CHAPTER XII

Training related to research capacity

M. Gonzalo CLAROS DÍAZ
University of Málaga (Spain)
Department of Molecular Biology and Biochemistry
claros@uma.es

Azucena STOLLE ARRANZ
University of Valladolid (Spain)
Library of the University of Valladolid. European Documentation Centre
azucenamaria.stolle@uva.es

Agustín MAYO ÍSCAR
University of Valladolid (Spain)
Department of Statistics and Operations Research
agustin@med.uva.es

1. Introduction

Research methodology is an essential area in the university environment and in fact constitutes the fundamental axis of master's and doctoral programmes in university and research centres. From a practical point of view, it is a set of techniques, procedures and instruments that are used in research in a broad sense: from the formulation of the object of study to the publication of the results and conclusions obtained through articles, reports, books, etc., and their dissemination by any academic, scientific and/or social means. Between the beginning and end phases, we must not lose sight of other intermediate ones that can be summarised as the search for information, as well as its evaluation, selection and organisation, the analysis of data from the qualitative and quantitative points of view, the elaboration and writing of publications, etc.

It is crucially important for those who are going to train the future leaders of the different areas of society to have extensive knowledge of research methodology. In this regard, it is seen as fundamental that, once a research question has been generated, they possess the necessary rudiments to plan and execute an efficient search for information among sources of proven quality. Future trainers must be aware of the important role that different types of scientific documents play in communicating research results. Among

them, articles have a decisive role and could be represented as the atoms of scientific matter or the bricks in the wall of knowledge. Likewise, it is essential that they bring students closer to the process of publishing works in scientific journals, mentioning the decisive importance of key elements such as anonymous peer review. In this contextual framework, it is essential to know and use the computer and technological tools available to respond to the different tasks related to research work, many of them today based on free-use computer programs accessible to everyone anywhere in the world.

A priority area in any research is related to statistics and quantitative analysis, which seek to answer any research question by analysing the information obtained from samples of the reality under study or from data produced experimentally in an attempt to reproduce a certain process of interest. In this research a fundamental role is played by statistical techniques for data summary and representation which describe the observations available and the inference techniques that facilitate the obtainment, from data collected from the sample, of approximations to the parameters specified by the models governing the real situations from which the observations come.

We cannot fail to mention the huge strides that statistics has taken in recent years. Although much statistical methodology was available in the 1980s, it was the emergence of the personal computer and its wide dissemination that made it possible for much of this methodology to be applicable in multiple fields of quantitative research and industry. The implementation of the statistical methodology available for its application in very different situations led to the development of new procedures that better meet the needs raised. In recent years, technology has facilitated the automated acquisition and storage of huge amounts of data of different kinds. The available data are not only numerical or qualitative observations, they are also images, sounds, videos that contain information from different realities and that must be analysed in an automated way to respond to different objectives. This means that statistical techniques are undergoing a great development process to meet these challenges that are posed. Trainers must be participants in this important development that is taking place in Statistics and know first-hand the practical applications of its use in the course of their research and training activity.

2. Framework of training and its characteristics

2.1. Context of training in research capacity

The training in research capacities was integrated into the second strategic line of the project "Support to the Ministry of Higher Education and Scientific Research for the reinforcement of the pedagogical skills of teachers-researchers and governance capacities of administrators." The purpose of this strategy was to support teachers in carrying out their pedagogical and research practice.

The training activities that made up this line of action were the following:

1. Structure of a training unit and design of teacher training plans.
2. Implementation of the hub for online teaching in Algeria: infrastructure for online teaching and training programmes for teachers.
3. Teacher training in teaching methodologies.
4. Teacher training in teaching technologies.
5. Teacher training in personal and social development.
6. Teacher training in research skills.
7. Teacher training in internationalisation.
8. Courses on demand for university centres.

The training to provide Algerian university teaching staff with knowledge and skills in research was the sixth line of action of those referred to above, and its purpose was to respond to the strategic objective of providing researchers with the basic knowledge allowing them to pose a research problem, contextualise it and find quality information, and to use this information in academic publication and dissemination: writing, structuring and publishing scientific documents.

Activity 6 had a transversal character within the entire strategic axis and, in this sense, transversal skills imply those acquired in three areas:

a. Knowledge: mastering and differentiating concepts, theories, models and methods in the areas covered
b. Implementation: knowing how to implement a treatment in practice, knowing how to develop a plan, knowing how to present a report or a publication
c. Attitude: having a proactive attitude, mastering social skills and communication techniques, acting ethically in the field of communication and scientific dissemination, etc.

In this contextual framework, the initial starting point was to provide teachers at Algerian universities with knowledge about:

— Use of statistical research programs.
— Use of bibliographical search tools in the different areas of knowledge.
— Use of database management programs.
— Publication of scientific articles in international journals.

These aspects were later specified in specific objectives that were adapted, firstly, to the starting situation of Algerian teachers and their most immediate needs and, secondly, to the trends, habits and customs of research at an international level according to the different areas of knowledge.

2.2. Objectives of research training

The specific objectives pursued were organised around different thematic areas that provided the theoretical cognitive framework for contextualisation and the practical skills necessary for Algerian teachers to acquire good research skills.

1. Knowledge and use of bibliographical research tools in different branches of knowledge.

a. Know the types of documents in the scientific field, based on their information content and the communication objectives.

b. Know the types of sources of academic and scientific information, nationally and internationally.

c. Know the main information resources both of a multidisciplinary nature and in each field of knowledge.

