is an immediate and automatic control for incomplete and implausible data;
- the direct data transfer to the centre of study (avoiding costs of printing and postage and organizational constrains, such as manual checks for missing and implausible answers or transfer of data to an electronic format);
- the direct aid on portion size estimation and recognition of food through real color illustrations;
- the small extra cost to add a few thousand participants to the study once the system for handling the questionnaires is developed and established.

In the literature, computerized questionnaires related to nutritional studies can be found. Section 3.1 present works that used computerized food frequency questionnaires (C-FFQs), whereas Section 3.2 describes computerized 24h dietary recalls (C-24HDRs).
3.1. Computerized Food Frequency Questionnaires (C-FFQs)

Research studies that developed and applied computer programs for automating the data extraction from Food Frequency Questionnaires (C-FFQs) can be found in literature[9, 16, 17, 10, 3, 11, 18, 7, 4, 12, 5].

Among all of them, some interesting studies that applied computerized FFQs are the following:

i. RIBEPEIX[4] is a computer program designed to quantify the intake of omega-3 fatty acids and chemical contaminants from the participants’ frequency of consumption of fish and shellfish and their meal size;

ii. a computerized FFQ developed and validated by Heath et al.[3] for estimating the iron, vitamin C and calcium consumption of participants using the New Zealand Food Composition Database [42], and the phytate consumption using the Canadian food composition database [43] and whose results were exportable to the Statistical Package for Social Sciences or SPSS\(^6\) v. 6.1.1;

iii. the computerized FFQ developed and validated by Vandelanotte et al.[18] that estimates the fat intake in Belgium using the Belgian [44] and the Dutch [45] food composition tables and whose results were also exported to SPSS v. 11 for analysis.

Although all these studies applied computerized FFQs, there are not technological descriptions of the design and development of those programs available.

From the point of view of engineering technologies, the most recent and relevant studies that applied computerized FFQs are the following:

1. CAFE[7]: The FFQ developed for the European Prospective Investigation into Cancer Study in the United Kingdom (EPIC-Norfolk and EPIC-Oxford) was automated for obtaining a nutritional analysis and for identifying extreme nutrient values using an exclusively designed computer program named CAFE (Compositional

\(^6\)SPSS Statistics: http://www-01.ibm.com/software/analytics/spss/
Analyses from Frequency Estimates). CAFE was written in SAS\textsuperscript{7} linked to an Oracle\textsuperscript{8} relational database which allowed predefined entries (numeric codes for frequency responses) and free text entries for registering new food items and nutrients. The EPIC nutrient database was built using the food composition tables by McCance and Widdowson and its supplements [46, 47].

2. NUTRISOL\cite{12}: a nutritional freeware program for analyzing dietary food intakes and translating them to nutrients using the Spanish CSIC food composition tables \cite{48} and domestic commonly used measures. It was developed using Visual Basic 6.0 and ran under Windows OS. The results produced are exportable to statistical programs.

3.2. Computerized 24h Dietary Recalls (C-24HDRs)

In the literature, few research studies that computerized 24h dietary recalls appear\cite{13, 19, 14, 15, 20, 21}:

1. EPIC-SOFT\cite{13}: a computerized questionnaire developed to obtain standardized 24HDRs between the nine countries involved in the European Prospective Investigation into Cancer and Nutrition (EPIC). This program was written in Clipper, it ran on MS-DOS v.3 and it was PC IBM-compatible needing 2 Mb of RAM and 5 Mb of hard disk space. The nutrient databases implemented in the system were not standardized, and temporary food composition tables, derived from national values, were used to calculate energy and macronutrients.

2. the Food Intake Recording Software System or FIRSS\textsuperscript{t}\cite{19}: a software program designed for use with fourth-grade children that uses interactive multimedia to facilitate a child’s self-report of diet by simulating a multiple pass 24HDR. FIRSS\textsuperscript{t} included a computerized tutorial about how to use the program and it organized foods within a group hierarchy in which commonly consumed foods such as pizza were also included. FIRSS\textsuperscript{t} was programmed in Director 6.5, Fox Pro was used to

\textsuperscript{7}SAS SW v.8, SAS Institute Inc., Cary, NC, USA
\textsuperscript{8}Oracle Corporation, Redwood Shores, CA, USA
create a database, and data were downloaded into MS Access. The Continuing Survey of Food Intakes by Individuals food coding system was used to identify foods and identify amount of foods in food groups for mixed dishes.

