

Data Collection in Hospital Emergencies: A Questionnaire Survey

Nouha Mhimdi, Wahiba Ben Abdesslem Karaa, Henda Ben Ghezala

Abstract—Many methods are used to collect data like questionnaires, surveys, focus group interviews. Or the collection of poor-quality data resulting, for example, from poorly designed questionnaires, the absence of good translators or interpreters, and the incorrect recording of data allow conclusions to be drawn that are not supported by the data or to focus only on the average effect of the program or policy. There are several solutions to avoid or minimize the most frequent errors, including obtaining expert advice on the design or adaptation of data collection instruments; or use technologies allowing better "anonymity" in the responses. In this context, and to overcome the aforementioned problems, we suggest in this paper an approach to achieve the collection of relevant data, by carrying out a large-scale questionnaire-based survey. We have been able to collect good quality, consistent and practical data on hospital emergencies to improve emergency services in hospitals, especially in the case of epidemics or pandemics.

Keywords—Data collection, survey, database, data analysis, hospital emergencies.

I. INTRODUCTION

NOWADAYS, a voluminous size of data is stored every day in data warehouses. The mass collected is from different sources, on the other hand, it is necessary to collect relevant, subject-oriented, integrated, non-volatile, historical and summary data, which will then help in the process of decision-making. The global volume of data will be multiplied by further than 3.7 between 2020 and 2025, then by 3.5 every five years until 2035, to reach the dizzying sum of 2,142 zettabytes [1]. However, several problems are encountered, we can cite, for example, the difficulties in extracting information [2] and/or knowledge, the increase in knowledge generation time from the warehouses, the information obtained is no longer usable, also the performances obtained are insufficient [3], [4]. This is due to the wrong choice of methods in data collection and analysis.

Currently, researchers are using online surveys as a data collection method for several reasons such as low cost, speed of data collection, decrease in errors during data entry [5]. On the other hand, the literature mainly highlights difficulties related to sampling control and the limitations such as representativeness, privacy, anonymity, self-selection bias, etc. too, the most important point is to ensure that the sample data can be correctly generalized statistically or analytically improper sampling can lead to erroneous conclusions [5]. These problems exist in all areas, in fact, in the medical field, the existing data are confidential or protected by copyright, it is a question of inadequacy of the available data (bad resolution, lack of adequate report). Also, collecting data for the benefit of

hospital emergencies becomes necessary to mitigate the problems encountered there, such as the significant congestion of the emergency services, the overload in services, the bad management of patient flows [6]. The resolution of these problems must be based on collecting relevant and efficient data leading to practical results [7].

Tunisia, like many countries, is not spared by the significant congestion of emergency services. However, even the resolution of these problems requires the collection of data; we are facing a big problem related to the availability, and the relevance of data. In fact, the existing relevant data sources such as survey data or the available time series are limited in terms of the time dimension, and the content is missing essential indicators [8]. In this context, we opted to collect relevant data, by carrying out an extensive survey based on developing a questionnaire containing targeted and practical questions.

This paper is organized as follows: Section II provides a brief overview of the work related to data collection; the proposed approach is described in Section III. Finally, in Section IV, the simulation and the results are presented, and then a conclusion and future work are given at the end of this article.

II. LITERATURE REVIEW

In the literature, many data collection methods exist, such as Individual face-to-face or remote interviews, focus group interviews [9], questionnaires, and written diaries [10]. The questionnaire is used in several ways manual writing or using an electronic medium with short answer questions. Several studies have used the questionnaire survey as a method of collecting data in various fields. In fact, in [11], the questionnaire was established open-ended by class students; they worked in groups to explore the results of 12 therapeutic sessions in a school context.

Many limitations were identified, for example, complications in transcribing recordings of groups made up of several participants and unnecessary discussions between participants, which wasted a lot of time. Similarly, [12] carried out a cross-sectional study in the emergency departments of four hospitals in Tunisia. This involves collecting data from a questionnaire from a sample of 1058 patients of both sexes, aged 18 and over. The objective was to collect information concerning the socio-demographic characteristics of responders, the motivations for choosing the emergency service, the time and reason for consultation, the diagnosis, and the severity of the state of health. The study aimed to analyze the reasons for resorting to emergency services and the seriousness of the patient's state of

Nouha Mhimdi is with Ensi, Tunisia (e-mail: nouhamhimdi@yahoo.fr).

health. The advantage is that the questionnaire is tested beforehand and administered by nursing investigators, the survey is systematic from the registration register by including prospectively the patients whose number is odd, and until the expected number of workers is reached, with stratification on the time slots. The downside is that using a face-to-face questionnaire to collect data near hospital emergencies can influence the quality of the obtained data [13].