2. Location of scientific and academic information in catalogues, databases, journal/review websites and other resources.

a. Learn the techniques of documentary and bibliographical search in the scientific and academic field.

b. Differentiate the various phases of an information search process.

c. Learn to describe a research problem and precisely define an information search equation.

d. Learn to efficiently use database search systems and other sources of information in order to obtain pertinent results.

e. Know the usefulness of informative alerts in the bibliographical research process.

3. Scientific citation, development of bibliographical references and bibliographical managers.

a. Know the characteristics of scientific publication and its objectives.

b. Reflect on the obligation to use citations correctly and ethically.

c. Distinguish the different cases of plagiarism that can occur in the academic field.

d. Know the most common citation standards in each field of knowledge.

e. Know tools to obtain bibliographical references in different styles automatically.

f. Know the bibliography managers and their use in the organisation of information and automated generation of citations and bibliographies.

4. Use of basic tools for the management, citation and publication of data and writing of academic and scientific papers: *LibreOffice* and *Zotero*.

a. Know the advantages of open source programs.

b. Know the importance of publishing scientific results, and how to do it.

c. Become familiar with the *Writer* word processing program, which is part of *LibreOffice*, learn how to use styles, and know what templates are, how to use them, and how to generate them.

d. Basic handling of *Zotero* to search for references and save them.

e. Using *Zotero* with *LibreOffice* to insert bibliographical citations in documents: how to add citation styles and change them in the word processing program.

5. Researcher profiles on the main international platforms and their importance for institutions and researchers: *Publons, ORCID, Google Scholar* and others.

a. Know the problems around researchers' digital authorship.

b. Understand the need for signature and authorship standardisation.

c. Know the importance of researcher profiles in the field of authorship, institutional linkage and the dissemination of scientific publications.

d. Know the main international scientific platforms on which to create a researcher profile: *Publons-ResearchID, ORCID.*

e. Know the main academic and general social networks for collaboration and dissemination of publications.

f. Recognise the importance of institutional or thematic open archives in the dissemination of information.

g. Learn to plan a strategy for creating, managing and maintaining researcher profiles and dissemination on social networks.

h. Know the characteristics of the specialised search engine *Google Scholar.*

i. Learn how to create and maintain a researcher profile on *Google*, its characteristics and the *Google H index*.

6. Statistical methodology

a. Introduce the role of statistics in the research methodology.

b. Introduce basic data description techniques.

c. Introduce basic inference techniques and their interpretation

7. Use of statistical software in research and practical application to real cases: R and Calc (*LibreOffice*).

a. To know the important role played by random assignment of individuals in clinical trials and the random selection of samples in sample-based research

b. Know the possibilities offered by spreadsheets to generate pseudo-random numbers that facilitate the tasks of randomisation and sample selection for the researcher.

c. Show the possibilities of spreadsheets as a teaching tool in the teaching of descriptive statistics and inference techniques.

8. Statistics with *R*

a. Introducing *RCommander* as a statistical package accessible to non-statistical researchers.

b. Apply the basic techniques of descriptive statistics and inference to compare two treatments.

c. Introduce R programming of inference tools.

d. Introduce clustering algorithms and their implementation in R.

e. Introduce the predictive rules based on logistic regression and their obtainment in R.

9. Graphic curriculum vitae with R and *Google Scholar*

a. *Google Scholar* utility to make yourself known.

b. Meaning and perversions of the H index and the impact factor.

c. Use of an R script based on the *scholar* package to obtain a graphic curriculum and compare the curricula of several people.

10. Dissemination of information through scientific publications and international journals, open access to science and dissemination on academic social networks.

a. Learn about the movement towards open access to science and information.

b. Know the field of scientific publication, its characteristics and the different means of publication.

c. Identify bad practices to avoid in scientific publication.

2.3. Methodology and expected results

The training methodology initially planned was based on face-to-face training using on-site electronic means to support the content. However, the COVID-19 pandemic made it impossible to travel to Algeria and, therefore for this direct and personal training to take place. For this reason, the *Moodle* training platform and the *Blackboard Collaborate* video conferencing platform were used.

1. The *Moodle* platform was used as the medium for the contents: basic and complementary reading information that was considered of interest (from the presentations to the bibliography, articles, links, etc.), and also for the participation in forums, as well as for the delivery and qualification of the mandatory tasks that the attendees had to deliver by the end of the week.

2. The *Blackboard Collaborate* platform was used daily for video conferences that had been previously set according to the usual morning and afternoon schedule in all the training activities of the project.

One of the main features of the video-conferences scheduled is that most of them were interconnected, either because they were given by the same expert and related to each other, or because, although they were given by different experts, they were closely related to the subject, being based on the same tools or programmes. The training schedule was established precisely to emphasise the connection between the contents and encourage the training to have a progression from theory to practice that facilitated the understanding and completion of the mandatory tasks that the participants had to present at the end of the week.

Although the training sessions were based on the materials previously uploaded to the *Moodle* platform, these were continuously combined with direct consultation of other sources prepared to show certain aspects, expand them, complement them with other information, etc. This method provided greater dynamism and also facilitated the

participation of the participants in the activities through small tasks that were entrusted to them at the time or through discussion through the chat about an issue.

As already indicated, the attendees had to deliver several tasks in which they applied the knowledge obtained throughout the week. The tasks that were requested were the following:

The first deliverable task was related to the search for bibliographical information, for which models, schemes, thesauri, etc. were provided.