3. Raper et al.[14] gave an overview of the Dietary Intake Data System by the US Department of Agriculture (USDA) which is composed of:

- the Food and Nutrient Database for Dietary Studies (FNDDS) which includes food descriptions, food portions and their weights, and their corresponding nutrients. It is reviewed periodically by dietary coders who make corrections if needed and assign codes to any unmatched responses.

- three separated computer systems: (1) the Automated Multiple Pass Method (AMPM) for collecting food intakes using a 24HDR programmed using a Microsoft Access database and Blaise programming language, (2) the Post-Interview Processing System (PIPS) for reformatting data and assigning food codes, and (3) the Survey Net system for final coding and editing, quality review, and nutrient analysis.

USDA is the leader in 24HDR methodology, having used 24HDRs as the primary dietary assessment method in American surveys of food consumption since 1965. Since 24HDRs are not considered representative of an individual’s usual diet, the USDA created an Automated Multiple Pass Method (AMPM) which was used by [15] for assessing nutrient intake and adequacy from the dietary interview component of the National Health and Nutrition Examination Survey (NHANES): What We Eat In America (WWEIA).

4. The self-administered computer dietary assessment program named ‘Young Adolescents Nutrition Assessment on Computer (YANA-C)’[20] which implemented and evaluated a single 24HDR structured in six meal occasions including 18 food groups and a 19th group for the items not listed in the menu. It also included photographs for portion size estimation which changed every time a participant pushed the ‘more’ or ‘less’ button and remaining information such as ‘Don’t forget mayonnaise if you ate French fries...’. YANA-C was developed using Microsoft Visual
Basic v. 6.0. The total energy and nutrient intakes were computed using the Unilever Becel Nutrient Calculation Program v. 5.03 and the Belgian [44] and the Dutch [45] Food Composition Tables (FCTs). Moreover, the YANA-C was also applied to the Healthy Lifestyle in Europe by Nutrition in Adolescence or the HELENA study [49, 50].

5. Toobert et al.[21] developed an interactive CD-ROM program to estimate fruit, vegetable and fat intake from 24HDRs with the objective of increasing portion size estimation accuracy.

4. Web-based or Internet-based Questionnaires

In the last few years, a relevant number of research works in nutrition that use Web-based or Internet-based questionnaires appeared in literature. The significance of this questionnaires increased with the broadcasting of information and communication technologies. Web-based questionnaires provide all the advantages of computerized questionnaires and others such as:

- the higher compliance, such as more flexibility of completion at any time and location, given personalized feed-back and interactive help features (self-administered questionnaires);

- the ability to communicate with geographically dispersed research groups (interviewer-administered questionnaires) or populations (self-administered questionnaires), potentially internationalizing research, and groups often difficult to sample.

Those Web-based questionnaires related to nutritional studies that are most popular in the literature can be classified as Web-based food frequency questionnaires (Web-FFQs) described in Section 4.1, and Web-based 24h dietary recalls (Web-24HDRs) described in Section 4.2.

---

9Unilever Becel Nutrient Calculation Program v. 5.03, Hartog Union and Van den Bergh, the Netherlands, 1995.
4.1. Web-based Food Frequency Questionnaires (Web-FFQs)

The significance of the research works that use Web-based food frequency questionnaires has increased in the last few years [6, 23, 24, 26, 29, 30] and the validity and reproducibility of Web-based FFQs has also been evaluated [22, 25, 27, 28].