The literature mainly highlights the difficulties related to data collection: significant problems with the quality of existing data, sometimes incomplete information about participants, an extension of the period or extension of the scope of data collection, limits of sampling such as confidentiality and self-selection bias [14]. Poorly performing sampling will lead to insufficient data and erroneous or incomplete and incorrect results, which leads to false conclusions and poor decision-making afterward. In addition, it is necessary to predict whether the data resulting from this sample can be generalized statistically or analytically.

The choice of responders is essential especially when it comes to the studied subject; their demographics, age, gender, and other relevant data can alter the results and ultimately lead to poor-quality data collection. Also, poorly designed questionnaires, the inability to allow participants to answer questions carefully, and incorrect data recording influence the achievement of consistent and complete data [14].

In the current research landscape, more and more researchers are using the online survey to collect data, disseminating it to relevant associations, the media, and interested individuals. The researchers noted several advantages: low cost, time-saving in data collection, reduction in errors during data entry, greater freedom of the respondent, large-scale survey dissemination [15].

A. Problem Description

Choosing to collect data not explicitly related to the critical evaluation questions will result in inadequate information and reporting. The online survey data collection method requires familiarity with computers and requires dexterity in handling the material [16]. It is more difficult for respondents to distinguish spam or advertising from serious mail. Some people are therefore suspicious of solicitation e-mails; others are also suspicious of computer security and fear viruses [17]. These elements make the selection of respondents even more complex and certainly affect response rates. Likewise, online surveys are generally non-probabilistic and uncontrolled; one cannot distinguish respondents from others. The non-probability sample is constituted and targeted in a manner that is not proportional to the general population, they are therefore unreliable [18]-[20]. Likewise, people living in remote areas, visible minorities, and low-income individuals are often under-represented, are not connected to the internet, and cannot participate in the survey, which will lead to poor results. The non-probability sample is constituted and targeted in a manner that is not proportional to the general population, they are

therefore unreliable [19]-[21].

B. Objective Work

To mitigate the limitations of data collection methods, we have opted for solutions to improve the collection of data. For this purpose, we first start by choosing the field of application. During this study, we are interested in the health field given the problems generated in recent years. Patient care is mainly done in hospital emergencies, which continue to see a sharp increase in the number of emergency department arrivals each year in most countries of the world.

Overcrowding and tension have become a real problem in the health system nowadays; insufficient human and material resources manifest them. Likewise, emergency physicians denounced flaws in the hospital system, mainly due to a lack of resources [22]. This leads to increased wait times that can worsen the condition of patients, causing dissatisfaction with patients, physicians, and increase the risk of mortality.

Our objective is to solve the problems afore mentioned to improve the quality of emergency services by carrying out a study allowing the collection of coherent, and interesting data, and to set up a process allowing their cleaning, their transfer between the different types of software analysis, and their availability, and accessibility at all times for verification purposes.

III. THE EVOLUTIONARY APPROACH

In all health systems, the emergency structure represents an important link in the chain of patient care. However, the number of arrivals to emergencies continuously increased, and enormous organizational problems were caused and negatively influenced the quality of services [23]. To remedy the problems that have been caused by the massive influx of emergency room visits, we propose an approach founded on a data collection system. The latter is based on a survey made using a questionnaire near hospital emergencies. Fig. 1 summarizes the steps to create a new database gathering the data necessary for optimized emergency management.

Our approach, Data Collection for Emergencies: DC-EMRG, allows us to collect data within the framework of hospital emergencies to have an interesting database that can attenuate the problems circumscribed these last decades in emergencies.

DC-EMRG is made up of three main stages:

- Creation of the DB-EMRG database
- Creation of the DB-CL database
- Merging of the two databases and production of the DB-RES database

A. Phase1: Creation of the DB-EMRG Database

The first step in the proposed approach is to collect data through a survey and create the DB-EMRG database. The sub-steps are summarized in Fig. 2.