The next one was looking for the application of style sheets and models to design a thesis model that also included how to cite scientific articles, using *LibreOffice*.

The third task was related to the application of statistics to a case using the *Calc* tool (*LibreOffice*).

Those attending this activity had a great responsibility in transferring the knowledge acquired to other teachers and researchers at their centres and universities, in addition to the teaching commitment they had towards their students, a fact that reinforced the interest and monitoring of the training sessions and practical activities proposed by the trainers.

In this context, the expected results in this training activity were, among other aspects, that the teachers of the Algerian universities strengthen the skills related to key aspects for university teaching and research, such as understanding research methodologies and techniques, scientific documentation processes, scientific communication techniques and methods, knowledge of statistical data processing, information management, analysis and interpretation, writing of research papers, and their dissemination to the scientific community and society in general.

To do this, not only were they provided with the basic knowledge to understand the procedures explained, but the practice of some of this knowledge was stimulated and they were provided with tools that they could adapt and use directly in both their teaching and their research activity. With all this, it was hoped that the participants would acquire the following work methodologies:

— Research methodology in the search and analysis of information.
— Methodology in the writing of academic and research papers.
— Methodology in the dissemination of publications.
— Statistical methodology.

3. Good practices

3.1. Search, use and dissemination of information

The Algerian teachers were provided with theoretical information that allowed them to identify the types of scientific documents and their communication objectives, the different sources of information and how to extract pertinent information from them, the use and ethics of information, and how to disseminate their own information through

academic and social profiles and networks. Likewise, they were provided with information on cases and methodologies easily adaptable to each area of knowledge and need.

3.1.1. *Sources of bibliographical information*

In order to properly search for information, it is essential to know the different types of scientific information that exist and what their informative purpose is in the academic-scientific field. Once the type of information and document that is required has been identified, we are ready to search and select the source of information that can provide it to us.

— Types of documents and their informative quality: journal/review articles, books, patents, reports, laws and regulations, factual data such as statistics or photographs, minutes of congresses and assemblies, theses and reports, etc. They were provided with different classifications with different degrees of completeness, among which the Classification et description détaillée de la Bibliothèque de l'Université de Montréal stands out: https://bib.umontreal.ca/guides/types-documents .

— Information sources: not only theoretical information on its content and use was provided, but the most useful Algerian, francophone and international sources of information on the different categories were indicated as exhaustively as possible:

- catalogues: they provide access to books, journals/reviews, theses, etc. Their main purpose is to allow documents to be located, and they are the fundamental basis of the inter-library loan service at a national and international level. Of special interest are the collective catalogues that pool the stocks of various specialised or university libraries. They were provided with specific catalogues such as the Catalogue Collectif d'Algérie (https://www.ccdz.cerist.dz/), the Catalogue du Système Universitaire de Documentation de la France SUDOC (http://www.sudoc.abes.fr/) and the global collective catalogue Worldcat (https://www.worldcat.org/).
- Likewise, the catalogues of national libraries were emphasised due to their usefulness as they contain the bibliographical heritage of each country: Bibliothèque Nationale d'Algérie (https://www.m-culture.gov.dz/index.php/fr/biblioth %C3 %A8que-nationale), Bibliothèque Nationale de France BNF (https://catalogueue.bnf.fr), or the compilation made by the Bibliothèque Nationale du Canada on Bibliothèques Nationales de la Francophonie (https://www.banq.qc.ca/documents/a_propos_banq/nos_publications/nos_publications_a_z/Bibliotheques_nationales_Francophonie.pdf),
- Databases: they can be bibliographical databases (they provide bibliographical references and sometimes access to the full text of articles, conference proceedings, book chapters, press releases, etc.) or factual databases (they provide information expressed in symbolic, graphic or numerical language, such as statistics, images or photographs). With few exceptions, most academic databases are subscription based.

- The most important databases in the international arena are multidisciplinary in nature and are the basis for obtaining impact metrics for publications and authors. These databases are: Clarivate's Web of Science and its InCites bibliometric indexes (https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-about.htm), and Elsevier's Scopus and its CiteScore indicators (https://service.elsevier.com/app/answers/detail/a_id/14880/c/10547/supporthub/scopus/).
- Finally, tools were provided for searching and locating databases by areas of knowledge and speciality, among which the selection made by the Bibliothèques de l'Université de Montréal stands out for its degree of detail and ease of consultation (https://bib.umontreal.ca/explorer).
- Thesis and end-of-studies projects: the sources of information for this type of document are characterised by their high dispersion, which makes it necessary to search in different sources. However, with the expansion of open access in the academic world, many of these documents are available in databases or repositories with free and open access such as Open Access Theses and Dissertations (OATD), DART- Europe E- theses Portal, TEL - Thèses en ligne, HAL archives-ouvertes.fr, thesis in OpenAIRE of the European Union, and many more.
- Journal websites: these are similar to databases, but give access only to journal articles. Access can be paid or free in the case of open access journal websites. In the French-speaking world, the best known are: Cairn.info, OpenEdition Journals, Revues.org ,, Revuesonline Persée, Les revues en libre accès-Science ouverte France, etc.,
- Open files, also called repositories: they are document files of all kinds in open access. The most widespread are files created by academic and scientific institutions, such as university repositories. There are very many repositories around the world, and the main tool to locate them is the Directory of Open Access Repositories (DOAR) (https://v2.sherpa.ac.uk/opendoar/), or also on the Dépôts institutionnels wiki mis en place par les universités et centres de recherche by disciplines at http://oad.simmons.edu/oadwiki/Disciplinary_repositories and, in Algeria, at ISTeMAG (Optimisation de l'Accès à l' I nformation S cientifique et Te chnique dans les Universités du Mag hreb).
- Locate specialised blogs and conferences: this is very scattered information, but very useful to follow specific lines of research or detect new trends, as well as to locate professional conferences and seminars to attend. They can be found in social networks but also in some sources of information that collect this type of publications, such as specialised blogs in OpenEdition (https://www.openedition.org/catalogueue-notebooks) and Congresos en Calenda (https://calenda.org/search).