Among all of them, some interesting Web-based FFQs are the following:

i. RIBEFOOD[24] is an extension of the computerized program RIBEPEIX by [4] for determining the dietary intake of a number of chemical contaminants (i.e., metals, dioxins and furans, PCBs, polycyclic aromatic hydrocarbons, etc.);

ii. Matthys et al.[22] implemented and validated a Web-based FFQ for analysing adolescents’ food habits and proved its validity and reproducibility. The security of the system was achieved by providing a login name and a session-specific password to the participants. This FFQ contained questions on the average consumption of 69 food items during the past month and the food items listed were based on the classification system described in the Flemish Food Guide, commonly referred as the Food Triangle [51]. The energy intake was calculated using the Belgian [44] and the Dutch [45] food composition tables [52].

iii. Probst et al.[26] developed a Web-based FFQ which automated the sequence of questioning that a dietitian usually takes with a client when conducting a diet history interview and they also video-recorded and analyzed the participants’ behaviors related to the type of foods appearing in the questionnaires;

iv. Vereecken et al.[29] developed an online FFQ for the Healthy Lifestyle in Europe by Nutrition in Adolescence or HELENA Study [49], the reproducibility of this FFQ was analyzed by comparison with the YANA-C computerized 24HDR [20] and it has been recently validated by Maes et al.[53].

From the point of view of engineering technologies, the most relevant developments of Web-based FFQs are:

1. the Diet History Questionnaire (DHQ) [54] funded by the American National Cancer Institute (NCI) with their DHQ Nutrient Database and data analysis used in cancer research.
2. A Web-based pictorial diet history questionnaire (PDHQ) [25] was developed by adding pictures to the National Cancer Institute’s DHQ in order to represent portion sizes. They also recruited participants from advertisements placed at *The Washington Post* website\(^{10}\) and the study was approved by the PICS Institutional Review Board Office in Reston, VA, USA;

3. The Web-based assessment tools in Epidemiology\(^ {11}\) by the German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE) [55]: (1) EPIC Postdam - Food Frequency Questionnaire (EPIC-Postdam-FFQ); (2) European Food Propensity Questionnaire (EFPQ); (3) German Food Propensity Questionnaire (GFPQ); (4) European-Hellenic Food Propensity Questionnaire (HFPQ). A lot of studies by DIfE applied these Web-based FFQs, an example is the EFPQ recently used by Illner et al. [30] in combination with three telephone-administered 24HDRs for comparing participation rates, acceptance of the instruments and data provided by Web analysis, etc.

4. The VioFFQ by Viocare enterprise [56]: a Web-based system to self-administer a FFQ that collects data on dietary behavior and food patterns, estimates nutrient intake, and delivers a dietary change report. The dietary analysis is done using the food and nutrient information from the Nutrition Coordinating Center (NCC) Food and Nutrient Database, which is developed and maintained by the NCC, located at the University of Minnesota Division of Epidemiology and Community Health in Minneapolis. The VioFFQ was used in the epidemiological studies carried out by [57].

5. The Block FFQ [58] commercialized by NutritionQuest enterprise [59]: a turn-key system, which provides an online structure for integrated data collection, nutrient and physical activity analysis, and data management. Questionnaires can be interviewer-administered or self-administered, the data stored, and nutrition or physical activity estimates calculated. Block FFQ accessed online and self-administered was used by Anderson-Bill et al. [60] for estimating the intake of fat, fiber, fruits and

\(^{10}\)[*The Washington Post*: http://www.washingtonpost.com/]
\(^{11}\)[DIfE Internet based assessment tools in Epidemiology: https://efbo.dife.de/portal/en]
vegetables of the participants of a study for examining the behavioral characteristics of Web-health users.

6. an online dietary questionnaire was developed and validated by Apovian et al.[27] for capturing food servings in the Dietary Approaches to Stop Hypertension or DASH diet recommended by the US Department of Agriculture (USDA). It was implemented on HTML using check boxes and entry fields for self-administration and it was placed on a secure Web server at Boston University’s Data Coordinating Center. Informal testing for ease of use and understandability was done on a subset of ten people.

7. the on-line semi-quantitative FFQ to evaluate calcium and iron intake by Galante and Colli[23]. This FFQ was located at Fohla journal website[12] for selecting participants. The contents of the study (questionnaires, forms, informed and written terms of consent) were stored in a database with a copy on a CD-ROM. To ensure the confidentiality of the information and the privacy of the individuals, access to the communication tool was made available through an individual password, and data transmission was performed through personal e-mails. The Virtual Nutri SW[13] was used for the evaluation of the data from the dietary records. Histograms were obtained to show the distribution of nutrient intake and the normality of the distribution curves was analyzed. The project was approved by the Ethics and Research Committee of the University of São Paulo.

