



**Software for quantitative and qualitative analysis of digital content:
from statistical processing to digital methods**

Kahina Berkounel*,
1(Algeria), kahina.berkoune@univ-batna.dz

Received on:29/ 06/2025

Accepted on :12/11/2025

Published on : 30/12/2025

Abstract:

It is important to distinguish between quantitative and qualitative analysis methods. Quantitative methods involve performing a series of manipulations and processing operations on data considered as quantifiable units (whether figures or texts). Qualitative methods are based on an interpretation of a corpus, which may result in a successive coding of the collected elements. In this case, the software offered supports this interpretation (visualization of the corpus, note-taking of the successive analysis carried out) and may also offer quantitative functionalities following the coding carried out.

This presentation presents some useful digital tools when analyzing your survey data, whether in a quantitative or qualitative approach. First, however, it provides some information on freely accessible and reusable online databases. This fact sheet also introduces the concept of data science, the associated interfaces/programming languages, and the working methodologies they involve. With an explanation of modern digital applications in the field of digital content analysis.

Keywords: Digital tools ; Analysis software ; Statistical processing ; Digital methods ; Digital content.

* *Corresponding author*



1. INTRODUCTION

A large part of doctoral research will consist of one or more phases of analyzing the information you collect. The terms "data," "database," or "empirical elements" are used to refer to these resources that serve as a basis for answering the research question(s) you are asking.

Whether researchers use qualitative or quantitative methods, several digital tools are available to analyze this data, but also, first and foremost, to access databases. This chapter will first discuss access to so-called open databases. Various software programs used to analyze your data, whether in a qualitative or quantitative approach, will then be described.

Access to data for all and their free reuse is part of a general open movement that has developed with the deployment of digital technologies and the use of the Internet and the web. The term open data refers more specifically to the opening of data from public administrations or data considered to be of general interest. Openness then consists of being able to access this data in open formats (non-proprietary), but also of being able to reuse and redistribute it. Within the research community, open access, translated as free access or open access, is part of this open movement. In this case, the data represents the scientific articles contained in bibliographic databases. There are also databases grouping together different data sets (for example, data relating to the fields of communication and information with their distinct sub-specialties). All of these elements can constitute resources for analysis corpus freely accessible online. Depending on the nature of the research (thematic, empirical data) researchers can consult different data portals referencing open data databases.

Below are some examples of freely accessible and reusable data (the open data movement)(NOBLE & SMITH, 2015, pp. 34-35) :

- For data from public administrations, more and more countries or national or supranational institutions are offering open data portals. The UNdata portal, for example, brings together all the databases of supranational organizations (UN,

WHO, FAO, etc.). The opendatasoft website lists several thousand open data portals worldwide, particularly in the Global South ;

When downloading datasets, be careful about the format in which you download them. Open data is normally offered in open formats (e.g., .csv, .json, .txt, .md), meaning that you can read and manipulate it with free and open source software (not proprietary software). Also note the download date and the author of the database (if indicated), which you will need to cite in your work. Attribution is an essential element in an open research approach.

- Please also note that several portals offer shared (and not open) data. In this case, the data is not freely accessible and downloadable. You must request access to this data and sign a user agreement stating the conditions of use of this data. The Quetelet network in SHS provides access to public national statistics data (major surveys, censuses, etc.). The BeQuali platform provides access to data from qualitative methods (interviews, observations, focus groups).

2. Analysis of data from qualitative and quantitative methods

2.1 Research corpus

Once your corpus is complete, you'll begin your analysis. Different software programs are available to help you do this, depending on the methods you want to use. Therefore, it's important to first have in mind the research protocol you want to follow to answer your research questions.

Be careful, software is not a miracle tool. Without a clear research plan/specification, using software will become a murky and often "bottomless" exploration!

Subsequently, you will find some software used for processing so-called quantitative data as well as tools to consolidate your qualitative analysis. It is important to distinguish between quantitative and qualitative analysis methods. Quantitative methods involve performing a series of manipulations and processing operations on data considered as quantifiable units (whether figures or texts). Qualitative methods are based on an interpretation of a corpus, which may result in a successive coding of the collected elements. In this case, the software offered supports this interpretation (visualization of the corpus, note-taking of the successive analysis carried out) and may also offer quantitative functionalities following the coding carried out (DESCAMPS, 2005, p. 210).