3.1.2. *Research techniques*

Special emphasis was placed on the need to establish a written search strategy prior to direct searching in any source of information. This phase of the research methodology

is essential because it allows the main research topic and secondary aspects to be specified, the specific terminology to be selected, its chronological, geographical or other limits to be defined and finally a search strategy to be established that can be evaluated according to the results obtained from the information sources. This methodological phase is important enough to have been the subject of one of the compulsory tasks of the course, and is related to the following aspects:

— Establishment of key research concepts: synonyms, complementary terms, geographical, chronological limiters, etc. For this, thesauri and specialised terminologies, encyclopaedias or manuals can be used.

— Identify the most appropriate sources of information to search for information and indicate which ones: catalogues, general and/or specialised databases, journal websites, open files, etc.

— Formulate the search equation with the selected terms by analysing the key concepts, and run it on the selected information sources.

— Evaluate the results obtained: refine the search equation in light of the results; to do this, apply limiters if the results are excessive, or broaden the search by replacing or eliminating any of the previously used terms.

— Create informative alerts in the information sources of interest to us: create alerts based on the optimal search equation to obtain information from the new publications that appear in our field of research.

3.1.3. *Citation. Bibliographical reference styles*

The citation of authors and works where new ideas, thoughts and empirical results have been published is a cornerstone of science. In fact, the recognition of a scientific work is measured mainly by the number of citations received by the scientific community and today constitute the basis of bibliometric studies that evaluate the impact factor of publications and researchers.

The citation of previous works has two aspects, one of an ethical nature and the other of a technical nature. From an ethical point of view, the citation must always be made with honesty and truthfulness, avoiding all kinds of plagiarism or improper mention of works not consulted. In this sense, sufficient resources were provided to identify, detect or correct plagiarism, for which we used as an example the Aix-en-Marseille Libraries biblioguide 'Plagiat, droit d'auteur, paraphrase, copier coller etc.: Plagier c'est…' (https://bu.univ-amu.libguides.com/plagiat). Automated programs to detect plagiarism, such as *Turnitin, Compilatio, Ephorus* or *Scribbr*, are increasingly being used in the educational field, although there are many more tools that teachers, researchers and students should also know.

Finally the *Subpœna tool from the University of Bordeaux was provided (https://www.u-bordeaux.fr/Actualites/Des-bibliotheques/Subpoena-un-serious-game-contre-le-plagiat)* which is a puzzle game based on the reconstruction of bibliographical references

and the detection of plagiarism, and which can be a useful tool to train undergraduate students in these subjects.

From a technical point of view, it is necessary to know the standards and styles of citation and bibliographical reference most used in each field of knowledge in order to follow the same conventions of the scientific community. Normalisation facilitates the unequivocal and precise identification of the sources used in our own study, analysis and research process, and this allows the construction of impact metrics on the one hand, or also indicators of use, consultation, downloads, etc., on the other. In order to know the most appropriate citation and reference standards and practices for each area and situation, various sources of information were provided, such as the guide provided in Scribbr by Justine Debret (https://www.scribbr.fr/citation-des-sources/apercu-des-styles-de-citation/). The "Cite" and "Create bibliographie" citation tools, which include some academic databases such as *Scopus* from Elsevier, *Web of Science* from Clarivate or all the databases hosted on the *Proquest* platform are also very useful.

And finally, they were given information and training on bibliographical managers, specifically *Zotero* but also *Mendeley* and *Endnote*, which allow the bibliography to be organised and automatically generate both the citations and the final bibliography of the cited documents in the works themselves; furthermore, these tools make it easier to change from one citation and reference style to another according to the researcher's needs, without having to master the corresponding norms and styles.

3.1.4. *Researcher profiles*

The management of the digital identity of the researcher through the personal profile is today a priority. This identity would be formed by the information that allows the person to be identified and distinguished in a lasting and unequivocal way, and is made through the signature and standardised name, a photograph, institutional affiliation, membership of research institutes, and professional category. Investigator profiles allow you to answer the questions: Who is it? Where does he/she work? What area does he/she specialise in? What projects does he/she have or in which ones does he/she collaborate? What has he/she published, and where?

One of the most outstanding qualities of the profiles is the dissemination of the publications and work carried out. In all academic and social platforms, researchers have the ability to integrate the metadata of the articles, as well as the full text, in accordance with the editorial policies on open access and intellectual property. The fact that the publications can be viewed by other users of the network increases the possibilities of citation and the impact of the researcher, and is a powerful incentive to stimulate scientific collaboration.

During the training workshops, detailed instructions and advice were given so that attendees could create their personal profiles in:

— Academic platforms, such as ResearchID/Publons (https://www.researcherid.com/#rid-for-researchers), linked to Clarivate's Web of Science, and ORCID (https://orcid.org/register), a global non-profit organisation supported by member organisations' fees and linked to Scopus de Elsevier in Spain.