8. FITUVEROLES[6]: a Webservice for calculating phytosterol intake in Mediterranean population [61] using an on-line FFQ developed inside the Spanish cooperative research thematic network on computational medicine *Combiomed*[14] made up by physicians, computer science engineers and other researchers in biomedicine. FITUVEROLES Webservice estimated the intake of phytosterols of the participants from previously validated FFQs that used three food composition tables (FCTs) for

---


13 Virtual Nutri SW by ST. Philipp, SC Szarfarc and AR. Latterza at Universidade de São Paulo, 1996.

14 *Combiomed*: http://combiomed.isciii.es/English Website/Index.html
comparing and providing the robustness of the estimation: the Finnish Food Composition Database[62], the National Nutrient Database for Standard Reference by USDA [63] and the FCT by Jiménez-Escrig[64]. For obtaining the energetic value of each food component independently of the name given in each composition table an intermediate ontology, NutriOntología(NGO), was developed for carrying out the corresponding alignments between food names [65, 66]. NGO was developed using the Ontology Web Language (OWL) and Protégé15 as front-end. The properties of the Webservice were: (i) dynamic generation of FFQs from the food and biochemical parameters of interest selected by the administrator; (ii) collecting, maintaining and organizing information derived from biochemical analysis made to the participants; (iii) reporting results of biochemical analysis and (iv) exporting the results obtained in a format suitable for SPSS.

4.2. Web-based 24h Dietary Recalls (Web-24HDR)

New approaches that automated 24h dietary recalls (24HDRs) using online technologies have appeared in literature [8, 31, 32, 33, 34]:

1. the Oxford WebQ[34]: a Web-based method for assessment of 24HDRs, which is self-administered and estimates nutrients automatically, providing a low-cost method for measuring dietary intake in large-scale studies.

2. ASA24 [8, 67] a Web-based Automated Self-Administered 24HDR developed by the American National Cancer Institute (NCI) which was inspired by the USDA’s Automated Multipass Method or AMPM[14, 15] and that incorporated and extended the Food Nutrient Database for Dietary Studies or FNDDS[14] and the food portion pictures of Food Intake REcording SW System or FIRSS[19].

3. the Synchronized Nutrition and Activity Program (SNAPTM)[31], a Web-based software designed and evaluated for obtaining self-reported 24h dietary recalls and physical activity questionnaires in school children. SNAPTM was written using PHP v. 4.0.1, MySQL v. 3.22 and JavaScript v. 1.3 to ensure data transfer and storage

15Protégé: http://protege.stanford.edu/
and its interface is intuitive and user-friendly for children to use. The dietary intake was measured using counts for 21 food groups and the list of commonly consumed foods and drinks was developed from a combination of findings from the National Diet and Nutrition Survey [68]. A free-text option box was also included to capture any unlisted food or drink. The data analysis was carried out using STATA\textsuperscript{16} v.8. software and the accuracy of the computer tool (systematic and proportional bias) was provided by a Passing-Bablok method comparison technique using the Analyze-it\textsuperscript{17} software.

4. the Web-Survey of Physical Activity and Nutrition (Web-SPAN) [33] presented and evaluated the appropriateness of a Web-based method for assessing dietary intake and physical activity in young students. The Web-based 24HDR data was transferred to ESHA Food Processor\textsuperscript{69} and the Canadian Nutrient File database\textsuperscript{70} was used for calculating the macronutrient and micronutrient of the intakes. Statistical analyses were performed using SPSS v. 15.

5. DietDay\textsuperscript{32, 71}: a fully automated, self-administered, Web-based 24HDR inspired on multi-passes similar to the AMPM by USDA and it is built by 9,349 foods and over 7,000 food images in 61 modules. Portion sizes are quantified by household measures using images of different amounts of food on a standard plate, glass, or bowl. Food preparation methods are also assessed, as well as condiments and additions. In DietDay usual consumption is asked by time of day (midnight to 11 a.m., 11 a.m. to 5 p.m., and 5 p.m. to midnight). Nutrient values in the program are based on USDA values and expanded to include mixed dishes and product labeling information.