2.2 Statistical analysis software

If your corpus consists of data organized in the form of a data table containing numerical information (.csv, .xls format, etc.), you will certainly want to perform a set of statistical processing on this data using different methods (linear regression analysis, factor analysis, multidimensional analysis, etc.). For this, different software programs are available that allow you to perform statistical analyses and visualize your results. The best known are, for example, Statistica, Stata or SPSS. These software programs also allow you to work on textual data, but other tools have been specifically developed in the humanities and social sciences for the statistical processing of text corpora. This is the case, for example, of Alceste, iRamuteq or Prospero (MOSCAROLA, 2019, p. 305).

These computer tools have been developed since the 1990s, but with the ever-increasing amount of so-called natively digital data coming from the Internet and the Web, new so-called "digital" methods have been developed to process and visualize these masses of heterogeneous data. These methods are grouped under the name of data sciences, they are at the crossroads between statistics and computer science. The approach is indeed more global than in statistics: it is no longer just a question of analyzing the data, but also of thinking about their collection, their formatting and "cleaning" in order to then be able to carry out the statistical processing as such. From a mathematical point of view, specific methods are being developed to analyze this large amount of data; this is what is grouped under the term machine learning or the general term artificial intelligence. Although these data sciences are mainly used in STM, new research streams in SHS are using these methodologies. These discussions are particularly at the heart of communities grouped under the name of Digital Humanities, Digital Methods (VERA & QUAN, 2019).

3. Quantitative analysis software: from statistical processing to numerical methods

3.1 Programming languages and data science

Without going into detail about these numerical methods and their theoretical mathematical substrate, we can cite some software based on free and open source programming languages used today in both STM and SHS. The R and Python programming languages are commonly used today and are accompanied by software development and modules to facilitate their use or offer specific treatments.

- The R language, for example, is associated with the R studio software, which is a working environment to facilitate the production of R scripts and the readability of

the results. Other software such as iRamuteq or RTemis are based on R, among other things, for the manipulation and analysis of textual data ;

- The Python language is widely used in STM but also in SHS. In addition to statistical analysis modules (also referred to as libraries), it also allows you to work on large databases (link with the SQL language), but also to apply a set of algorithms (sequence of logical instructions understandable by a machine) derived from machine learning. Python can be used on the command line or via different applications, notably the Jupyter Notebook (VERA & QUAN, 2019).

Some specificities of programming languages and organization of free and open source communities.

Whether it's Python or R, both languages were developed using a free and open source approach. The organizational arrangements are different from proprietary software development. Anyone can contribute to the development of the language, but also propose specific libraries/modules containing a set of basic scripts that will later become easily usable functions. It's important to know the practices of your research community (do they use R, Python, or another language?) to be able to benefit from the work already done by other researchers, but also to ask for help (via forums or at conferences)[♦]

Other software can be used for specific analyses, such as network analysis and mapping. This is the case, for example, with Gephi software (also free). It allows you to build graphs and perform a series of statistical treatments from a data set. However, be careful not to fall into the trap of the "wow" effect of visualizations. This application offers an interface where it is easy to click on buttons, apply different treatments and obtain visualizations without understanding the successive steps carried out on the data[♦]

Keep in mind that it is important to take a critical look at these methodologies (avoid the "black box" effect) and the influence of visualizations on the interpretation that one makes of one's data corpus.

Other software is available for managing and analyzing data, as well as easily obtaining visualizations. This is the case with Data Studio or Tableau (mainly for visualizations), or the CorText platform, which offers a set of text analysis processing

[♦] Cassandra: online logbook platform adapted to a grounded theory approach.

[♦] Gephi: software for network analysis and mapping.

tools. These tools are easier to use and intuitive in terms of interface, which is certainly a good thing. It's not a question of turning you into an engineer. Nevertheless, use your critical eye as a social sciences and humanities researcher to reflect on the use of such socio-technical devices*.

3.2 Qualitative and mixed analysis software

Several software programs can also be used to consolidate your qualitative research approach. This isn't just about performing quantitative processing on your empirical (textual) elements, although most of these programs also offer quantitative analysis features. The goal is to assist in the interpretation of your corpus. These programs allow you to code and annotate documents, images, and even web pages. They also offer search options and tracking of all actions taken to ensure the transparency of the analysis process. These programs can, for example, be used in a grounded theory approach, in which different successive coding stages are proposed, or even in the use of mixed methods (VERA & QUAN, 2019).

