NARRATIVE REVIEW
Volume 6 (2025), No. 1, pp. 27-46
DOI: 10.15367/ch.v6i1.770

AI-Empowered Qualitative Data Analysis: Training Future Public Health Researchers

CHRISTOPHER W. WHELDON, PhD, MEd, MSPH[1], RYAN W. MCKEE, PhD, MEd, MS[1]
[1]Department of Social and Behavioral Sciences, College of Public Health, Temple University, Philadelphia, PA, USA

Correspondence: chris.wheldon@temple.edu (Christopher W. Wheldon)

Advances in generative artificial intelligence (AI) are transforming qualitative data analysis in public health research. By enhancing the efficiency and scalability of data analyses, AI-empowered tools allow researchers to handle larger datasets (i.e. interview transcripts) and uncover deeper insights than traditional methods of manual or computer assisted analysis. However, while AI tools like NVivo, Atlas.ti, and MAXQDA enhance efficiency by automating coding and theme identification, they are still limited in their ability to replicate the nuanced interpretative role of human researchers. In addition, the adoption of these technologies necessitates careful consideration of ethical implications. Educating public health students on the use and limitations of AI tools, alongside traditional qualitative methods, will prepare them for the evolving research landscape. In this article, we describe current AI-empowered tools for qualitative data analysis, review their strengths and weaknesses, and outline learning objectives with sample lesson plans to prepare public health students to critically employ these tools in their own research endeavors.

***Keywords**: Generative AI, Qualitative Data Analysis, Public Health Education, CAQDAS, Ethical Considerations, AI Tools*

Background

Qualitative methods have become an integral part of public health research,[1] allowing researchers and practitioners to achieve a wide range of objectives (See Stickley and Colleagues for a recent overview).[2] Qualitative data collection methods can include interviews, focus groups, participant observation, and textual analysis. Qualitative designs are flexible and iterative, and help researchers understand the meaning behind people's health decisions and behaviors, as well as the symbolic meaning behind participants' words and actions. In addition, qualitative researchers attempt to understand the structural barriers that lead to negative health outcomes, and evaluate interventions designed to promote health and wellness through the analysis of rich data sources.[3] Qualitative study designs have gained increasing acceptance as stand-alone methodologies, driven by a growing emphasis on cross-cultural health and health equity.[4,5]

Qualitative data analysis encompasses a diverse set of methods designed to systematically interpret and make meaning from data (e.g., textual, audio, visual data).1 While approaches vary depending on the study's epistemological framework, most qualitative analyses involve an iterative process of data immersion, data reduction through qualitative coding, and making meaning from the data.[6] Common analytic approaches include thematic analysis (i.e., methods to identify themes from the data), grounded theory (i.e., identify patterns and develop new theories), framework analysis (i.e., apply existing theories to reduce data into meaningful chunks), and content analysis (i.e., categorizing data by counting the frequency of words, phrases, or concepts), each offering distinct strategies for identifying and interpreting patterns within the data. Data analysis is usually conducted in teams, where discussions and reflexivity play a crucial role in ensuring rigor, reducing bias, and enhancing the credibility of findings. Since the 1980s, researchers have utilized computer assisted qualitative data analysis software (CAQDAS), significantly enhancing the efficiency of qualitative data analysis (e.g., coding data and identifying themes).[7,8] More recently, the rise of inexpensive and accessible generative artificial intelligence (AI) tools has presented the opportunity to transform many industries and academic fields. The disciplines of public health and qualitative research are equally poised to gain from these groundbreaking technologies.[9,10] However, these advancements also raise important questions regarding the effectiveness of AI tools in research and the ethical considerations surrounding their use.[9–12]

In this article, we provide a brief history of CAQDAS, explore how generative AI can support qualitative researchers, assess the strengths and limitations of select widely available AI tools for data analysis, and discuss the ethical considerations of using generative AI in qualitative research. Finally, we will share and describe a lesson plan aimed at helping public health students learn about and use generative AI tools for qualitative data analysis.

The Role of Generative AI in Qualitative Data Analysis

The widespread availability of generative artificial intelligence tools is opening a new world of possibilities for qualitative researchers.[9,10] By leveraging advanced algorithms, generative AI can process and analyze vast amounts of qualitative data with unprecedented speed and efficiency.[12] Generative AI can quickly identify patterns, themes, sentiments, and trends in data. It can also be used to develop coding schemes (a set of codes used to help categorize information) and, potentially, reduce biases inherent in human analysis.[13] By automating the coding process, generative AI tools can expedite tasks that traditionally required extensive manual effort, such as categorizing and organizing open-ended responses, making qualitative research more efficient and scalable.

Generative AI can transcribe interviews, or even provide real-time analysis of audio and video interviews.[14] These tools can be fitted to traditional coding methodologies like deductive (top-down coding organized by pre-existing themes) and inductive coding (bottom-up coding, in which new codes are developed from data), enhancing the objectivity and scalability of qualitative research.[15] Generative AI can support complex research frameworks such as grounded theory, which involves multiple stages of coding and the identification of connections between codes.[16] Further, some researchers even use generative AI to create synthetic data that can be used for training students or in situations where real data is difficult to obtain or sensitive in nature.[17] By leveraging these capabilities, generative AI can complement human researchers, allowing them to focus on the more nuanced aspects of qualitative analyses, such as interpreting results, formulating conclusions, and determining broader implications.