4.3. Systems Integrating Both: Web-FFQs and Web-24HDRs

In literature, approaches that integrate Web-based FFQs and Web-based 24h dietary recalls are those by [35, 37, 36, 38].

\textsuperscript{16}Stata Corp. USA: http://www.stata.com/
\textsuperscript{17}Analyze-it SW, UK: http://www.analyse-it.com/
Hanning et al.[35] developed a Web-based Food Behaviour Questionnaire (FBQ), which included a 24HDR and a FFQ which analyzed nutrients using ESHA Food Processor[69] and the Canadian Nutrient File database[70]. This FBQ was used by [72] for assessing dietary energy as a function of gender and weight status among Ontario and Alberta adolescents. And later it was validated by [73] with students with grades six to eight obtaining a positively correlated energy and nutrient intakes between FFQ and dietitian interviews.

myDIDeA [38]: a Web-based dietary intervention for type 2 diabetes patients which included a semi-food frequency questionnaire (SFFQ) for recording the participants’ dietary intake and two days 24HDR for analyzing their nutrient intake using the Axxya Systems Nutritionist Pro™ Diet Analysis

OBENUTIC[36, 37] Web service for automating FFQs and 24HDRs designed and developed for investigating the relation between the participants’ food intake and obesity and its related diseases (diabetes, hypertension, cancer, atherosclerosis, etc.) inside the Spanish cooperative research thematic network on computational medicine Combiomed. The properties of OBENUTIC Webservice were: (1) dynamic generation of FFQs and 24HDRs from the food and biochemical parameters of interest selected by the administrator; (2) collecting, maintaining and organizing information derived from biochemical analysis made to the participants; (3) reporting results of biochemical analysis and (4) exporting the results obtained in a format suitable for the SPSS. OBENUTIC Webservice estimated the energy intake of the participants using one of the following 4 food composition tables (FCTs) or combinations of them: the Spanish tables by Mataix[74] and CESNID[75], the Fineli or the Finnish Food Composition Database[62] and the USDA’s National Nutrient Database for Standard Reference[63]. For obtaining the energetic value of each food component independently of the name given in each FCT an intermediate ontology, NutriOntologia, was developed for carrying out the corresponding alignments between food names [65, 66]. Moreover, a light ontology of recipes named OntoReceta [76] based on common Spanish recipe

books [77, 78] was developed using Resource Description Framework\(^{19}\) (RDF) on a taxonomy of foods in order to decompose usual dishes (i.e. Paella) in kinds of foods and their quantities and the dressings used in their preparation. OBENUTIC database engine was MySQL using XML as the interchange language among databases. For building interfaces and programming modules in the Web architecture, the following Web technologies were applied: (i) PHP, JavaScript and XHTML as Web programming languages; (ii) the RDF Application Programming Interface (API) for PHP (RAP) included in pOWL\(^{20}\) was used for the ontology definition; (iii) nuSOAP\(^{21}\) was applied for building Web services; (iv) the frameworks XAJAX\(^{22}\), Prototype\(^{23}\), Scriptaculous\(^{24}\) and ExtJS\(^{25}\) were applied for designing and developing asynchronous interfaces. For example, the framework Scriptaculous was used for allowing the food intakes in the 24HDRs to be organized by drag\&drop in an easy and intuitive way for the interviewers. The results obtained by the developed Web-FFQs and Web-24HDRs for the OBENUTIC study were validated by [79] comparing the results of an aleatory sample of on-line questionnaires with the results obtained in paper. Finally, users also evaluated this Web platform as useful, reliable and easy to use.

5. Discussion

After analysing all the works appeared in the literature related to the automation of questionnaires in nutrition, a discussion is presented comparing: (i) automated vs. paper questionnaires (Section 5.1); (ii) computerized vs. Web-based questionnaires (Section 5.2); and (iii) self-administered vs. interviewer-administered questionnaires (Section 5.3).