— Manage the Elsevier Scopus ID using the Author Feedback Wizard tool (https://service.elsevier.com/app/answers/detail/a_id/26216/supporthub/scopus/kw/id+scopus/).

— Google Scholar (https://scholar.google.com/), use of which, including creation of the researcher profile, constituted a specific workshop with the aim of all participants ending the day with their website created.

— Specialised social networks, such as Academia.edu (https://www.academia.edu/) or Researchgate (https://www.researchgate.net/).

— General social networks, such as LinkedIn (https://www.linkedin.com/), Twitter (https://twitter.com/home?lang=fr), or Facebook (www.facebook.com).

Among good practices in this area, it is essential to detail a written action plan indicating the form of signature chosen, which personal profiles are the most appropriate according to interests and needs, a regular and periodic update calendar of the information that is disseminated through them, etc.:

— Objectives pursued: disseminate projects and research, improve professional reputation, meet colleagues from other institutions, more visibility…

— Scope: what is expected to be achieved through the presence in these areas in terms of other researchers, the area of specialisation, the institution they belong to, society as a whole…

— Lines of action: how to approach the creation and maintenance of personal profiles, decide and manage the unique signature, indicate the institutional and/or professional affiliation, control and correct the way in which the researcher appears in the media and existing information networks, etc.

— Profiles and dissemination networks chosen: which ones will be created or used according to the own needs and interests of the research area. The proactive attitude is essential as well as dedicating time per week to updating the information of each researcher profile.

— Evaluation: citations received, impact factor, altmetrics from social networks, indexing in databases, etc.

3. 2. Writing academic and research papers

To master the main aspects of the preparation of scientific papers, training and good practices were given on the aspects detailed below.

3.2.1. *Tools for writing articles and thesis*

We particularly stressed the use of open source or public tools, since there is no reason to think that everyone has to buy licences. *LibreOffice* was recommended for writing texts for the following reasons:

— A stable interface that does not change depending on the size of the window, the operating system, or the version of the program. This greatly facilitates teaching (students see the same thing as the teacher) and learning (the student will always see the same thing on any computer, and even with different versions of *LibreOffice*).

— Continuous updates and read and write support for the most common formats, including *Microsoft Office* documents and *Google Docs*.

— Contains everything needed: writing, spreadsheet, mathematical formulas, layout, slides, and database.

— It is translated into 119 different languages.

— It can be used on Windows, macOS and Linux without problems and with full compatibility.

The other important part of a scientific document is the bibliography. It was explained that there are tools to manage the bibliography that greatly simplify the work of the teacher and researcher. In keeping with the above, we recommend *Zotero* for the following reasons:

— A DOI, a PMID, or reading an article on the web is all that is required to include it as a reference without having to type anything.

— Although it is only in English, it has a native application for Windows, macOS and Linux; the bibliographical database is fully compatible between the three; in addition, it can be used from any browser, without installing anything.

— It is also constantly updated and improved.

— The bibliography is not stored in encrypted form (it can always be retrieved with a text editing program).

— The *Zotero* engine has also been incorporated into *Medeley*, *EndNote*, *Papers* and others, which guarantees its continuity and fundamental nature.

— It is designed to work with *LibreOffice*, but it also works very well with *Microsoft Word*, *Google Writer*, or any document saved in RTF format.

— Virtually all journal styles (at least 10,000+) are defined in a public repository, so you just have to "select and use", with nothing else to worry about.

3.2.2. *Organisation of a scientific work*

It must follow the IMRaD structure (introduction, material and methods, results and discussion) of any scientific article. The basic style rules that will facilitate its understanding in any language were explained: short sentences to describe clear ideas, respect the rules of the International System, begin each section with the main finding, include

graphical representations and tables to lighten the text, etc. It was also indicated that the order of writing began with the tables and figures, followed by the methods and the results, to continue with the introduction and end with the discussion. This is when you have to consider writing the synopsis (which will be read by thousands of people) and, finally, the title (which will be read by millions of people). It was also shown that each journal and each scientific document has a model that should be followed, so we indicated where to get these models. Given the importance and convenience of having models, part of the final evaluation consisted of creating a model that can be used to write theses in their corresponding faculties.

3.2.3. *Good practices that will facilitate writing and formatting*

To do this, you have to know how *LibreOffice* works and what its main contributions are. Special emphasis was placed on the use of styles and the ad hoc modification of fonts, size, margins, etc. was discouraged). Visualisation of hidden characters, opening and recording of documents in standard format (ODT and PDF) as well as other compatible proprietary formats was recommended. It was indicated how to insert special characters, the tracking of changes when the writing is collaborative, the inclusion of comments outside the text, the automatic creation of the index when styles have been used, the combination of pages with different orientation in the same document, and what a header and a footer must contain to properly fulfil their purpose. It was stressed that putting all this together in one model is really convenient and advantageous. They were invited to consult and use templates already designed for *LibreOffice* and even those for *Microsoft Office* (which are 100% compatible with *LibreOffice*).

3.2.4. *Optimal bibliography management with Zotero*

They were shown that the standalone version (Windows and macOS) of the program and the web version were identical, and that the only thing that changed slightly was the appearance of the application. They were shown in detail the interface, which is very simple, comfortable and intuitive —especially in relation to the addition of new references and their organisation in folders, —the meaning of the icons to know if the element is a book, an article, a website, a thesis, etc. It was explained how to synchronise (or disconnect) the content of the *Zotero* account with the document on your computer, as well as the most common ways to add references when we have a PMID, a DOI, an ISBN, a URL, even a PDF of an Article. We also showed how to add the articles that are consulted on the web, either in *PubMed*, in *Google Scholar*, or in websites of journals/reviews and publishers (*Scopus*, *Loop*, *OUP*…). Manual entry was discouraged because it is very prone to errors, especially typographical ones.