In table 1, we provide specific examples of how AI-empowered tools can be used at each step of the thematic data analysis process, as described by Braun and Clarke (2006). This includes automatic transcription of audio files, generating word counts and concept clouds, auto-coding transcripts, generating summative themes, cross-referencing themes with coded segments, and creating tables and figures to illustrate findings. These capabilities can enhance the efficiency and credibility (i.e., validity of conclusions drawn from the raw data) of qualitative research, enabling researchers to manage larger datasets and gain deeper insights.[18]

But while generative AI offers exciting possibilities for advancing qualitative research in public health, it also presents challenges that need to be addressed to fully realize its potential benefits. The key will be to leverage AI as a tool to complement human expertise, not replace it. As AI becomes more integrated into public health research, there will be a need for increased education on its capabilities and limitations, as well as the development of rigorous regulations to guide its use.

Table 1.
Examples of AI Use in the Thematic Data Analysis

Steps in Qualitative Data Analysis	Examples Using AI-Empowered Tools
Data familiarization: Preliminary ideas start to emerge through repeated exposure to the data	Automatically transcribe audio files in real time; Generate word counts and concept clouds; Generate and review AI-derived summary codes without specific prompts; Ask for narrative summarizes of the data; Summarize, edit, and synthesize analytic memos in real time.
Generating initial codes: Initial process to reduce textual data into smaller meaningful coded segments	Auto code transcripts; Prompt more specific coding with a detailed codebook; Specify coding approach (e.g., sentiment, in-vivo codes); Interact with data using conversational AI features.
Searching for themes: categorizing codes into meaning summative themes	Automatically collate data relevant to each code; Generate summative themes with associated text extracts; Explore coded segments using different layouts and visualizations; Ask conversational AI to list patterns, themes, and key summaries.
Reviewing themes: Mapping themes back onto original coded extracts	Cross-reference themes to coded segments and specific quotes from the text; Validate themes with AI-driven data checking.
Defining and naming themes: Edit and refine themes producing clear definitions of each theme	Generate automatic code and document summaries; Cross-reference theme descriptions with published research; Generate alternative definitions and names.
Reporting: Creating tables and figures that provide evidence for how themes relate to codes and to extracted quotes from the original data	Generate summary tables and figures to illustrate findings.

AI and CAQDAS

Commercialization of CAQDAS tools increased in the following decades, yet integration of these software products into qualitative research methods pedagogy has been slow despite the availability of a wide variety of user-friendly packages.[7,19] Many CAQDAS software packages have been available for decades and have continued to add features (Table 2). Packages such as NVivo, ATLAS.ti, and MAXQDA were originally developed in the 1990s, but continue to adapt. These packages now have web-based applications joining newer web-based only applications (e.g., Dedoose). The rapid emergence of generative AI technologies represents the latest advancement impacting CAQDAS. Established software packages have begun to incorporate AI features into existing web-based and desktop applications.

Table 2.
Examples of AI-Empowered Tools for Qualitative Data Analysis as of the Writing of this Article

Name	Type	Release Date	Platform	Student License	Free Trial
ATLAS.ti	QDAS	1993	Windows Mac Web-based	Yes	Yes
Chat-GPT	General purpose AI	2020	Web-based	No	Yes
Claude AI	General purpose AI	2023	Web-based	No	Yes
CoLoop	AI specific text analysis	2022	Web-based	No	No
MAXQDA	QDAS	1989	Windows Mac Web-based	Yes	Yes
Nvivo	QDAS	1981	Windows Mac Web-based	Yes	Yes

Examples of Current AI-Empowered QDA Tools

We conducted a cursory examination of existing AI-empowered QDA tools; however, the availability of these tools and their specific features are rapidly changing. At the time of this writing, Atlas.ti has deployed features on their cloud-based application that can, if desired by the researcher, fully automate the coding process. This tool can also suggest codes during manual coding and allow the user to ask a chatbot questions about specific documents (e.g., "What are the main themes from this transcript?"). The desktop version has these features in addition to what is called "AI summaries," which provide a concise description of the text documents. All these features are in beta testing and use OpenAI's GPT models.

MAXQDA has also added features powered by OpenAI to their desktop application. Some features are like those available in Atlas.ti, such as suggested codes. However, MAXQDA also offers the ability to suggest subcodes to manually coded data. The software can also automatically transcribe audio and video in over 45 languages. The rest of their AI features are focused on summarizing, explaining, and paraphrasing coded segments.

NVivo, also leveraging the power of OpenAI's ChatGPT, offers a similar set of AI tools to process and analyze qualitative data. As with the previously mentioned software packages, NVivo's Autocode Wizard enables the user to analyze text documents (e.g., interview transcripts) by applying codes. However, Nvivo's AI tool can learn coding strategies from previously analyzed documents that were manually coded (referred to as "pattern coding"). This enables the user to create and apply a codebook using traditional methods and then have Nvivo's tool apply that codebook to new documents. Users can also apply sentiment analyses and control how much of the text segments to code, such as sentences, paragraphs, or whole transcripts. The results can then be displayed as a coding hierarchy or a matrix.