The steps outlined in Fig. 2 are detailed in the following subsections.

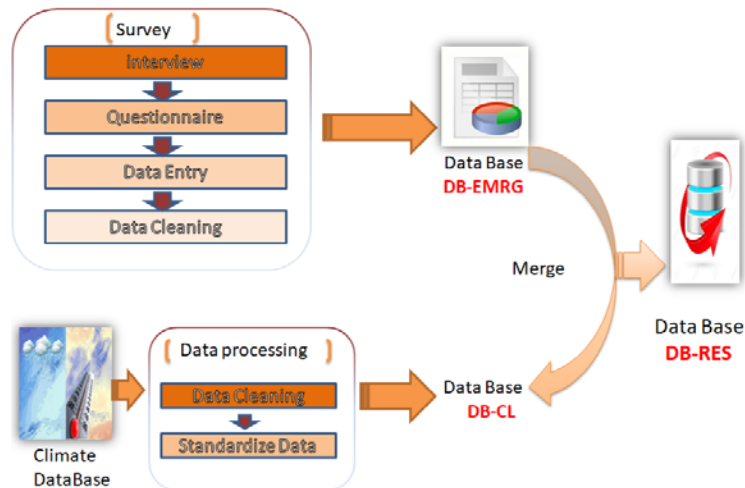


Fig. 1 Approached data collection DC-EMRG: Database creation DB-RES

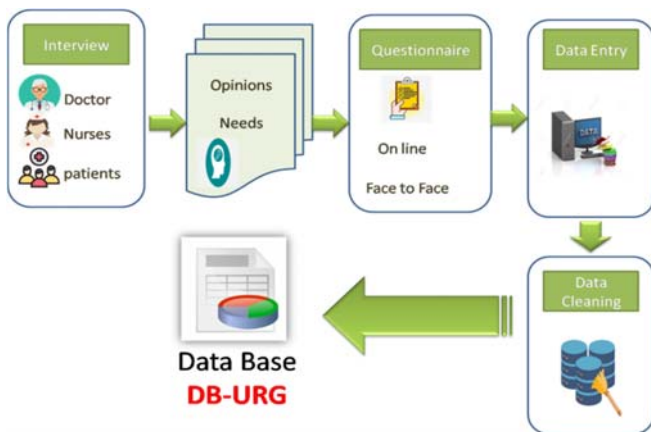


Fig. 2 Phase1: investigation and creation of DB-EMRG database

1. Interview

This step begins by determining the data collection problems and the tasks carried out to obtain quality data and clear and precise information. To this end, it might be useful to interview individuals affected by this study, such as emergency physicians, nurses, and patients visiting emergency departments.

The interview will answer a set of questions clarifying and indicating the main checkpoints that prevent the most common erroneous practices that can affect the quality of the data and give recommendations to prevent these pitfalls [24]. First, it is necessary to check whether other surveys on the same subject are made or done during the same period to see if the data in the emergency rooms are useful and can be exploited. Then, we collected a set of data after the interviews, we had a whole set of points of view, opinions, gaps, missing needs. We identified the set of requirements, on one hand, for the medical staff, and on another hand for the patients. Therefore, this step is considered as preparation for the next step.

2. Questionnaire Survey

The questionnaire survey is an interesting method of data collection; it encompasses any measurement strategy that

requires asking questions of respondents. Having reliable and relevant data depends essentially on the quality of the questions asked the knowledge of the responders and the credibility of the used methods [25]. The questionnaire survey consists of short questions to be completed with paper and pencil or a detailed one-on-one interview; also, it is done online via the internet or face-to-face.

The literature review shows that the limits of the questionnaire are linked globally to the sampling: addressing an uncontrolled sample and individuals who are always connected will influence the quality of the responders, people with low incomes are often under-represented, sometimes responders do not distinguish between spam e-mails and advertisements, and others fear viruses [26]. To overcome these limitations, we have opted for an approach that combines the two survey methods face-to-face and online to reach the collection of reliable, complete, and good quality data. Our work begins by first determining a sample of interviewees: We opted for a large sample size to achieve satisfactory results: 8829 individuals. We reached out to the population of northern Tunisia given the geographic, economic, and social specificities common to this region.

