Development of Web-based System for a Cohort Study
The Brazilian e-NutriHS

Luciana Dias Folchetti, Isis Tande da Silva, Bianca de Almeida Pititto and Sandra Roberta G Ferreira
Department of Nutrition, Faculty of Public Health, University of Sao Paulo, Av. Dr. Arnaldo, 715, Sao Paulo, Brazil

Keywords: Electronic Health Records, Health Information on the Web, Web-based System, Cohort Study, Nutrition.

Abstract: This study describes a web-based system developed to gather online information on health of college students and graduates in nutrition, the e-NutriHS. The Nutrititionist Health Study – NutriHS is planned to be a cohort study aiming to collect health-related data at a 3-year interval. The e-NutriHS consists of 6 questionnaires regarding demographic and socioeconomic data, dietary habits, physical activity, alcohol and tobacco uses, anti-fat attitudes and personal and family histories. Validated and internationally recognized lifestyle instruments were used. Our software and respective database are hosted in the School of Public Health server; the software is based on free programming languages. e-NutriHS data obtained from questionnaires can be transferred to excel format. An e-NutriHS prototype was created preceding online attachment. An improved version of website was released based on 20 volunteers' opinions. 503 users were already registered. Our initiative of building a website designed for collecting data for epidemiological studies, tailored to our local reality, is innovative under the perspective of the health informatics available in the developing world. Considering that web-based systems produce reliable data, are easy to use, less costly and less time-consuming, we conclude that our experience deserves to be shared, particularly with middle-income economy countries.

1 INTRODUCTION

In recent years, the use of the Internet by the populations, in particular for healthcare information provision, is markedly increasing. This represents a desirable scenario, but it is worrying that sometimes there is a lack of scientific evidence which could have harmful impact in individuals. In light of this growth, researchers of health sciences have invested in learning how to use the Web to spread high-quality evidence-based data. Such initiatives have developed their abilities to deal with electronic tools for data collection and report, recognizing benefits as well as pitfalls on using the Web. Since nutrition is a key element for health, this issue represents one of the most explored.

Research studies in epidemiology are often conducted in hundreds or thousands participants, followed for long periods of time. These studies generate a significant amount of data, requiring trained professionals with specialized skills for data collection and processing. Health surveys have been mostly based on paper questionnaires. Their quality depends on accuracy in collecting, sorting, coding, typing and checking of data to create a consistent database before analysis. These steps are expensive and time-consuming, prolong the study duration, but are essential for the adequate data analysis and results interpretation to achieve reliable conclusions. Until recent years, relevant epidemiological studies were based on paper-and-pencil surveys, particularly those concerning lifestyle data (Colditz and Hankinson 2005, Wilson et al. 2012, Hu et al. 1997).

Technological advances have allowed obtaining standardized and reliable data in a faster and less expensive way. The completion of data electronically can ensure almost immediate construction of database, reduce costs, errors and biases and minimize duplicate information. This has largely motivated the use of the Web, for instance in nutritional epidemiology. Researchers in nutrition commonly deal with analysis of dietary intake, usually collected by questionnaires and/or food recalls. Several softwares are available for data analysis, tailored to dietary habits of a given population. In contrast, a direct link between questionnaires completion and data processing for analysis is missing. Few softwares, specifically designed to play this role, are found in literature and none in developing countries. Web-based question-

D. Folchetti L., Tande da Silva I., de Almeida Pititto B. and R. G. Ferreira S..
Development of Web-based System for a Cohort Study - The Brazilian e-NutriHS.
DOI: 10.5220/0005277805740579
In *Proceedings of the International Conference on Health Informatics* (HEALTHINF-2015), pages 574-579
ISBN: 978-989-758-068-0

naires applied in the nutrition field would facilitate the "dialogue" of data collected with software to dietary reports, as well as to statistical procedures.

The development of web-based online self-administered systems, employing validated health-related surveys, sociodemographic status, dietary habits and personal and family medical history, with a link for online scheduling for face-to-face data collection is warranted. Such system to obtain health-related data would be particularly important in middle-income economy countries where resources for researches are limited. It is anticipated that these methods will become a more feasible way of implementing surveys, providing a number of advantages over traditional methods, including convenience for the participant, potentially large cost savings for the researcher, efficiency in data collection, higher data quality, a degree of perceived anonymity for the participant, and high response rates (Best et al. 2001, Hewson et al. 1996, Krantz et al. 1997, Schaefer and Dillman 1998, Schmidt 1997, Shettle and Mooney 1999).