New software packages developed with the sole focus of leveraging AI for CAQDAS are also being produced. In contrast to traditional QDAS that have incorporated AI powered features, CoLoop is designed specially to leverage AI technologies for qualitative data analysis. Developed by Genei.io, the tool is described as an "AI Copilot," that allows the researcher to ask questions of the data through a chat function. CoLoop can also transcribe audio and video files, analyze text data, and output results in a table format. The tool can also provide answers to prompts, summaries for individual interviews, and thematic summaries across multiple transcripts.

A third category of software utilized for CAQDAS includes more general-purpose AI systems such as ChatGPT and Claude. These systems are not solely designed for qualitative data analysis but are being used to support a range of research tasks. For instance, ChatGPT 4.0, with its advanced natural language processing capabilities, can assist researchers in summarizing the content of interview transcripts, abstracting meaning from the text, and generating codebooks. Similarly, Claude, another generative language model, is geared towards understanding and creating human-like text responses. It can be used in qualitative research to simulate interactions, analyze sentiment, and assist in the thematic analysis of large datasets. Both ChatGPT and Claude can be integrated into academic and research-focused products to tailor AI capabilities needed to enhance the efficiency of qualitative analyses. These integrations demonstrate the potential of generative AI tools to significantly impact the field of qualitative research by providing robust, scalable solutions for data interpretation.

Strengths and Limitations of Current Tools

The integration of AI into CAQDAS offers significant strengths that enhance the research process. Efficiency and speed are among the most evident benefits, as AI tools like ChatGPT and Claude can quickly process large volumes of data, reducing the time researchers spend on manual coding. This increased efficiency enables researchers to handle larger datasets than traditionally possible in qualitative research, addressing scalability issues and allowing for the study of broader sample sizes.

Additionally, AI-driven CAQDAS can enhance insight and improve accuracy. These tools may help uncover patterns and connections that human coders may overlook, thus providing a novel method for member checking and triangulation. This capability not only bolsters the reliability of research findings but also expands the analytical possibilities of text data (e.g., open coding of electronic medical records).[20]

Despite these strengths, the integration of AI into CAQDAS is not without limitations. Many AI features in tools like Atlas.ti, MAXQDA, and CoLoop are still in beta testing, which may pose reliability issues. The rapid incorporation of AI functionalities often means these features are underdeveloped and may not meet all research needs immediately.

A gap exists between the current capabilities of AI-driven CAQDAS and more advanced qualitative coding methods. Qualitative coding methods are complex and varied. Saldaña and colleagues described more than 32 coding methods.[6] For instance, methodologies such as constant comparative analysis by document groups or emotion coding require nuanced coding methods. For example, understanding how participants describe vaccine hesitancy requires more than identifying keywords—it involves interpreting shifts in reasoning, emotional tone, and underlying social influences. A researcher often wants to systematically apply different coding methods to a dataset, synthesize codes into broader categories and themes, and engage in iterative interpretation of the data. Existing AI-driven CAQDAS tools cannot, at this time, fully replicate these varied systematic approaches.

Currently, AI-assisted CAQDAS tools are limited in their ability to extend beyond text analysis. Qualitative data can include the coding and analysis of images and other visual mediums. For example, qualitative research has been used to analyze internet memes to surveil trends public perceptions of pandemics (e.g., COVID-19).[21] These methods are commonly used to understand contexts contained in images and video.[22] They are also central to the evaluation of public health programs employing novel methods like PhotoVoice.[23] However, AI image processing is rapidly advancing, and innovative qualitative research applications will have the possibility to yield tremendous benefits (e.g., using AI-driven software to identify pain and emotions from facial expressions).[24]

There are also limitations inherent to existing large language models that should be considered, such as their ability to understand subcultural, regional and generational slang (i.e., ubiquitous informal language that is highly culturally context dependent).[25] Interpretive methods, such as ethnography, are highly focused on understanding emic knowledge (i.e., descriptions of a culture from the point of view of those who use a common language and have a common lived experience), which require an analysis of subcultures that might be marginalized, and whose discourse may be outside the purview of the data corpus used to train the large language models. The potential inability of large language models to understand the use of slang, and the likelihood of bias, is likely greater for non-English languages. This is particularly concerning for public health research that is often focused on global health issues (e.g., HIV prevention in sub-Saharan Africa) where there are already language barriers that need to be addressed using qualitative methods (e.g., HIV prevention in African countries).[26] Further, the lack of human emotion and cultural nuance (emic perspectives) in AI analysis could result in misinterpretations of qualitative data, particularly when analyzing slang or culturally specific discourse.

This also raises a larger issue related to potential bias in large language models. Their use introduces the risk of "inheriting" the biases of those who develop and program the software, or biases that are present in the training data.[27] Such inherent bias could skew research outcomes (particularly in health disparities research involving marginalized populations). While human coders are not immune to bias, reliance on large language models like those developed by OpenAI, which are used in commercial software such as Atlas.ti, MAXQDA, and CoLoop, could impact analysis and findings, particularly in sensitive research areas. It is crucial for researchers to remain vigilant and critically evaluate AI-generated insights, ensuring that the final interpretations accurately reflect the data's context and inherent meanings. Such human oversight is of scientific importance and has ethical implications as noted below.