Below are descriptions of some programs to give you an idea of the available features. To discover other software, please consult the Wikipedia page entitled Computer-assisted qualitative data analysis software.

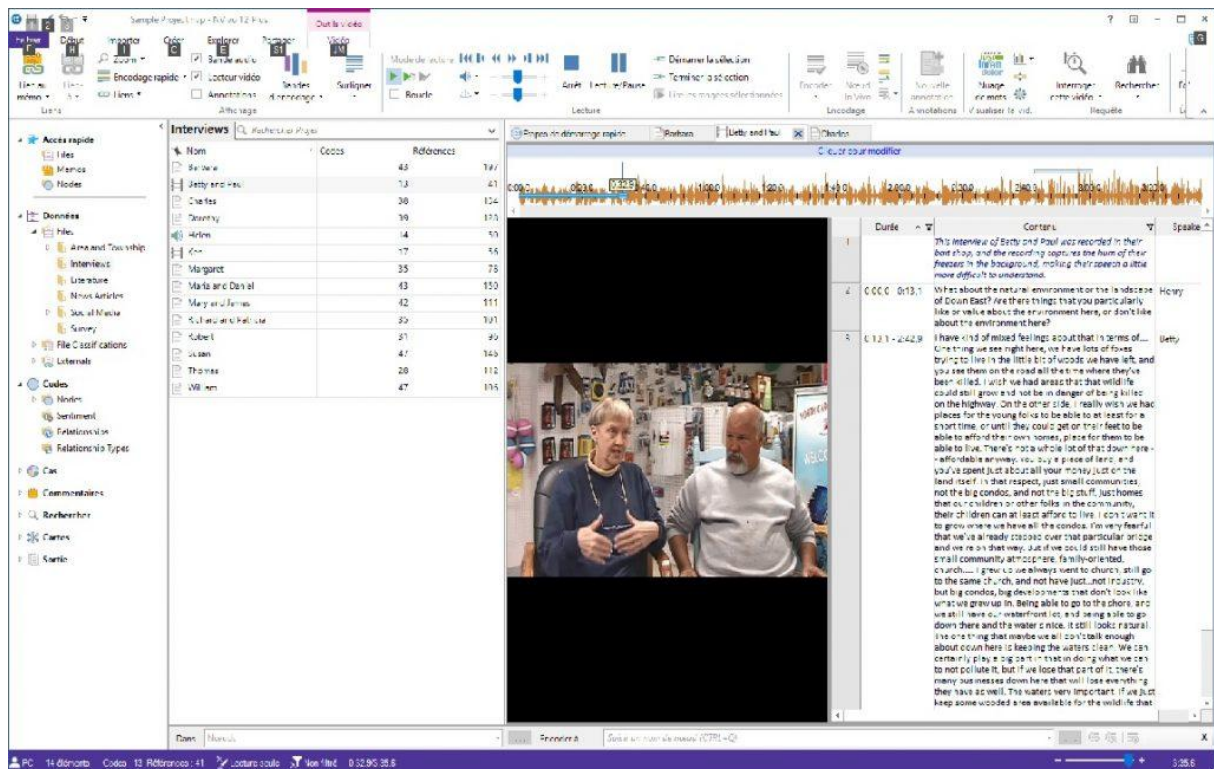
- **NVivo** : is a proprietary software that allows you to tag elements of your corpus, organize it into different subcategories, search on the tags you have made, take notes on each defined tag, etc. Quantitative options are also offered (keyword clouds, indication of the most used words, etc.). NVivo also allows you to take a screenshot of web pages (N-catpure) and then analyze them in the software, users can export their data in various formats, compatible with applications such as Microsoft Excel, Microsoft Word, IBM SPSS Statistics, EndNote, Microsoft OneNote, SurveyMonkey, and Evernote. Easy outsourcing of video transcription to TranscribeMe*.

* Data Studio or Tableau: online data analysis and visualization tools.

* NVivo: proprietary software to assist the qualitative analysis process with the provision of quantitative text analysis functionalities.

Please note, however, that this software is proprietary; you will then need to use this software to conduct all your research and pay a license fee. It will not be possible to open .nvpx/.nvex file formats in other software.