In Brazil, initiatives for electronic collection of data for monitoring health aspects are rare. A national health survey, the Telephone-based Surveillance of Risk and Protective Factors for Chronic Diseases, under the coordination of the Ministry of Health and supported by the University of Sao Paulo and the Center for Disease Control, uses desktop software to obtain data of the Brazilian population but requires trained interviewers (Azevedo e Silva et al. 2011). To our best knowledge, self-administered online questionnaires have not been reported.

The purpose of the current study was to describe a web-based system developed to gather information on health of a specific subset of the Brazilian population, allowing participants to complete data online. Characteristics of the participants were established based on their potential skill to navigate on the Internet and on the expected high quality of data provided. The choice of nutrition college students and graduates fulfills these criteria.

2 LITERATURE REVIEW

2.1 Epidemiological Studies of Health Professionals

Monitoring of nutritional and lifestyle factors associated with changes in health conditions in epidemiological studies has contributed to identify modifiable risk factors, to deepen the understanding of the pathophysiology of diseases, especially non-communicable chronic diseases (NCD) and to propose interventions. In this sense, experiences involving health professionals have brought significant contributions to the knowledge about the role of behavioral factors on health outcomes of Americans (Colditz and Hankinson 2005, Wilson et al. 2012, Hu et al. 1997). Particularly, cohort studies due to their longitudinal design have provided the most relevant evidence.

Approaches to gather information in cohort studies have been mainly based on face-to-face and/or telephone interviews, using paper-and-pencil questionnaires. The Nurses' Health Study I and II and the Health Professionals' Follow-up Study including thousands of participants in North America, are remarkable considering their scientific contributions for identifying lifestyle risk factors for NCD (Colditz and Hankinson 2005, Wilson et al. 2012, Hu et al. 1997). Despite the use of paper-and-pencil questionnaires, those studies had the strength of involving professionals that guarantee a high quality of data collected. Other epidemiological studies with similar objectives were conducted in America and Europe, including smaller samples and non-restricted to health professionals (Turner et al. 2009, Mikkelsen et al. 2009).

In this century, researches using web-based questionnaires have gained significant popularity (Couper 2000). More frequently, they have been used in psychological studies and marketing research, but in the field of epidemiology, Internet tools were employed in less than 1% of the reported studies (van Gelder et al 2010). The Millennium Cohort Study (Smith et al. 2007), the Nurses and Midwives e-Cohort Study (Turner et al. 2009), and the Danish Web-based Pregnancy Planning Study (Mikkelsen et al. 2009) are examples of successful large cohort studies where de recruitment and follow-up occur over the Internet.

In developing countries, similar initiatives involving health professionals are rare and, considering the use of Internet as a tool for data collection, are inexistent. In countries with a large territory and limited resources like Brazil, the development of a cohort study able to identify environmental risk factors, based on web-based tools would be highly desirable. Therefore, in 2014, our research group, working for a recognized Brazilian reference academic institution, decided to develop a study with those characteristics, the Nutritionist Health Study – NutriHS (www.fsp.usp.br/nutrihs).

2.2 Web-based Data Collection

Epidemiologists conducting cohort studies often work with huge amount of data to evaluate the relationship between exposures and outcomes at multiple points in time. Application of these methods based on paper-and-pencil questionnaires requires trained professionals for collecting, sorting, coding, typing and checking data. These studies frequently are relevant for actions in public health demanding agility to be implemented. Therefore, researchers have been abandoned traditional methods, replacing by Web-based data collection.

Web-based questionnaires have several advantages compared with traditional tools, including data quality, reduction of costs and time from the study initiation to the receipt of analysable data (Schleyer and Forrest 2000, Wyatt 2000). The interactive element of this system allows the inclusion of visual aids; design issues can be simplify for responders; and also the scripts of the website can be immediately adjust to solve unforeseen problems or to incorporate preliminary results (van Gelder et al. 2010). Costs for printing, postage, and data coding and entry are avoided. On the other hand, the set-up costs, including programming and design, may be too high in studies with small sample sizes or in populations with low response rates (Adams and White 2008, Rodriguez et al. 2006, Dillman 2007).

During survey implementation on the Web, there are unknown issues prompting some authors to express concerns (Best et al. 2001, Schaefer and Dillman 1998, Smith 1997). These concerns include sampling problems, lack of participant access to computers with Internet connections, security, and response inconsistency across different media. The possibility of bias associated with collecting information over the Web previously raised is no greater than that introduced by traditional paper methods (Ekman et al. 2006).