Ethical Considerations

The integration of AI technologies into qualitative data analysis necessitates unique ethical considerations. AI-powered tools like Atlas.ti, MAXQDA, and NVivo store and process data through cloud computing systems. This raises significant concerns about data security and confidentiality. Additionally, general-purpose AI systems such as ChatGPT and Claude may use the data to train their language models, posing additional risks. It is crucial for researchers to understand these implications and take appropriate measures to protect participant data in compliance with ethical standards. The Belmont Principles (i.e., Respect for Persons, Beneficence, and Justice) provide a useful framework to consider some of these issues.

Respect for Persons requires that participants are fully informed about the use of AI in data analysis and the potential risks. This includes transparency about how AI will process and interpret their data, with a focus on ensuring participants comprehend the implications of AI involvement. Given that not all participants may be familiar with AI technologies, researchers must provide clear, simplified explanations. Additionally, privacy and confidentiality must be rigorously maintained. Again, all the AI

CAQDAS tools reviewed in this article utilize cloud computing to store and process data. This inherently involves third party companies. In many cases, the ways in which data are protected, accessed, and used by these companies are not explicitly stated. For example, Atlas.ti indicates that they "...have concluded a Data Processing Agreement, including the Standard Contractual Clauses (SCC) with OpenAI" to ensure "the safety of your personal data," users should be aware that this data is transferred through an application program interface (API) to run the OpenAI models. Furthermore, Atlas.ti "may also send select portions of content to third-party contractors (subject to confidentiality and security obligations) for data annotation and safety purposes."

Participants should also be made aware if their data will be used to train AI models. This is essential to ensure that the respect for persons criteria is met. Researchers must approach the use of AI-assisted CAQDAS with caution, ensuring that participants are fully aware of the risks of, and provide explicit consent for, their data to be used in training AI models. This transparency is vital to maintain the integrity of the research and respect participants' rights.

Under the principle of Beneficence, researchers must conduct a thorough risk-benefit analysis to ensure that the integration of AI does not introduce risks that outweigh the benefits. In qualitative research, a main concern is often the breach of confidentiality. Qualitative methods often involve the collection and analysis of sensitive data, which in public health research may include personal health information, sensitive participant narratives (e.g., sexual health, substance use), and other personal information. Data editing and the removal of identifiers are crucial steps in safeguarding sensitive information. Researchers should be cautious about using cloud-based AI transcription services for sensitive data unless they can guarantee that audio files are thoroughly de-identified. Alternatively, using local, or "offline", AI-based transcription tools (e.g., noScribe) can mitigate the risks associated with cloud storage. This approach helps prevent unauthorized access and helps ensure that participant confidentiality is maintained. Additionally, researchers and IRBs must consider the unique ways AI threatens confidentiality (e.g., voice recognition, triangulation of multiple datapoints to re-identify a participant), which could inadvertently lead to the identification of participants. Evaluating and mitigating these risks is essential to protecting participants' privacy.

The principle of Justice emphasizes that those who bear the burdens of research should receive the benefits in equal measure to the burdens. AI algorithms are only as good as the data they are trained on. If the training data lacks diversity or underrepresents certain population subgroups, the large language models will not be able to accurately analyze or reflect the experiences of these populations. The potential for AI to make errors or misinterpret data of vulnerable populations is an ethical consideration. Thus, it is crucial to implement robust validation protocols that can be used to check AI outputs and mitigate the risk of harm caused by erroneous data interpretation. Using existing qualitative data validation techniques such as member checking, triangulation, audit trails, and peer debriefing can significantly enhance the ethical principle of justice in AI-driven qualitative data analysis. These techniques can be applied to ensure a more equitable representation of all participant groups, particularly marginalized and vulnerable populations.

In addition to these principles, algorithmic transparency is crucial. AI systems used in research should be transparent in their operations, allowing researchers to understand and explain methodological decisions and interpretations. However, many AI models, particularly deep learning systems, function as "black boxes," where decision-making processes are not easily interpretable even to developers. This presents a significant challenge for ensuring full transparency in qualitative research applications. Collaboration with AI experts is essential to navigate these complexities and promote ethical implementation, but achieving true interpretability remains an ongoing challenge.

Using AI Empowered CAQDAS in the Classroom

As with past technological advancements, the rise in generative AI will change the way scientific research is conducted. Public Health programs, including qualitative researchers and research methods instructors, must become adept at utilizing these tools so that they may prepare students for their roles as future researchers and program evaluators. Understanding the utility and ethics of AI-assisted qualitative research will enable students to engage critically with contemporary scholarship. Furthermore, the next generation of public health researchers will need to be able to make the most of generative AI to contribute to a field that is already being shaped by these tools.

As generative AI become ubiquitous, public health research will demand a skillset that is inclusive of generative AI tools for data analysis and interpretation. Additionally, only those with a comprehensive understanding of the ethical issues surrounding the use of generative AI in research, and their responsibilities as researchers, will be able to ensure that participants and their data are protected during the qualitative research process. Those who are highly skilled at working with this technology to increase productivity, creativity, and participant safety will be sought after in the job market, and on the forefront of innovation in qualitative research and analysis.

Related competencies are already guiding curriculum at all levels of public health education. In our own masters and doctoral programs, a core competency is to identify and evaluate emerging methods and technologies that can advance public health research and practice. Students must also establish ethical principles in research and practice. The field is rapidly advancing, and our curriculum needs to keep pace.