Responders are of different age categories, two genders, different living standards, living in remote rural areas and urban areas. We launched the questionnaire in parallel, online, and face-to-face. The questionnaire was prepared on the Google Forms platform and sent to relatives, friends, colleagues, and all acquaintances asking them to complete and share the form. We also printed the questionnaire; we visited all the regions of northern Tunisia and asked individuals to complete the form either on paper or online.

We proposed short, understandable, and targeted questions to allow respondents to respond more spontaneously to the questionnaire. The choice of question type will also affect the survey results; for this, we used qualitative questions based on impressions, and opinions to collect information intended to describe the topic rather than quantitative measurement questions used to collect raw, concrete data, in digital form. We used different types of questions: closed single-choice

questions, closed-ended multiple-choice questions, open-ended questions and ordered qualitative questions.

3. Data Entry

Collected answers following the paper questionnaire must be entered carefully to avoid double entry. Therefore, we used the online platform Google Sheets [27] that allows adding lines of answers. This step consists of adding to the online database the responses relating to minors, children, the elderly, and those who cannot answer the online questionnaire. Once the questionnaire has been entered, it must be classified and highlighted with a perceptible color.

After the data entry, the long-awaited results can be expressed in figures to facilitate analysis and guarantee optimal reliability, integration, and ergonomic.

4. Data Cleaning

The biggest challenges in data entry are human error, incomplete and inconsistent data, typos when entering data, and irrelevant responses. If the data are not clean, decisions based on those data might be wrong and affect the results [28]. Therefore, it is important to distinguish how to find errors and correct them to make the data more useful for good decisions. Therefore, to have a consistent, useful, and reliable database, we processed the responses line by line to correct errors, examine incomplete data and erroneous data. In addition, in this step, we have merged the columns concerning people who give answers instead of children or instead of individuals who do not have access to the internet. Also, the answers to qualitative questions are transformed into quantitative data to be represented in the database.

5. Creation of the DB-EMRG

The result of the presented steps is a DB-EMRG database gathering structured, coherent data, useful in the third phase of the proposed approach. This database contains a set of data relating to hospital emergencies in Tunisia, which will help to alleviate existing and future problems in emergencies.

B. Phase 2: Creation of the DB-CL Database

The second phase consists of building the DB-CL database gathering the climate values for 2020. Fig. 3 shows the architecture of the followed steps:



Fig. 3 Phase 2: Development of the climate database DB-CL

- 1) Download the existing national climate database: This database brings together Tunisian climate data (temperature, humidity, visibility, heat index, weather conditions, etc.)
- 2) Clean the data: Keep only useful data, needed for this study and remove the unnecessary data. Also, correct some

errors, get rid of inconsistent data, and delete missing values.

- 3) Standardize the data: put the values in a standard format, for example, put the dates in standard format (dd-mm-yyyy), all decimal numbers of the same format, and all integers of the same format.
- 4) Create a database that we named DB-CL, gathering all the necessary and useful meteorological values needed in our study.

C. Phase 3: Creation of the DB-RES Database

This phase is the final step of our proposed approach to merge the two created databases DB-EMRG and DB-CL, after putting the data in a standard format. We named the resulting database DB-RES: it contains 49 variables and 8829 structured rows, ordered and sorted in ascending order of the date. All the data are interesting and useful in our work; they concern hospital emergency services, as well as patients. These data will be of great interest for a decision support system aimed at improving the quality of services and alleviating existing problems in emergencies in hospitals.

IV. SIMULATIONS AND RESULTS

We have opted for the analysis of our data obtained in the database DB-CL by using the programming language R [29], which is a statistical language with built-in statistical and graphical data processing functionalities. Also, we used the IBM SPSS statistical [30] software to ensure advanced statistical analysis through its scalability and adaptation to any project regardless of their size and level of complexity.

We used the "pie chart" [31] frequently to express proportional data or percentages to visualize the frequency distribution of patients visiting hospital emergency rooms by gender in approximately 8829 patients. Fig. 4 shows this graphical representation.