3.2.5. *LibreOffice+Zotero, the ideal partner*

To make them aware of the advantages of using this pair of programs (and not others based on payment and encryption programs), the use of the *Zotero* tab that appears in any text document opened with *LibreOffice* was illustrated. This implies having the *Zotero* reference collection open to select the ones to be included in the text. We showed how to select the citation styles for different journals/reviews, as well as where to find the style needed from the *Zotero* public repository of styles. It was made clear that the best practice is to insert the bibliography with connection to the *Zotero* library so that simply by changing the journal style, the article bibliography is automatically reformatted. Although *Zotero* allows the inclusion of citations and references offline, this use eliminates all the advantages of combining the text with a bibliography manager. A practical case was created in which the references had to be included in the bibliographical collection in all the ways that had been taught before, and then cited in the text sometimes one by one, and at other times several at the same time. Then the reference list was created automatically and it was noted how easy it was for the list and citations to be made in *Cell* journal format and then to quickly switch to *Nature* without rewriting anything.

3.3. Statistics

The approach to good practices in the field of statistics was structured around data summary and inference techniques, the use of random procedures for the selection of individuals in research and, additionally, randomisation techniques in research that seek to compare treatments. Additionally, students were introduced to statistical procedures available to identify groups in data, encompassed in supervised and unsupervised learning and in the interest of using statistical techniques with robustness properties and simulation techniques to obtain empirical approximations to the properties of statistical procedures.

3.3.1. *Data summary techniques*

The mean and standard deviation were presented as the most commonly used data summaries to capture the location and variability of the distribution of numerical variables. When the distribution that we want to summarise is asymmetric, which happens, for example, when studying characteristics linked to the size of expenses or salaries and with characteristics related to time periods, it is more convenient to use the median and the interquartile range (difference between the 75th and 25th percentiles) in the characterisation of its distribution. Students were encouraged to routinely use percentiles to summarise distributions of values. In this regard, they were told that the joint use of the 25th, 50th and 75th percentiles is very frequent, which divide the sample into four intervals and which are the basis of the popular box plot. In the case of qualitative variables, we will summarise the observed frequencies with percentages.

3.3.2. *Inference techniques*

In most research, it will not be possible to study the entire set of individuals whose behaviour we are interested in, the population, it will only be possible to obtain measurements of a subset of these individuals, which we will call a sample. If this subset of individuals participating in the research, the sample, has been randomly selected, then from the observations collected in them, we can use inference techniques not only to approximate the population summary values but also to obtain measures of the error of estimation associated with these values. This is a property that makes statistical procedures unique when applied to randomly obtained samples, which provide us with bounds for the error we make in the estimation. The most widely used inference techniques are confidence intervals and hypothesis tests. With confidence intervals we try to enclose population summaries in intervals constructed from sample information. These ensure that those population values will be contained with a pre-set level of assurance, usually 95%. We told the students how to interpret them when they find them used in scientific literature articles or technical reports. To do this, they must identify the population summary that is associated with these intervals, which is the objective of their use, and we ask them to avoid a relatively frequent misinterpretation related to their use, related to thinking that they contain the values, in the characteristic of interest, of most of the individuals in the sample.

Regarding hypothesis contrasts, it should be mentioned that they respond to the classic research paradigm related to establishing a hypothesis about the operation of a phenomenon and deciding on the veracity of this hypothesis based on partial information, corresponding to a sample, collected on the operation of this phenomenon. We indicated that the construction of a hypothesis test requires that we know, if the hypothesis were true, the distribution corresponding to statistical summaries of the information from random samples. We provided them with information about the protocol related to its application in practice to make a decision on a hypothesis. This corresponds to comparing the statistical summary obtained from the sample with the aforementioned corresponding distribution of this summary. We did this by calculating the probability of observing, for that distribution, values as extreme or more than that observed value. If this probability is low, it is necessary to doubt the veracity of the hypothesis and reject it. The underlying reasoning is that if the observed value of the behaviour of the studied phenomenon is far from the interval in which we would expect it with high probability when the hypothesis is true, we will doubt that this hypothesis is true.

3.3.3. *Importance of randomisation*

It was explained to the students that when studies are carried out to find out the efficacy of a new treatment, for a certain disease the possibility of including in the research a control group that corresponds to the most frequently used treatment should always be considered or, if no effective treatment is available, to the administration of some type

of placebo. The usefulness of this proposal seems initially restricted to the evaluation of medical treatments, but it is much more general and can also be applied, for example, to all situations in which the evaluation of a new educational strategy is of interest. Assignment of individuals to the groups corresponding to the new treatment or to the control group, whenever possible, should be done randomly. Randomisation guarantees that, if differences appear between the two groups in the response variable, these differences can be attributed to the different treatment received and not to other factors. Randomisation is much more powerful than any strategy that we carried out aimed at balancing the two groups in characteristics that we had collected from the individuals. This is because the random assignment equalises the two groups on known and unknown factors. We illustrated the application of these procedures using scientific articles in which two of the first vaccines available for COVID-19 were evaluated for the first time.