Screenshot N°1: Reflects the transcript of a scientific interview

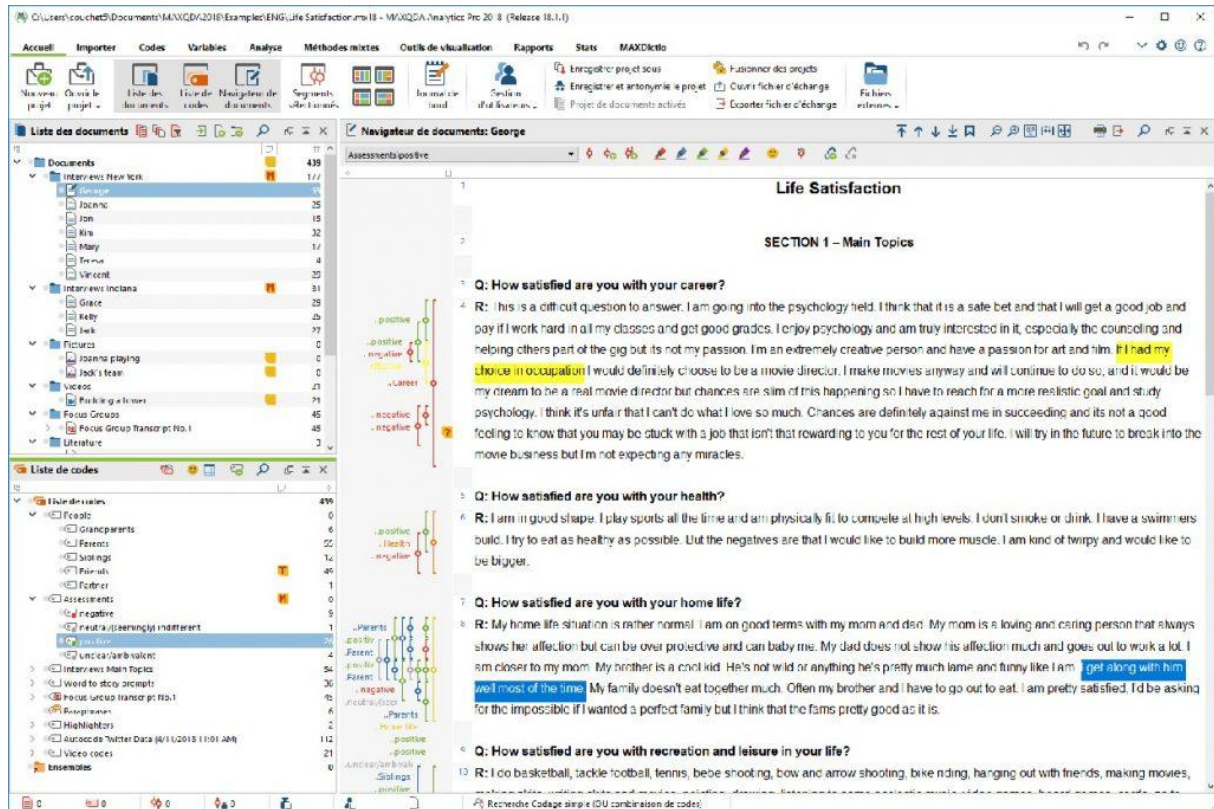


Source : a transcribed interview with Betty and Paul, with NVivo

- MAXQDA, paid software for Windows or Mac, 40-day free trial, first released in 1995. All versions of MAXQDA (MAXQDA Standard, MAXQDA Plus, MAXQDA Analytics Pro) offer tools for organizing and analyzing qualitative and mixed data. A variety of files: text, audio, images, video, and bibliographic files, as well as survey data, tweets, or focus group transcripts are managed. Data can be analyzed using codes and memos. MAXQDA's visualization and export functions (DOCX, XLSX, HTML) facilitate presentations. Several modules available only in certain versions: MAXDictio, Stats, TeamCloud, add additional capabilities.

Recently released, an online version to promote teamwork and data backup. Humanum offers a floating license for the use of MAXQDA Analytics Pro.