In summary, it is notable that the majority of respondents prefers Web-based version to postal questionnaires or telephone interviews (Rankin et al. 2008, Touvier et al. 2010). This is particularly true for young people who were born during the Internet era. Being a more attractive method of data collection, more Web-based systems tended to be developed for epidemiologic studies.

3 MATERIAL & METHODS

This chapter includes details on the users of the e-NutriHS, as well as the instrument (e-NutriHS) and its prototype for pre-testing. The NutriHS was conceived during the year 2013 by our research group from the School of Public Health of the University of São Paulo, Brazil. The NutriHS is planned to be a cohort study mainly involving students of Nutrition Graduation Courses from Universities located in the State of Sao Paulo, but also graduates. For its first phase, launched in 2014, they are being invited to participate in the study, in which health-related data will be collected at a 3-year interval.

A web-based system was developed to gather information online of this specific subset of the Brazilian population, the e-NutriHS. Our software is unique, tailored to the researchers' purposes and to local requirements of the university server.

3.1 e-NutriHS Users

The choice of nutrition college students and graduates was based on four presumed conditions: a) potential skill to navigate on the Internet; b) availability of computer and Internet connections; c) expected high quality of data provided; d) local facilities in our University.

Nutrition students and graduates are being invited to this first phase of NutriHS. The ethical committee at FSP-USP approved this study. All student or graduate who accepts participation has signed Statement of Informed Consent electronically prior the fulfilment of questionnaires.

The ones who agree to participate are being investigated about the role of lifestyle habits on health conditions. Instruments used to collect feedback from users were e-mail and Facebook.

3.2 Survey Instruments

The e-NutriHS includes of 6 questionnaires regarding demographic and socioeconomic data, dietary habits, physical activity level, alcohol and tobacco use, anti-fat attitudes and personal and family medical histories. Contact information was obtained to help the participant tracking during the follow-up. Validated and internationally recognized lifestyle tools were used, since standardization is necessary for comparisons with other populations. Main tools employed in the NutriHS include the International Physical Activity Questionnaire (IPAQ) (Craig at al. 2003), Food Frequency Questionnaire (FFQ) (Fisberg et al. 2008) and the Antifat Attitudes Questionnaire (Lewis et al. 1997).

3.3 Online Social Network and Website

Colors for the website were carefully selected; green and orange are related to health, nutrition and youth. During the software building process, ease of use was frequently reviewed until become a friendly version. All the questionnaires were inserted in the website exactly as the printed version, except for essential instructions for completion. The length of each questionnaire page was defined based on a comfortable time for filling. Our logotype was created reinforcing the impact of increasing knowledge about nutrition on health.

In order to increase awareness of the NutriHS a social page in the Internet was created before starting the data collection. The most popular social network – the Facebook – was chosen to publicize the study (Figure 1).

Figure 1: Social online network of NutriHS (https://www.facebook.com/nutrihs).

Our page in Facebook includes a link to the e-NutriHS homepage. This place has also represented an important vehicle to disseminate scientific knowledge on nutrition for Internet users and to support the dialogue among researchers and users.

With the University of São Paulo Informatics Center support, the e-NutriHS website (Figure 2) was developed to expedite and further analyze data generated from the first phase of the NutriHS.

As our software and respective database would be hosted in the School of Public Health server, our team had to build the web system which was based on free programming languages, such as HyperText Markup Language (HTML), Cascading Style Sheets (CSS), JavaScript, Hypertext P reprocessor (PHP) and Structured Query Language (SQL), in database management system MySQL.

Before fulfilling the questionnaires, users are required to create a unique login and password. Afterwards, their access to questionnaires is released. When users enter incomplete or implausible answers, prompts alert them to complete

Figure 2: The e-NutriHS website (http://www.fsp.usp.br/nutrihs/index.html).

or to check information. Each page must be saved at the end in order to feed database. User workflow is depicted in Figure 3. e-NutriHS data obtained from any questionnaire can be transferred to excel format and provide print reports. Dietary data are immediately processed permitting estimates of daily energy and nutrient intakes, based on the nutritional composition data of the USDA National Nutrient Database for Standard Reference.