3.3.4. *Statistical procedures with robustness properties*

We talked to students about the importance of using statistical procedures that resist the presence of outliers in the sample. It is known that many data summaries can perform aberrantly when the sample contains some outlier observation. Among them, the most frequently used data summaries, such as mean, standard deviation, correlation coefficient can be severely affected by a single observation. We explained to the students that there are statistical procedures that are known because they resist the influence of outliers, among them are those based on ranges or percentiles. We also made a brief introduction to the techniques based on the clipping of observations that allow robust properties and resistance to the presence of outliers in the sample to be given to the statistical procedures on which they are applied. These techniques are capable of identifying observations with discordant behaviours, in relation to the behaviour shown by the majority of observations, while at the same offering summaries of the data based on the majority behaviour observed in the sample.

3.3.5. *Statistical simulation*

Students were encouraged to use simulation and resampling techniques for approximating probabilities and for evaluating statistical procedures. In the first case, the use of these techniques will be of interest when the calculation of a certain probability is complex and, on the other hand, it is easy to reproduce the underlying random phenomenon with the computer. For this, obtaining independent realisations of this experiment would allow obtaining an estimate of the probability and, also, a 95% confidence interval for that unknown probability. Here we stressed again that the application of statistical techniques to samples obtained at random allows not only to give approximations to unknown values, but also to obtain limits for the error associated with those approximations. As for the second case, it will be possible to repeatedly apply a statistical procedure to data, obtained

by simulating a pre-set model, and study the behaviour of this procedure based on the values obtained in those repetitions. This allows an empirical evaluation, an alternative to mathematical evaluation, which is simple to obtain and, what is more important, within the reach of many users of statistical procedures. On the contrary, mathematical assessment usually requires significant mathematical skills and abilities, only available to those who have received specialised training in this field.

3.3.6. *Supervised and unsupervised classification*

One of the areas of statistics that has developed most in recent years is that concerning learning techniques, which are related to the identification of groups in samples of individuals. Among them is supervised learning, which corresponds to a situation in which the groups are known at the beginning of the research, but in which the diagnosis is difficult. In other words, given a new individual, it is difficult to correctly assign him to the group to which he belongs. In these cases, the strategy corresponds to having a sample of individuals in whom a collection of characteristics has been measured, which are called explanatory variables, and have a known membership group, and to apply statistical methodology that allows, given the values of new individuals in the explanatory variables, to assign them correctly to their group of origin. A large number of different methods are available for this.

An important classification for these methods corresponds to separating those that are based on probabilistic models, and therefore their operation will depend on whether these models are true for the situation in which they are applied, and those that we could call non-parametric and are usually based on the non-parametric estimation of densities and do not depend on the assumption of previous hypotheses about the models that verify the data. The other large area in which learning can be subdivided is related to unsupervised learning that responds to a need to identify groups in a situation in which they were never defined. Ideally, there are individuals in whom characteristics have been measured and we are interested in defining groups that integrate individuals who jointly take similar values in the multiple variables available. There will be two main types of approximations, those based on the assumption of models and those that are sequentially added by individuals.

3.3.7. *Statistical Computing with LibreOffice, SPSS, and R*

Statistical programs that allow us to apply them play a key role in the application of statistical techniques. Among them we have general-purpose programs, such as spreadsheets, which, although they are not specifically oriented to perform statistical analysis, incorporate many simple data analysis and representation techniques. Among the spreadsheets, we highlighted to our students the one included in *LibreOffice*. We used it to simulate the random extraction of samples from a population and to apply random

assignment of treatments to individuals in a sample. Likewise, we applied, in a simple estimation problem, confidence intervals and hypothesis tests.

Another type of software that allows the application of statistical methodology corresponds to statistical packages, generally paid, specialised in solving this type of problem. Among them, we introduced our students to *SPSS*, which is one of the most widely used statistical packages. The program is attractive because it is designed to be easy for non-professional users to use. With this program we handle several classic data sets to motivate the use of some statistical procedures that it incorporates.

Finally, we cannot fail to mention R, which is a free-to-use computer program resulting from a universal collaboration. Contrary to what was mentioned for SPSS, this program is not designed to be handled by non-professional users, except for one module, *R Commander*, which contains the simplest statistical procedures and which can be managed through menus. In any case, the main attraction of R is the possibility of programming, in its own high-level language, with the possibility of creating statistical meta-procedures, a mixture of other programmed procedures, which can be applied to different data sets in an automated way, available in files or generated by simulation, redirecting the output to different files.

3.4. Graphic curriculum vitae

In this topic, an attempt was made to integrate all the teachings received so far: bibliographical search, use of *ORCID, Publons and Google Scholar*, as well as programming in *R*, to obtain a curriculum from a researcher who is in *Google Scholar*. It was also seen how to compare the curricular evolution of two or more researchers.

The first thing that was taught was how to locate the researchers who have their CV in *Google Scholar*, and how to obtain their identifier. They were explained what it consists of and how to interpret the Hirsch index (H) to measure the quality of the scientist based on the citation of their works, since the impact factor is a measure of the quality of the journal, not of the published works (and in journals such as *Nature*, *Science*, *Cell* and other high impact journals, only 20% of the articles have a great impact, while 75% have much less relevance, as if they had been published in another journal of much less 'quality'). In addition, presenting a resume with images is surely going to be more impressive than based on text. The students installed the R *scholar* library and executed a *script* that was provided to them, in which they only had to change the IDs of the researchers. They were shown how to create the curriculum of the teacher who explained it to them, that of the course director, and they compared that of two very significant Spanish bioinformatics scientists.