\(^{19}\)RDF: http://www.w3.org/RDF/
\(^{20}\)pOWL: http://powl.sourceforge.net/overview.php
\(^{21}\)nuSOAP: http://sourceforge.net/projects/nusoap/
\(^{22}\)XAJAX: http://www.xajax.net/
\(^{23}\)Prototype: http://www.prototypejs.org/
\(^{24}\)Scriptaculous: http://script.aculo.us/
\(^{25}\)ExtJS: http://www.sencha.com/products/extjs/
5.1. Automated vs. Paper Questionnaires

In a comparison between the use of a Web-based questionnaire with a similar printed questionnaire, [80] found that the willingness to answer the second part of the questionnaire was higher with a Web-based questionnaire than with the printed questionnaire, which suggests that the participants responding to the Web-based questionnaire found the process more appealing than those who responded to the mailed questionnaire. Therefore, they recommended the use of Web-based questionnaires if the Internet access is high for the population involved. Another study by [81] for comparing Web-based versus paper questionnaires in adolescents concluded that most of the participants required less time to fill in the Web-based questionnaire and they also felt less observed and more independent while completing it. [82] and [83] compared Web-based and paper versions of self-administered questionnaires used in the NutriNet-Santé Study concluding that the quality of information provided by the Web-based questionnaire was equal to, or better than, that of the paper version, with substantial logistic and cost advantages.

5.2. Computerized vs. Web-based Questionnaires

After revising all the works in the literature we note that the first trend was developing computerized questionnaires that were used by the researchers to administer questionnaires to the participants and helped them to store all the dietary information and to calculate the corresponding amount of energy from it.

Then, the trend was developing Web-based questionnaires that may be accessible around the world using Internet and storing all the information in a centralized Web-server. Because of the widespread of the World Wide Web (WWW), these new applications allowed the possibility of self-administration of questionnaires to a wide range of participants. And since the Semantic Web gained importance, developments of ontologies related to food intake, recipes and nutrition were developed. Pattern recognition algorithms were also improved allowing the introduction of free entrances in 24HDRs questionnaires because the contents of the answers were automatically interpreted.

Finally, the most important advantage of Web-based questionnaires is their portability because they can be used on standard PCs or on tablet computers or mobile
phones only using Internet connection.

5.3. Self-administered vs. Interviewer-administered Questionnaires

Regarding the type of administration to participants computerized and Web-based questionnaires can be classified into: (i) designed for self-administration and (ii) designed to be administered by interviewers to participants.

Interviewer-administered questionnaires avoid excluding from the studies segments of population without Internet access or ability to use computers. Moreover, as interviewer-administered questionnaires are made in front of the participants, an expert dietitian can be able to recognize nonverbal behavior that potentially may aid in minimizing social desirability bias. As [26] mentions, recognizing that a client is shifting in the chair, touching their face or moving their head around when asked about dessert, for example, may assist in determining the appropriate types of questions to be asked, how they are asked, and whether or not to ask additional probing questions in order to gain more detail.

On the other hand, as the information and communication technologies have been broadly extended in the last years, self-administered questionnaires are becoming more popular. A comparison by [84] between interviewed-administered 24HDRs and interactive Web-based self-administered 24HDRs concluded that the second ones permit considerable logistic simplification and cost saving and that they may be highly advantageous for large population-based surveys. The feasibility of self-administration (self report) of YANA-C was analyzed [49] by comparison with administration by a dietitian (interview) concluding that a good agreement between both administration modes was found. Moreover, Web-based self-administered questionnaires also can be completed at any time from any location with centralized monitoring of participant completion. However, from the point of view of information technologies, a self-administered Web-based questionnaire must be designed and developed more accurately because it is not completed by an expert dietitian or computer scientist. Therefore, some considerations have to be taken into account:

1. The participants must receive information regarding: (a) the Web address, and (b) a unique ID-number and a secure password for each participant in order to ensure
the confidentiality of the data provided to the system. This information can be
provided by e-mail for accelerating the process of data gathering.

(2) The interface must guide equally all the participants and minimize potential errors
derived from their varying degrees of knowledge. For that, help screens, a tutorial
or an animated agent who guides participants through the Webside can be used.

(3) In order to obtain trustworthy information regarding the food portion-size it is neces-
sary to include colour photographs showing different portion sizes per food items
or/dishes.