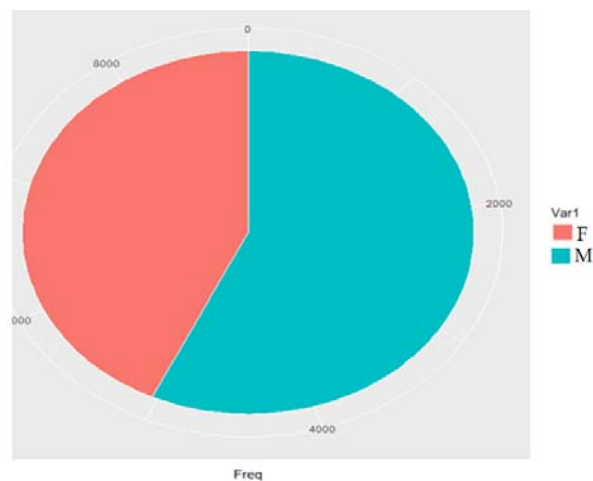


Fig. 4 Distribution of patients by gender

As shown in Fig. 4, the number of women represents 43% compared to men patients' percentage 57%; this shows that the number of male patients is higher than that of females. We

validated the data by nurses and emergency physicians who confirmed these percentages and justified the responses by the influence of the schedule of emergency room visits, generally at night, and the age groups of sick patients. The distribution of the number of patients by gender and by age category is presented in Fig. 5.

According to Fig. 5, the number of male patients is higher than that of women for ages between 35 and 45 years and those between 55 and 65 years. According to our study, this comes down to undercurrent health problems generally affecting men, such as digestive problems, respiratory problems due to

frequent smoking, cardiovascular problems, and pain from work accidents or road accidents. This appeared in particular periods; for this, we have represented the distribution of patients per month in Fig. 6.

Fig. 6 shows the activity peaks in emergencies in particular periods; the curve reaches its limits in March and April, between May and June, August, and December. This comes down to several external factors such as climate change which makes you feel bronchitis and flu, the appearance of an epidemic such as COVID-19 and school holidays which influence the number of accidents on the road.