4. Conclusions

Participants have received complete training on general and specialised information resources with open access programs and websites (*Google Scholar Metrics* and others).

They have also received additional training related to writing scientific articles and the best tools to do it (*LibreOffice*, *Zotero*, again all of them in the public domain). Attendees' scientific communication skills and competencies have been strengthened. Statistical skills have been highlighted, including handling the *LibreOffice* spreadsheet and the R programming language, given their enormous importance in the treatment of research data.

The attendees have acquired knowledge about bibliographical and document management, the monitoring of computer alerts and how to achieve a greater projection of their contributions in the academic and scientific fields, with which we hope that they will optimise the time and work dedicated to their teaching and research activity. They have received training on good practices when it comes to writing their own works and publications, editing them effectively and providing them with bibliographical references in accordance with international standardised styles according to the different fields of knowledge in which their activities are encompassed. More indirectly, they have also known general research means and procedures (how to cite, standards, editing and publication, revision, evaluation procedures…) based on the main international trends. This will have contributed to promoting e-learning and digital literacy in Algeria. Although the COVID-19 pandemic forced the training to take place remotely, we are convinced that, had it been face-to-face, it would have been even more beneficial for the participants and the activities could have been better supervised, since we know that some have had more difficulties and have not informed us, and so the opportunity to solve the problem has been lost (hence, perhaps, the very few evaluation exercises that were received).

We think that it would be a good to involve the libraries of Algerian universities in the training of teachers and students in some of the transversal skills to free teachers in relation to these general teachings. In this way, training could also be accessible to students, libraries would participate more actively in teaching and research, and the budget for the acquisition of bibliographical and documentary resources would be optimised.

We hope that the transversality of this module has contributed to a higher quality training and the creation of a competitive space on a global scale in the knowledge economy, thanks to which further development of the Algerian economy is expected. We are convinced that the course participants will be able to transmit the knowledge acquired to the different spheres of university institutions. But a one-off effort is not enough, and we urge the Algerian authorities to come up with a plan of continuous and permanent training of teachers to enhance their development as teachers and researchers; We support this recommendation in that the participants clearly showed their interest in the application of the lessons learned to verify its applicability in their own universities. Many aspects of the course could be addressed in much more depth (for example, bibliometric indicators of research and their use in the evaluation processes of researchers and institutions, the management of specific bibliographical databases by subject areas, or the fluent management of R and *LibreOffice*) and include new aspects, such as non-parametric alternatives

and robust approaches to estimation and inference problems, as well as the ideal discursive forms in scientific texts or the development of citation styles in *Zotero* from scratch. The overall impression of the three experts of the mission is very satisfactory and we hope that the impression of the participants was also very satisfactory.

5. References

Guides officiels de LibreOffice https://wiki.documentfoundation.org/Documentation/Publications/fr#Guides_utilisateur_officiels [access 16-12-20].

Get Started with Zotero https://libguides.stkate.edu/zoteromendeley/zotero [access 17-12-20].

Roberd A. Day, Barbra Gastel (2016) How to write and publish a scientific paper (8th ed.) ABC-CLIO, Santa Barbara.

Harris JK, Johnson KJ, Carothers BJ, Combs TB, Luke DA, Wang X. Use of reproducible research practices in public health: A survey of public health analysts. PLoS One 13 (2018) e0202447.

Ziemann M, Eren Y, El-Osta A. (2016) Gene name errors are widespread in the scientific literature. Genome Biol. 2016 Aug 23;17(1):177.

Ewen Callaway (2016) Beat it, impact factor! Publishing elite turns against controversial metric. Nature 535 (7611), 210-211.

Pochet, B. (2018). Comprendre et maîtriser la littérature scientifique. Presses agronomiques de Gembloux. https://orbi.uliege.be/handle/2268/186181.

Bibliothèques de l'Université de Montréal. (s. d.). Méthodologie de recherche d'information. Consulté 10 juin 2021, à l'adresse https://bib.umontreal.ca/guides/methodologie.

Martinolli, P., & Université de Montréal. Bibliothèques. (2018). Rechercher et exploiter la documentation. Consulté 10 juin 2021, à l'adresse https://bib.umontreal.ca/multidisciplinaire/plu6058.

Gestion de la formation: Trucs et astuces. (2020, November 26). Guides Thématiques Du Service Des Bibliothèques et Archives de l'Université de Sherbrook. Consulté 10 juin 2021, à l'adresse https://libguides.biblio.usherbrooke.ca/c.php?g=172727&p=3344343.

Girard, A. (s. d.). Guides thématiques: Plagiat, droit d'auteur, paraphrase, copier coller etc.: Plagier c'est… Bibliothèques de l'université d'Aix-Marseille. Consulté 9 décembre 2020, à l'adresse https://bu.univ-amu.libguides.com/plagiat.

Outils de citation: The Learning Portal at Ontario Colleges Library Services. (2020, septembre 24). Le Portail d'Apprentissage, Bibliothèques des Collèges de l'Ontario. Consulté 10 juin 2021, à l'adresse https://tlp-lpa.ca/recherche/outils-citation.

Bibliothèque de Polytechnique Montréal. (2020, juillet 8). Mise en valeur de la recherche. Consulté 10 juin 2021, à l'adresse https://guides.biblio.polymtl.ca/mise_en_valeur_de_la_recherche.

Martinolli, P. (2019). Gestion d'identité et de profil en ligne pour jeune chercheur universitaire. Consulté 10 juin 2021, à l'adresse https://github.com/pmartinolli/TM-SchoProMa.