Moreover, Web-based self-administered questionnaires are not usually designed to
capture the social desirability of a particular food. However, they may incorporate
video recording of participants’ body movement and facial expression at the time of
food reporting. And further on, this video may be processed by intelligent systems
for identifying such desirability. An initial study by [26] described the differences in
the observed behaviors according to the type of foods selected by participants using a
prototype version of a Web-based video-recording dietary assessment.

Technologies

In the symposium by the European Nutrigenomics Organisation (NuGO) [85] the fol-
lowing innovative applications were presented:

1. the ACASI by [86, 87], an audio computer-assisted program for interviewing
participants so that they fill in a diet history questionnaire listening to the ques-
tions;

2. the IMM by [88], an Interactive Multimedia computerized recall that incorpo-
rates a touch screen and audio functions to help data input;

3. the Wellnavi by [89, 90], a Pocket-PC with digital camera and mobile phone for
direct electronic data transfer to a dietitian in the study centre;
4. the Technology Assisted Dietary Assessment (TADA) project by [91, 92, 93] used a mobile device with an embedded camera to estimate daily food and nutrient intake from digital images taken by the participants. Food recognition in digital images is achieved using segmentation methods combined with Gabor filters for texture comparison and statistical pattern classification techniques such as Support Vector Machines (SVMs). Food portion size is estimated in cm$^3$ and the Food and nutrient Database for Dietary Studies (FNDDS) is used for obtaining weight measures. Moreover, X-ray computerised microtomography (XMCT), 3D laser imaging, and other techniques are used to measure food density (grams/cm$^3$) to allow for conversion of portion estimates to a weight measure.

Other innovative applications can be found in literature [94, 95]. Image-Diet Day[94], an automated image-capture method to aid dietary recall consisting of a user-initiated camera-equipped mobile phone programmed to automatically capture and transmit images to a secure website in conjunction with computer-assisted and multipass 24HDRs. Participants used the device during eating periods on three independent days and then they used the captured images using ImageViewer Software while completing the 24HDR on the following day. Image processing filters successfully eliminated underexposed, overexposed and blurry images. Participants concluded that the images were helpful but that wearing the device around their neck was not easy and they had also problems such as limited battery life, self-consciousness about wearing the device in public and concerns about the field of view of the camera. Therefore, there is still work to be done to solve this problems and to met the challenge to manage the thousands of images generated. Lambert et al.[95] tested how useful are smart cards for assessing food intake. They provided smart cards to school children and when they paid for their meal, the foods on the tray were recorded at the cash desk and sent to a central computer. Moreover, the date and time of the meal was stored and also the data on the computer can be linked to a nutrient database. However, they observed that this method was useful for collecting information about participants’ food selection but not for their food intake because they sometimes exchanged trays or paid each other’s meals and the real
portion of the intake was not known and it had to be estimated.

7. Conclusions and Future Prospective

FFQs and 24HDRs dietary recalls have been automated in a variety of ways in different studies. Depending on the aim of the study and on the population analysed, a method for automating nutrition questionnaires would be more suitable than others.

In this state-of-the-art-review, nutritional studies conducted or published since 1980 in which the use of automated (computerized or Web-based) nutrition questionnaires were described. In order to identify relevant studies that describe the design, development and/or use of food questionnaires, a comprehensive search procedure was developed using the electronic databases PubMed, Embase and Web of Science and hand searches of published conference proceedings, key nutrition journals and reference lists of retrieved articles. The works obtained were classified from a nutritional point of view and regarding the type of administration to participants. Then, a discussion is given regarding the classification made and the works reported. And finally, some studies that apply innovative technologies are outlined too.