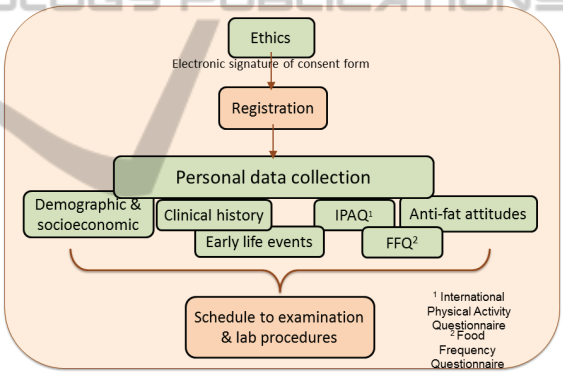

Figure 3: User online workflow.

3.4 The e-NutriHS Prototype

A prototype of e-NutriHS was created preceding online attachment for NutriHS users and hosted on a temporary url in order to test the system and to check its ease-of-use, data entry and website design.

Twenty volunteers, graduates at the School of Public Health who have been worked in the health area, aged 25 to 60 years old, were invited to test the prototype. They did the online registration and answered to all questionnaires.

To check if answers were accurate, one questionnaire per person was randomly chosen for paper responses, sent by e-mail. The volunteers also answered four simple questions about the design, method preferred and ease-of-use of the website:

1. Which method do you prefer?

2. Is attractive the design of the website? Would you change anything?
3. Were the questionnaires easy to answer?
4. Did you have any kind of problem when navigating the site?

4 INITIAL FINDINGS

4.1 Prototype Results

In average, volunteers fulfilled the web-based questionnaires 30 minutes after they had received the temporary url, meaning that these data were already ready to be analyzed. Regarding the paper version, they returned it at least one day after e-mail delivery. It is important to take into consideration that those paper questionnaires still needed to be coded to enter the database.

Based on the comparative analyses of the prototype with paper questionnaires, some technical problems were detected in the FFQ. These script errors were properly adjusted and solved.

Users preferred web-based method over the traditional one, considered the website attractive, made favourable comments about the colours used and the interactive elements and suggested no change. The questions were considered simple and the prompts convenient, facilitating fulfilment. No problem was mentioned related to navigation. An improved version of website was released based on the volunteers' opinions.

4.2 NutriHS Preliminary Results

In March 2014, NutriHS participants have started to use the online web-based system. No complain on e-NutriHS access and use has been reported. In general, users' satisfaction seems to be positive since only favourable comments have been received.

A total of 503 users were registered, being 459 college students and 44 graduates. Predominance of female sex was detected (93%) which was expected since the majority of nutrition students are women. Mean age of the sample is 23.8 (SD 6.6) years; 63.5% are Caucasian and 84.6% single.

Preliminary data on energy and nutrient intakes, physical activity levels, alcoholic beverages and tobacco uses and others are available allowing comparisons of students' profile at the beginning and end of the course. These conditions are in contrast with the nutritional epidemiological studies, which require a long period to get results.

Summaries of scientific papers on nutrition recently published should be responsible – at least in part – to the high access to our social network.

Taken together, these preliminary results should bring relevant information not only for the student, graduate and researcher but also to the academic institutions' staff.

5 LIMITATIONS & STRENGHT

There is a lack of originality in using web-based system to collect and disseminate health information. However, its use in cohort studies, conducted in emerging countries, exploring the lifestyle impact on health conditions, is rare. As far as we know, in Brazil, this kind of web-based research in nutritional epidemiology is a novelty.

The applicability of the current instrument is limited since this was specifically created for a subset of professionals. However, similar web-based systems could be adapted for other practices. Particularities of our design site limit comparisons with other similar web systems available.

6 CONCLUSIONS

Our initiative of building a website designed for collecting data for epidemiological studies, tailored to our local reality, is innovative under the perspective of the health informatics available in developing world.

Preliminary findings suggested that e-NutriHS may be useful to obtain data with potential to identify gaps and/or problems in graduation process as well as health disorders. In parallel, we speculate that this website might encourage students to search for e-learning on nutrition. Also, it is expected that college authorities could obtain some subsidies to review aspects of the academic curriculum.

Considering that web-based systems produce reliable data, are easy to use, less costly and less time-consuming, we conclude that our experience deserves to be shared, particularly with middle-income economy countries.

ACKNOWLEDGEMENTS

The authors have no conflict of interest to disclose. This study was supported by a scholarship grant from Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

Adams, J. & White, M. (2008) Health behaviours in people who respond to a web-based survey advertised on regional news media. *European Journal of Public Health*. 18 (3). p. 335–338.

Azevedo e Silva, G., Valente, J. G. & Malta, D. C. (2011) [Tendências do tabagismo na população adulta das capitais brasileiras: uma análise dos dados de inquéritos telefônicos de 2006 a 2009] *Revista Brasileira de Epidemiologia.* 14 (1). p.103-114.