To assist students in learning how to use generative AI tools, instructors must stay up to date on advancements in platforms. Finding a community of researchers to converse with through academic societies, social media or their university library can help instructors stay current in their understanding of generative AI tools. In the classroom, instructors can utilize free trial periods offered by many of the computer assisted qualitative analysis software programs. Most of the programs offer lower-cost student licenses as well (Table 2). Additionally, we have included sample objectives and activities (see Table 3). Instructors can utilize these objectives and activities to develop new lesson plans or integrate them into their existing classroom strategies.

To further assist instructors, we have included two sample lesson plans aimed at integrating generative AI into qualitative methods courses. The first sample lesson plan (Lesson Plan 1; Appendix A) introduces students to the fundamentals of using QDAS with generative AI capabilities. The lesson, designed for undergraduate students, allows students to explore the possibilities and limitations of using AI tools for qualitative analysis and discuss the ethical considerations involved.

Lesson Plan 2 (Appendix B) is an interactive and comprehensive session focused on the possibilities of using AI-empowered tools to conduct qualitative analysis of textual data. It begins with an introduction, followed by hands-on segments for manual coding and employing AI tools. The workshop emphasizes understanding the strengths and limitations of these approaches, incorporating discussions on ethical considerations in AI usage. It concludes with a reflective group discussion and Q&A, aiming to deepen the understanding of qualitative data analysis in the modern research landscape. This lesson plan was created for a graduate-level research methods course for public health students focusing on social and behavioral sciences. When it was first delivered in the fall semester of 2023, most students described very limited exposure to AI tools like ChatGPT. There was a rich discussion around research ethics and the unique issues raised by using AI-empowered tools in qualitative data analysis.

Table 3.
Learning Objectives and Strategies to Incorporate AI-Empowered Qualitative Data Analysis in Undergraduate and Graduate Courses.

Learning Objective	Sample Activities
Describe what generative AI is and how it can be used in qualitative data analysis	**Interactive Demonstrations:** Use simple, interactive demonstrations of generative AI tools to show how they can process and analyze qualitative data. **Case Studies:** Present case studies where generative AI has been used in qualitative research, highlighting the process and outcomes. **Critical Analysis:** Engage students in critically analyzing the strengths and limitations of generative AI in qualitative research.
Recognize ethical issues that arise from using AI empowered tools for qualitative data analysis	**Ethical Debates:** Facilitate debates on the ethical implications of using AI in research, promoting awareness of responsible use. **Scenario Analysis:** Present hypothetical scenarios where AI tools are used in qualitative research, prompting students to identify potential ethical issues. **Policy Review:** Analyze existing policies on AI ethics in research, guiding students to understand the regulatory landscape.
Use AI empowered QDAS to generate initial first cycle codes of qualitative data and generate summary themes	**Interactive Demonstration**: Demonstrate how to use specific AI-empowered tools to analyze a qualitative data set. Focus on different coding strategies (e.g., descriptive, sentiment, etc.) and themeing methods (e.g., axial coding, pattern coding). **AI Coding Workshop:** Provide students with a dataset of qualitative responses. They will use an AI-powered Qualitative Data Analysis Software (QDAS) to input the data and generate initial codes. The activity will guide them through refining these codes into coherent themes.
Compare and contrast AI generated results to manually coded qualitative data	**Dual Coding Challenge:** Give students a set of qualitative data to code manually. Afterward, they will use an AI tool to code the same data. The class will then compare results, discussing similarities and differences. **Accuracy Assessment Task:** This task involves students assessing the accuracy of AI-generated codes against a 'gold standard' set of manually coded data. They will evaluate precision, recall, and overall accuracy.

	Reflective Analysis Seminar: Students will reflect on their experiences with both manual and AI coding, considering factors like time efficiency, ease of use, and perceived accuracy. They will share insights in a seminar format.
Create a data analysis plan to integrate AI empowered tools into a qualitative methods research project	**Mock Group Research Project:** In groups, students will design a mock qualitative research project, incorporating AI tools into their methodology. They will present their plans, highlighting how AI will be used at each stage.
	Individual Grant Proposal Project: Students will individually draft a sample grant proposal for a qualitative research study that incorporates the ethical use of AI-empowered CAQDS

Conclusion

The integration of AI tools into qualitative data analysis represents a transformative shift in public health research. By enhancing the efficiency and scalability of data processing, AI-empowered tools allow researchers to handle larger qualitative datasets and uncover deeper insights than traditional methods. However, the adoption of these technologies necessitates careful consideration of ethical implications, including data security, bias, and the need for transparency. Educating public health students on the use and limitations of AI tools, alongside traditional qualitative methods, will prepare them for the evolving research landscape.

Acknowledgements

We acknowledge the use of AI tools, including Microsoft Copilot and ChatGPT, in the development of this manuscript. Microsoft Copilot was utilized to brainstorm objectives and to compare its generated lesson plan to our initial draft, ensuring comprehensiveness. ChatGPT was also used early in the process to brainstorm and outline ideas, which were then compared with our independently developed concepts to refine the manuscript's overall structure and content.