In conclusion, the identified future trends for administration of questionnaires from the point of view of computing engineering and nutrition are: (1) using mobile devices such as mobile telephones, PDA-like devices or tablets with a built-in camera, network connectivity, and a microprocessor for integrating image analysis, visualization tools, etc; (2) programming questionnaires as mobile phone applications which the participants may download from public markets on their personal mobiles and use for self-administration; (3) introducing multimedia functions in questionnaires automated for mobile devices such as: selection of foods by touching the screen, naming the food appearing in photographs using a handwriting-note application; speech synthesizers for reading the questionnaire, etc; (4) introducing computer vision technologies for recognition of foods (i.e. Gabor filters and Support Vector Machine algorithms, etc.), for calculating food portions and weight in the photos made by the participants’s mobile devices; (5) using computer vision methods for identifying the barcode of some foods difficult to decompose in components (i.e. precooked foods) in order to get the specifi-
cation of the exact ingredients used by the manufacturer; (6) using pattern recognition algorithms and language processing techniques for identifying the contents of free text entrances; (7) developing and using ontologies for assigning meaning (interpretable by Web agents) to the food, recipes and other nutrition information extracted from the questionnaires; (8) cameras (in portables or mobile phones) for video-recording and recognizing participants’ behaviors when completing questionnaires with the aim of minimizing social desirability bias.

Acknowledgments

This work has been partially supported by: (1) Spanish Ministry of Science and Innovation (MICINN) under project named GEWIMICS and referenced as SAF2009-12304/GEN; (2) the Cooperative Research Thematic Network (RETIC) on Computational Medicine (COMBIOMED) referenced as RD07/0067/0006 and funded by Carlos III Spanish Health Institute (ISCIII); (3) the Biomedical Research Centre Network (CIBER) named Fisiopatología de la Obesidad y Nutrición (CIBERobn) as an initiative of ISCIII; (4) Generalitat Valenciana under grant numbers BEST/2010/211, BEST/2011/261, and ACOMP/2011/145; and (5) Spanish National Centre of Cardiovascular Research (CNIC) under project number CNIC06.

References


Table 1: Classification of Computerized Questionnaires: (i) Food Frequency Questionnaires (C-FFQs); (ii) 24 hour Dietary Recalls (C-24HDRs); and both combined FFQs and 24HDRs.

<table>
<thead>
<tr>
<th></th>
<th>C-FFQs</th>
<th>C-24HDRs</th>
<th>C-FFQ + C-24HDRs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smith et al.[10]</td>
<td>AMPM[14, 15]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Geekie et al.[11]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAFE[7]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RIBEPEIX[4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NUTRISOL[12]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Administered</td>
<td>Engle et al.[16]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suitor&amp;Gardner[17]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heath et al.[3]</td>
<td></td>
<td>FIRSS[19]</td>
</tr>
<tr>
<td></td>
<td>Vandelanotte et al.[18]</td>
<td></td>
<td>YANA-C[20]</td>
</tr>
<tr>
<td></td>
<td>Domingo et al.[4]</td>
<td></td>
<td>Toobert et al.[21]</td>
</tr>
</tbody>
</table>
Table 2: Classification of Web-based or Internet-based Questionnaires: (i) Food Frequency Questionnaires (Web-FFQs); (ii) 24 hour Dietary Recalls (Web-24HDRs); and both combined Web-FFQs and Web-24HDRs.

<table>
<thead>
<tr>
<th>Interviewer-Administered</th>
<th>Web-FFQs</th>
<th>Web-24HDRs</th>
<th>Web-FFQ + Web-24HDRs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FITUVEROLES[6]</td>
<td></td>
<td></td>
<td>OBENUTIC[36, 37]</td>
</tr>
<tr>
<td>RIBEFOOD[24]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matthys et al.[22]</td>
<td></td>
<td>ASA24[8, 67]</td>
<td></td>
</tr>
<tr>
<td>Galante&amp;Colli[23]</td>
<td></td>
<td>SNAP[31]</td>
<td></td>
</tr>
<tr>
<td>Probst et al.[26]</td>
<td></td>
<td>DietDay[32]</td>
<td></td>
</tr>
<tr>
<td>Beasley et al.[25]</td>
<td></td>
<td>Web-SPAN[33]</td>
<td></td>
</tr>
<tr>
<td>Martí-Cid et al.[24]</td>
<td></td>
<td>Oxford WebQ[34]</td>
<td></td>
</tr>
<tr>
<td>FFQ-DASH[27]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vereecken et al.[29]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labonté et al.[28]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPIC-Postdam-FFQs[55]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VioFFQ[56]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BlockFFQ[59]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Administered</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBQ[35]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>myDIDeA[38]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>