Best, S. J. et al. (2001) An assessment of the generalizability of Internet surveys. *Social Science Computer Review*. 19. p.131–45.

Cade, J. E. et al. (2004) Food-frequency questionnaires: a review of their design, validation and utilisation. *Nutrition Research Reviews*. 17 (1). p. 5-22.

Colditz, G. A. & Hankinson, S. E. (2005) The Nurses' Health Study: lifestyle and health among women. *Nature Reviews Cancer*. 5. p. 388-96.

Couper, M. P. (2000) Web surveys – A review of issues and approaches. *Public Opinion Quarterly.* 64. p. 464–494.

Craig, C. L. et al. (2003) International Physical Activity Questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*. 35 (8). p. 1381-95.

Dillman, D. A. (2007) Mail and Internet Surveys: The Tailored Design Method. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.

Ekman, A. et al. (2006) Feasibility of using web-based questionnaires in large population-based epidemiological studies. *European Journal of Public Health.*21. p. 103–11.

Fisberg, R.M, et al. (2008) [Questionário de freqüência alimentar para adultos com base em estudo populacional]. *Revista de Saúde Pública*, 42. p. 550-554.

Hewson, C. M., Laurent, D. & Vogel, C. M. (1996) Proper methodologies for psychological and sociological studies conducted via the Internet. *Behaviour Research Methods, Instruments, & Computers.* 28. p. 186–91.

Hu, F. B. et al. (1997) Dietary fat intake and the risk of coronary heart disease in women. *New England Journal of Medicine*. 337. p. 1491-9.

Krantz, J. H., Ballard, J. & Scher, J. (1997) Comparing the results of laboratory and World Wide Web samples of the determinants of female attractiveness. *Behaviour Research Methods, Instruments, & Computers.* 29. p. 264–9.

Lewis, R. J. et al. (1997) Prejudice toward fat people: the development and validation of the antifat attitudes test. *Obesity Research.* 5 (4). p. 297-307.

Mikkelsen, E. M. et al. (2009) Cohort profile: the Danish Web-based Pregnancy Planning Study-'Snart-Gravid'. *International Journal of Epidemiology.* 38 (4). p. 938–943.

Rankin, K. M. et al. (2008) Comparing the reliability of responses to telephone-administered versus selfadministered web-based surveys in a case-control study of adult malignant brain cancer. Cancer Epidemiology, Biomarkers *& Prevention.* 17 (10). p. 2639–2646.

Rodriguez, H. P. et al. (2006) Evaluating patients' experiences with individual physicians: a randomized trial of mail, Internet, and interactive voice response telephone administration of surveys. *Medical Care*. 44 (2). p. 167–174.

Schaefer, D. R. & Dillman, D. A. (1998) Development of standard e-mail methodology: results of an experiment. *Public Opinion Quarterly.* 62. p. 378–97.

Schleyer, T. K. & Forrest, J. L. (2000) Methods for the design and administration of web-based surveys*. Journal of the American Medical Informatics Association*. 7 (4). p. 416–425.

Schmidt, W. C. (1997) World-Wide Web survey research: benefits, potential problems, and solutions. *Behaviour Research Methods, Instruments, & Computers.* 29. p.274–9.

Shettle, C. & Mooney, G. (1999) Monetary incentives in U.S. government surveys. *Journal of Official Statistics.* 15. p. 231–50.

Smith, B. et al. (2007) When epidemiology meets the Internet: Web-based surveys in the Millennium Cohort Study. *American Journal of Epidemiology.* 166 (11). p. 1345–1354.

Touvier, M. et al. (2010) Comparison between web-based and paper versions of a self-administered anthropometric questionnaire. *European Journal of Epidemiology*. 25 (5). p. 287–296.

Turner, C. et al. (2009) Cohort profile: the Nurses and Midwives e-Cohort Study - a novel electronic longitudinal study. *International Journal of Epidemiology*. 38 (1). p. 53–60.

van Gelder, M. M., Bretveld, R. W. & Roeleveld, N. (2010) Web-based questionnaires: the future in epidemiology? *American Journal of Epidemiology*. 172 (11). p. 1292-8.

Wilson, K. M. et al. (2012) Dietary acrylamide and risk of prostate cancer. *International Journal of Cancer.* 131 (2). p. 479-87.

Wyatt, J. C. (2000) When to use web-based surveys. *Journal of the American Medical Informatic Association*. 7 (4). p. 426–429.