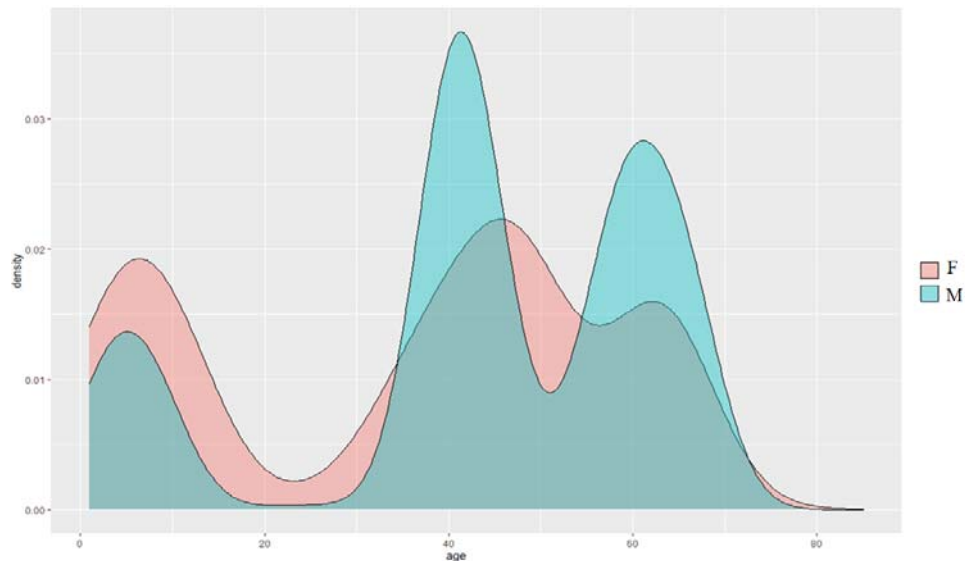


Fig. 5 The distribution of the number of patients by gender and by age category

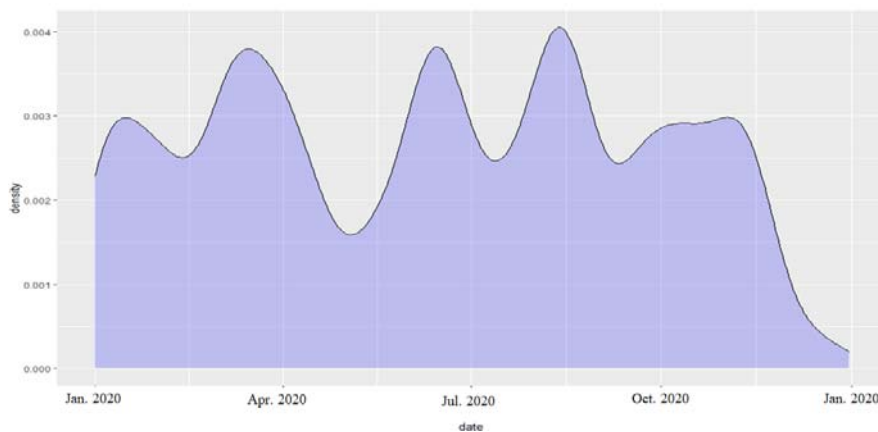


Fig. 6 Breakdown of the number of patients by date

V. CONCLUSION AND FUTURE WORK

Guaranteeing good quality data collection is becoming necessary given the large mass of data that appears every day and comes from several sources: messages, signals, climate information, recording, video, etc. [32]. However, as the existing data in different fields do not meet the needs of researchers in health, enormous organizational problems have forced emergency structures to adapt to various changes and

seek solutions. Furthermore, the management of emergencies becomes one of the main missions of health establishments that must respond to an exponential demand for unscheduled care, which has posed the problem of dimensioning the structure and its resources [33]. In this context, we proposed an approach for data collection to address existing problems and gaps. Our idea consists of creating a new database, the DB-RES database resulting from merging of the two databases named DB-EMRG and DB-CL, created in the whose creation steps are presented

in this paper. The database obtained DB-RES can be used in future work to resolve the problems generated in the health system such as the anticipation of tension in hospital emergencies, also we can use the data collected for the optimized management of material and personal resources in hospital emergencies.