Statement of Contributions

Both authors contributed equally to this manuscript. Dr. Wheldon led the writing of the sections detailing the current tools, their strengths and limitations, and the ethical considerations surrounding the use of AI in qualitative data analysis. Dr. McKee lead the sections focused on the overview of qualitative methods, the impact of AI on qualitative research, and the pedagogical strategies for teaching public health students about AI in qualitative data analysis. Both Drs. Wheldon and McKee contributed example lesson plans. Dr. Wheldon took the lead in organizing the manuscript and preparing it for publication.

References

1. Padgett, D. (2012). Qualitative and Mixed Methods in Public Health. https://doi.org/10.4135/9781483384511
2. Stickley, T., O'Caithain, A., & Homer, C. (2022). The Value of Qualitative Methods to Public Health research, Policy and Practice. Perspectives in Public Health, 142(4), 175791392210838. https://doi.org/10.1177/17579139221083814
3. Tolley, E. (2016). Qualitative Methods in Public Health: A Field Guide for Applied Research. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/templeuniv-ebooks/detail.action?docID=7104122
4. Taylor, B & Francis, K. (2013). Qualitative Research in the Health Sciences: Methodologies, Methods and Processes. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/templeuniv-ebooks/detail.action?docID=1244832
5. Griffith, D. M., Shelton, R. C., & Kegler, M. (2017). Advancing the Science of Qualitative Research to Promote Health Equity. Health Education & Behavior, 44(5), 673–676. https://doi.org/10.1177/1090198117728549
6. Saldaña, J. (2009). The Coding Manual for Qualitative Researchers. SAGE Publications. https://doi.org/10.1109/test.2002.1041893
7. Wolski, U. (2018). The History of the Development and Propagation of QDA Software. The Qualitative Report. https://doi.org/10.46743/2160-3715/2018.2984
8. Gilbert, L. S., Jackson, K., & di Gregorio, S. (2013). Tools for Analyzing Qualitative Data: The History and Relevance of Qualitative Data Analysis Software. Handbook of Research on Educational Communications and Technology, 221–236. https://doi.org/10.1007/978-1-4614-3185-5_18
9. Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare, 11(6), 887. https://doi.org/10.3390/healthcare11060887
10. Biswas, S. S. (2023). Role of Chat GPT in Public Health. Annals of Biomedical Engineering, 51(5). https://doi.org/10.1007/s10439-023-03172-7
11. Computer Does Qual: Avoiding AI-overclaim & false equivalency. (2023). Greenbook.org. https://www.greenbook.org/insights/the-prompt-ai/computer-does-qual-avoiding-ai-overclaim-and-false-equivalency
12. Hryciw, B. N., Andrew, & Kwadwo Kyeremanteng. (2023). Guiding principles and proposed classification system for the responsible adoption of artificial intelligence in scientific writing in medicine. PubMed, 6. https://doi.org/10.3389/frai.2023.1283353
13. AI is Transforming Qualitative Research Coding. (2024). Greenbook.org. https://www.greenbook.org/insights/qualitative-market-research/ai-is-transforming-qualitative-research-coding
14. Atefeh Jebeli, Lujie Karen Chen, Guerrerio, K., Papparotto, S., Berlin, L. J., & Brenda Jones Harden. (2023). Quantifying the Quality of Parent-Child Interaction Through Machine-Learning Based Audio and Video Analysis: Towards a Vision of AI-assisted Coaching Support for Social Workers. ACM Journal on Computing and Sustainable Societies. https://doi.org/10.1145/3617693
15. Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. 28th International Conference on Intelligent User Interfaces. https://doi.org/10.1145/3581754.3584136
16. Muller, M., Min Kyung Lee, Eric, Mimno, D., & N. Sadat Shami. (2016). Machine Learning and Grounded Theory Method. https://doi.org/10.1145/2957276.2957280
17. Guo, X., & Chen, Y. (2024). Generative AI for Synthetic Data Generation: Methods, Challenges and the Future. ArXiv.org. http://arxiv.org/abs/2403.04190