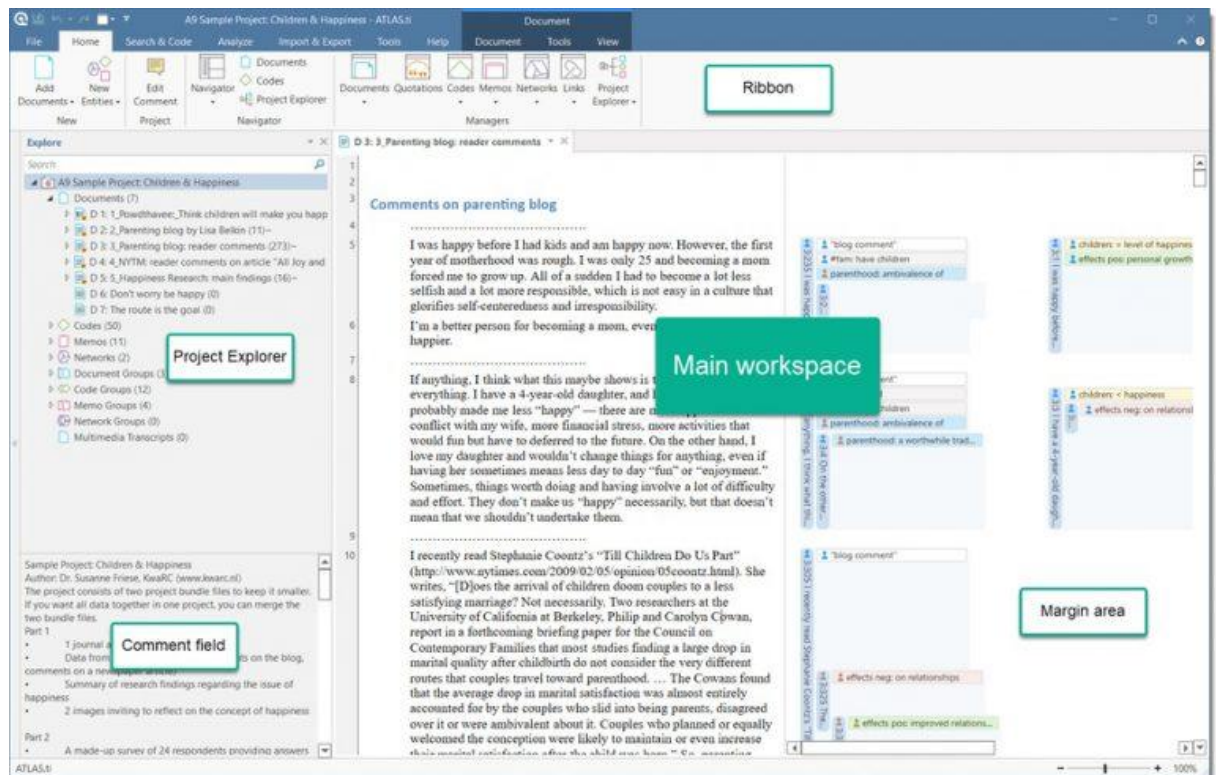
Screenshot N°2: transcription of interview (content)



Source : MAXQDA: Transcription and Annotation of an Interview

- ATLAS.ti offers four products: ATLAS.ti Windows, ATLAS.ti Mac, ATLAS.ti Mobile, and ATLAS.ti Web. 5-day free trial. The software helps researchers discover and analyze complex phenomena hidden in unstructured data such as text, multimedia, and geotagged data. The program provides tools that allow the user to locate, code, and annotate primary data, weight and assess its importance, and visualize the often complex relationships between documents. An entire project can be exported in the XML-derived QDPX format, which can promote interoperability and data storage in dedicated repositories. SPSS (IBM Statistical Analysis) format is also supported.

Screenshot N°3: text of the interview



Source : Main workspace

- Sonal is a free interview transcription software that offers qualitative analysis features. It allows you to easily transcribe, categorize, and analyze interviews from WAV audio files. Each successively transcribed portion is categorized by assigning them a specific color. Statistical analysis is also possible. Windows only.

Screenshot N°4: Software of Sonal's



Source : Sonal's interface

- The Rotterdam Exchange Format Initiative (REFI) aims to promote the development of the REFI-QDA standard for exchanging acquired data (transcription, indexing, categorization) between qualitative analysis software. It was at the 2016 conference "Reflecting on the Future of QDA Software" (Erasmus University Rotterdam, Netherlands) that the directors of eight software vendors expressed their desire to cooperate to create such a standard. These companies were: ATLAS.ti, Dedoose, f4analyse, MAXQDA, NVivo, QDAMiner, Quirkos, and Transana. The REFI format was created shortly thereafter. An expert group made up of users, trainers, and standardization specialists oversees its evolution. A class diagram that can lead to various implementations resulted from their work.
- QDA Miner and Dedoose are available and position themselves as software or applications suitable for mixed methods. Dedoose offers the possibility of collaborating on the same project. Unlike other software programs where you pay a license to access it, Dedoose offers paid subscriptions (monthly or annually) with separate pricing depending on the features. The use of these different software

programs must be accompanied, at all costs, by documentation of the various stages of your investigation, which may itself be supported by digital tools[♥].