REFERENCES

- [1] <https://www.journaldunet.com/statista/>
- [2] Y. Feng, D. Duives, W. Daamen, S. Hoogendoorn, "Data collection methods for studying pedestrian behaviour: A systematic review", *Building and Environment Volume 187*, 107329, 2021
- [3] B. Hensen, C. R. S. M. Young, M Simwinga, N Abdelmagid, J. Banda, C. Mavodza, A. M. Doyle, C. Bonell, H. A. Weiss, "Remote data collection for public health research in a COVID-19 era: ethical implications, challenges and opportunities", *Health Policy and Planning, Volume 36*, Issue 3, Pages 360–368, 2021
- [4] K. Rosenthal, S. Strecker, O. Pastor, "Modeling Difficulties in Data Modeling", International Conference on Conceptual Modeling, pp 501-511, 2020
- [5] C. Andrad, "The Limitations of Online Surveys", *Indian Journal of Psychological Medicine*, Volume: 42 issue: 6, page(s): 575-576, 2020
- [6] Pang, P., Chang, S., Verspoor, K., Clavisi, "The use of web-based technologies in health research participation: Qualitative study of consumer and researcher experiences". *Journal of Medical Internet Research*, 20(10), e12094, 2018
- [7] J. Guo, H. Xie, M. Liang & H. Wu, "COVID-19: a novel coronavirus and a novel challenge for oral healthcare", *Clinical Oral Investigations volume 24*, pages 2137–2138, 2020
- [8] S. Jribi, H. Ben Ismail, D. Doggui, H. Debbabi, "COVID-19 virus outbreak lockdown: What impacts on household food wastage?", *Environment, Development and Sustainability volume 22*, pages 3939–3955, 2020
- [9] N. Hopwood, "Conceptions of geography: research findings and methodological implications". *International Research in Geographical and Environmental Education*; 18(3): 185-197, 2009
- [10] E. Vo-Jutabha, K. Dinh, J. Mchale, and al. "A qualitative analysis of Vietnamese adolescent identity exploration within and outside an ethnic enclave.", *Journal of Youth & Adolescence*; 38(5): 672-690, 2009
- [11] P. T. R. Malott, K.M., J.M. Maldonado "Exploring group activity therapy with ethnically diverse adolescents". *Journal of Creativity in Mental Health*, 3(3), 285-302. 2008
- [12] H.S khiri, N. Somrani, M. Hsairi, "Motifs du recours aux services d'urgence des principaux hôpitaux du Grand Tunis", *Eastern Mediterranean Health Journal*, 10.26719, 2012
- [13] E. Braekman, R. Charafeddine, and al., "Mixing mixed-mode designs in a national health interview survey: a pilot study to assess the impact on the self-administered questionnaire non-response", *International Journal of Public Health volume 65*, pages 5–16 (2020)
- [14] J. Denis, F. Guillemette, J. Luckerhoff, « Introduction : les approches inductives dans la collecte et l'analyse des données", *the journal Approches inductives Volume 6*, Issue 1, p. 1–, 2019
- [15] V. Braun, V. Clarke, E. Boulton, L. Davey, C. McEvoy, "The online survey as a qualitative research tool", *International Journal of Social Research Methodology*, 2020
- [16] S. Johnson and K. Rao. "Survey Mode Effects on Data Quality: Comparison of Web and Mail Modes in a US National Panel Survey". *Social Science Computer Review* 30 (2): 212-228, 2012
- [17] D. Frippiat and N. Marquis. "Les enquêtes par Internet en sciences sociales : un état des lieux". *Population* 65 (2): 309-338, 2010
- [18] H. Lindhjem and S. Navrud. "Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes". *International Review of Environmental and Resource Economics (Forthcoming)*, 2011
- [19] D. Syeager, J. A. Krosnick, and al. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples", *Public Opinion Quarterly* 75 (4): 709-747, 2011
- [20] L. Bstephenson, and J. Crête, "Studying Political Behavior: A Comparison of Internet and Telephone Surveys". *International Journal of Public Opinion Research* 23 (1): 24-55, 2011
- [21] A.-M. Dussaix, "La qualité dans les enquêtes". *Revue Modulad* 137 (39), 2009
- [22] G. Tpisheh, E.; and al. "Challenges for Hospital Resilience in Emergencies and Disasters: A Qualitative Study in Iran". *Journal of Clinical & Diagnostic Research*. Vol. 13 Issue 9, p1-8. 8p, 2019
- [23] A. Asheim, L P. Bache-Wiig, N. Pleyim, O. Uleberg, J. Dale, S M. Nilsen, "Real-time forecasting of emergency department arrivals using prehospital data", *BMC Emergency Medicine volume 19*, 42 , 2019
- [24] C. Son, S. Hegde, A. Smith, X. Wang, F. Sasangohar, "Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study", *Journal of Medical Internet Research*, Vol 22, No 9, 2020.
- [25] M. Ikart, "Survey Questionnaire Survey Pretesting Method: An Evaluation of Survey Questionnaire via Expert Reviews Technique", *Asian Journal of Social Science Studies* ISSN 2424-8517, ISSN 2424-9041, 2019
- [26] C. Carpinio, D. Morab, M. Simone, "On the use of questionnaire in residential buildings. A review of collected data, methodologies and objectives", *Energy and Buildings Volume 186*, Pages 297-318, 2019.
- [27] L. W. Tim, M. and al., "Googling your hand hygiene data: Using Google Forms, Google Sheets, and R to collect and automate analysis of hand hygiene compliance monitoring", *American Journal of Infection Control Volume 46*, Issue 6, Pages 617-619, 2018
- [28] X. Chu, I. Ilyas, S. Krishnan, J. Wang, "Data Cleaning: Overview and Emerging Challenges", *SIGMOD '16: Proceedings of the 2016 International Conference on Management of Data June*, Pages 2201–2206, 2016
- [29] F. Husson, S. Lê, J. Pagès, "Analyse de données avec R", *HAL Id : hal-01292429*, 2016
- [30] D. Lawrence, S. Meyers, G. C. Gamst, A. J. Guarino, "Performing Data Analysis Using IBM SPSS", *Hoboken: Wiley*, 2013.
- [31] D. Zheng, M. Plaisent, C. Zuccaro, P. Bernard, N. Daghfous, S. Favreau, « L'analyse des données de sondage avec SPSS: Un guide d'introduction", *Québec (Québec): Presses de l'Université du Québec*, 2018
- [32] B. Gauthier, I. Bourgeois, " de la problématique à la collecte des données", *Presses de l'Université du Québec*, ISBN 276054396X, 9782760543966, 2016
- [33] A. Brown, M. D. Cadogan, "Emergency Medicine: Diagnosis and Management", *Taylor & Francis Group*, ISBN 036746991X, 9780367469917, 2020.