18. Korstjens, I., & Moser, A. (2017). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. European Journal of General Practice, 24(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092
19. Luiz Rafael Andrade, António Pedro Costa, Ronaldo Nunes Linhares, Azevedo, C., & Luís Paulo Reis. (2018). Qualitative Data Analysis Software Packages: An Integrative Review. Advances in Intelligent Systems and Computing, 279–290. https://doi.org/10.1007/978-3-030-01406-3_24
20. Knevel, R., & Liao, K. P. (2022). From real-world electronic health record data to real-world results using artificial intelligence. Annals of the Rheumatic Diseases, annrheumdis-2022-222626. https://doi.org/10.1136/ard-2022-222626
21. Al-Rawi, A., Siddiqi, M., Li, X., Vandan, N., & Grepin, K. (2021). A thematic analysis of Instagram's gendered memes on COVID-19. Journal of Visual Communication in Medicine, 44(4), 1–14. https://doi.org/10.1080/17453054.2021.1941808
22. Covert, H. H., & Koro-Ljungberg, M. (2015). Layers of narratives, images, and analysis. Qualitative Research Journal, 15(3), 306–318. https://doi.org/10.1108/qrj-08-2014-0042
23. Seitz, C. M., & Muhsin Michael Orsini. (2022). Thirty Years of Implementing the Photovoice Method: Insights From a Review of Reviews. Health Promotion Practice, 23(2), 281–288. https://doi.org/10.1177/15248399211053878
24. Park, I., Jae Hyon Park, Yoon, J., Song, I.-A., Hyo-Seok Na, Ryu, J.-H., & Oh, A.-Y. (2023). Artificial intelligence model predicting postoperative pain using facial expressions: a pilot study. Journal of Clinical Monitoring and Computing, 38(2), 261–270. https://doi.org/10.1007/s10877-023-01100-7
25. Prescott, M. R., Yeager, S., Ham, L., Rivera, C. D., Serrano, V., Narez, J., Dafna Paltin, Delgado, J., Moore, D. J., & Montoya, J. (2024). Comparing the Efficacy and Efficiency of Human and GenAI Qualitative Thematic Analyses (Preprint). JMIR AI, 3, e54482–e54482. https://doi.org/10.2196/54482
26. Duby, Z., Hartmann, M., Mahaka, I., Munaiwa, O., Nabukeera, J., Vilakazi, N., Mthembu, F., Colvin, C. J., Mensch, B., & van der Straten, A. (2015). Lost in Translation: Language, Terminology, and Understanding of Penile–Anal Intercourse in an HIV Prevention Trial in South Africa, Uganda, and Zimbabwe. The Journal of Sex Research, 53(9), 1096–1106. https://doi.org/10.1080/00224499.2015.1069784
27. Morgan, D. L. (2023). Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. International Journal of Qualitative Methods, 22. https://doi.org/10.1177/16094069231211248

Appendix A:
Lesson 1: Introduction to Generative AI for Qualitative Data Analysis in Public Health Research

Objectives:
- Understand the role of generative AI in enhancing qualitative data analysis within the context of public health research.
- Explore practical applications of generative AI tools specific to public health studies.

Duration:
- 60 minutes

Materials Needed:
- Projector or screen for presentation
- Laptops or computers with internet access
- Access to AI enhanced QDA software for student use

Activity Steps:
1. **Introduction to Generative AI (15 minutes)**
    - Briefly define generative AI as a subset of artificial intelligence that generates new content based on patterns learned from existing data.
    - Present specific use cases of generative AI tools applicable to public health research:
        - Text summarization: How generative AI can summarize lengthy qualitative interviews or reports.
        - Topic modeling: Identifying key themes and topics within large datasets.
        - Sentiment analysis: Analyzing public health-related social media posts.
    - Discuss potential benefits of using these tools in qualitative or mixed-method public health studies.
        - Efficiency and speed
        - Improved accuracy
        - Scalability
        - Enhanced insights
        - Predictive modeling
        - Others?
2. **Hands-On Exploration (30 minutes)**
    - Form small groups of students.
    - Provide access to a generative AI platform (e.g., OpenAI's GPT-3 Playground).
    - Assign a public health research question (e.g., "Explore barriers to vaccination acceptance") to each group.
    - Instruct students to use a generative AI tool to generate relevant content related to their research question (e.g., prompt AI to "draft ten 200-word first-person vignettes from the perspectives of different individuals who were hesitant to get vaccinated or had trouble accessing them).
    - Instruct students to use an AI tool to analyze the vignettes: searching for themes/codes/sentiment
    - Share findings within the class.
    - Proposed discussion questions
        - How might using AI to generate the data have influenced the findings?

        - What similarities and differences were there in the AI themes/codes and how you would interpret the data using manual qualitative coding?
3. **Ethical Considerations (10 minutes)**
    - Discuss ethical implications of using AI-generated content in public health research:
        - Bias: How generative AI models may inherit biases from training data.
        - Transparency: The importance of transparency when reporting AI-generated results.
        - Privacy: Considerations when handling sensitive health data.
        - IRB: What questions might be raised by an IRB about AI assisted qualitative research?
    - Brainstorm strategies to mitigate these ethical challenges.
4. **Conclusion (5 minutes)**
    - Summarize the potential of generative AI in advancing qualitative data analysis within the field of public health.
    - Encourage students to explore further and critically evaluate the role of AI in their research endeavors.
        - Students evaluate the work
        - Choose 2 (1 from each group) to demonstrate that have free versions

Appendix B:
Lesson 2: Using AI for qualitative data analysis

This is an interactive and comprehensive session designed for undergraduate or graduate students, focusing on the possibilities of using AI-empowered tools to conduct qualitative data analysis of text data.

Learning Objectives:
1. Apply manual template and emotion coding to excerpts from qualitative interviews.
2. Describe current capabilities of AI-empowered qualitative data analysis software (e.g., Atlas.ti).
3. Explore text analysis capabilities in ChatGPT and discuss ethical issues using AI tools.

Preparation:
1. Register for a Free Trial of Atlas.ti Cloud https://atlasti.com/free-trial-version
2. Register an account with ChatGPT at https://chat.openai.com/

Materials Needed:
- Sample qualitative interview excerpts.
- Access to Cloud based Atlas.ti and ChatGPT.
- Projector and screen for demonstrations.
- Handouts with step-by-step instructions.
- Students needs computers connected to the internet.