Qualitative analysis software designed for mixed methods research. It is both user-friendly and easy to use, allowing the coding, annotation, mining, and analysis of small and large quantities of documents and images. QDA Miner can analyze transcripts of individual and focus group interviews, documents, reports, journal articles, books, as well as images, photographs, and any other type of visual material. Its seamless integration with SimStat, a statistical analysis tool, and WordStat, a content analysis and text mining module, gives you unprecedented flexibility to analyze text and highlight the relationships between its content and structured information, including numerical and categorical data. A free version of QDA Lite is available for download.

- The Cassandra platform (free software) positions itself as a logbook that allows you to write reports as you go along, schematize the conceptualizations resulting from your analysis and work in a team*.
- Kwalitan. The software, designed for Windows, provides a variety of tools for storing and analyzing text documents, images, and videos. Text encoding, text fragment searching, categorization codes, and memo writing.
- Quirkos. The software provides a graphical interface in which the analysis nodes or themes are represented by bubbles. The software is designed for beginners in qualitative analysis.
- Transana. Three products are offered: Transana Basic, dedicated to the transcription of audio and video documents; Transana Professional, for processing complex data, analyzing multimedia documents, texts and images; and Transana Multiuser, with the possibilities of Professional accessible to groups analyzing the same data.
- Sphinx Software. Comprehensive, powerful, and innovative data analysis and research solutions for teaching and research, Jean Moscarola.
- Taguette is easy to use, this simple, free software comes in two versions.

1/The "Centralized Server" version, hosted by Taguette or on your servers, allows a project manager to log in, create a project, and import document sets. These documents can then

[♥] QDA miner or Dedoose: software and online application for qualitative and mixed analysis.

*Cassandra: online logbook platform adapted to a grounded theory approach.

be tagged using a set of keywords specific to each project. Participants with specified rights can intervene to facilitate indexing or simply view the analysis results.

2/ The "Local Server - Single User" version runs on a single computer. It is also offered free of charge. After installing the software on Windows, Mac, or Linux, a local web interface is launched. The same features as before are available. Data is saved in SQLite3 format.

- Celluloid is a web platform dedicated to annotation and commentary, alone or in teams, on digital audiovisual documents (films, series, shows, or your own videos). Developed by the Canevas d'Huma-Num consortium, the freely distributed software is based on a PeerTube instance.

4. CONCLUSION

In conclusion, the use of data analysis software or access to databases will be essential activity in scientific research. However, do not forget that a good understanding of how these software programs work (the processing they perform) as well as a critical look at the constitution of the data sets are necessary to justify their use and guarantee the quality of doctoral project. The "digital" is in itself a subject of reflection in Information and Communication Sciences, and it leads us to question our own research practices and, more generally, the contemporary methods of producing and disseminating knowledge.

It is important to note that software and platforms do not provide sources, annotations, or the study itself. This task of creating meaning is the sole responsibility of specially trained researchers and analysts. Excellent preparation of your groups, conduct of your surveys, annotation of your documents, and preservation of your qualitative data.

5. References

DESCAMPS, F. (2005). *L'historien, l'archiviste et le magnétophone : De la constitution de la source orale à son exploitation*. Paris : Institut de la gestion publique et du développement économique.

MOSCAROLA, J. (2019, 04 7). Méthodes d'enquête et d'analyse de données à l'usage des Managers-Chercheurs. *CAIRN.Info* , pp. 295 - 316.

NOBLE, H., & SMITH, J. (2015, April). Issues of validity and reliability in qualitative research. *BMJ Journal* , 18 (2), pp. 34-35.

VERA, G., & QUAN, N. H. (2019). *Comprendre la recherche. Méthodologie, Unité de soutien SRAP du Québec*. Consulté le 06 1, 2025, sur <https://www.comprendrelarecherche.ca/>