Activities:
1. **Introduction (10 minutes)**
    - Brief overview of qualitative data analysis.
    - Introduce the approaches: manual coding and AI auto-coding and summarizing.
2. **Approach 1: Manual Coding (30 minutes)**

- **Lecture (10 minutes):** Explain template and emotion coding, their importance in qualitative analysis.
- **Practice (20 minutes):** Hands-on activity with interview excerpts.
  - Provide excerpts and guide students through the coding process.

3. **Approach 2: AI Tools in Atlas.ti (30 minutes)**
   - **Lecture (10 minutes):** Overview of AI-empowered tools and their applications in Atlas.ti.
   - **Practice (20 minutes):**
     - Demonstration of Atlas.ti features.
     - Discussion on how AI tools aid in qualitative data analysis.
4. **Approach 3: ChatGPT for text analysis (30 minutes)**
   - Lecture (10 minutes): Overview of ChatGPT and large language models
   - Practice (20 minutes):
     - Demonstration of ChatGPT capabilities and the importance of well-designed prompts.
     - Discussion on how ChatGPT compares to AI tools in Atlas.ti.
5. **Q&A and Discussion (10 minutes)**
   - Address questions and discuss practical applications.

Group Discussion Questions:

- What was most surprising about what you learned today?
- What are some strengths and limitations of the approaches we demonstrated today?
- What would you want to learn more about?
- What are the key ethical considerations using AI tools for QDA?

Post-Workshop:

- Provide an online resource list for further learning.
- Share the presentation slides and practice data sets with attendees.

Approach 1: Manual Coding

The **purpose** of this assignment is to **practice qualitative coding** and compare the results of the qualitative analysis with your classmates. Below you will find **open-ended responses** that were provided in a study about HPV vaccination among gay and bisexual men (ages 18-26). In this study the participants completed a questionnaire about HPV vaccination. For those who indicated that they intended to get vaccinated within the next 12 months, they were asked the following open-ended question: **Why did you decide to get the HPV vaccine?** Your task is to read the responses and generate qualitative codes. You can apply more than one code per response. The codes should descriptive and concise. You can reuse codes when appropriate.

Instructions:

First Cycle Coding

1. Deductively code responses using established codebook/framework. Each open-ended response should be coded as a behavioral belief, normative belief, or control belief. If these don't fit code it as 'something else'.
2. Inductively code using emotion and open coding.

3. Compare your codes with your group. Did you generate any of the same codes? Where did you differ?

***A priori* codebook based on Theory of Planned Behavior**

Construct	Definition
Behavioral Beliefs	An individual's belief about consequences of particular behavior. The concept is based on the subjective probability that the behavior will produce a given outcome.
Normative Beliefs	An individual's perception about particular behavior, which is influenced by the judgment of significant others.
Control Beliefs	An individual's beliefs about the presence of factors that facilitate or impede performance of the behavior.

Second Cycle Coding

The second round of coding allows the researcher to refine codes and categories in order to develop themes and align data with research and theory. Some approaches to second cycle coding include:

- **Pattern coding**, which groups similar codes/categories into a smaller number of sets, themes, or constructs.
- **Focused coding**, which searches for the most frequent or significant codes/categories to create or align with conceptual themes.
- **Axial coding**, which explores how codes, categories, and subcategories relate to each other.

Open-Ended Response	Framework Coding	Emotion Coding	Open Coding
1. Because vaccines are safe and if there is a significant benefit for me, I believe the benefits outweigh the insignificant costs.			
2. So I can hookup with one less worry about getting an STD.			
3. I believe in preventative medicine, where doing something now to prevent something very serious down the road is the ideal course of action.			
4. I Would rather be safe than sorry.			

5. It protects me and others from getting ill, so it should be a no-brainer.			
6. I want to live a healthy life and safeguard myself from preventable diseases. This would all be for my peace of mind, and for the safety of my future sexual/romantic partners.			
7. It's one less disease to worry about contracting from sexual contact. There are generally no downsides to getting vaccinated aside from occasionally feeling a little sore at the location of the shot.			
8. If it was easy enough. I've talked to my doctor about an HPV test before but he didn't really get it. Many doctors don't seem to know that gay/bi men are at risk and need to be tested/ vaccinated. My experience so far has been rather dismissive. When I last got a round of STD testing, my doctor, by default, excluded the HPV test. I asked about it but was unable to convince them to order the test.			
9. The increased risk of cancers, possibility of unknowingly transferring something to someone, and anal warts don't really sound like a great time to be honest either.			
10. Anything that would help me stay healthy I am all for it.			

11. I could be spreading it without even realizing it.			
12. My mom wanted me to get vaccinated.			
13. My doctor recommended I get vaccinated, so it was really convenient to get it then.			
14. My insurance paid for it so it was free to me. So why not!			

Approach 2: Exploring AI in Qualitative Data Analysis with ATLAS.ti

Objective:
- Understand the application of AI tools in qualitative data analysis, focusing on sentiment analysis and AI-driven coding.

Duration:
- 25 minutes

Tools Required:
- Cloud-based ATLAS.ti

Thought Questions:
Before starting the tasks, consider the following questions:
1. **Strengths and Limitations:**
   - What strengths do you anticipate AI will have in qualitative data analysis?
   - What limitations or challenges might AI face in this context?
2. **Ethical Considerations:**
   - What ethical issues could arise from using AI in qualitative data analysis?
   - How might AI impact the interpretation of qualitative data?
3. **Comparison with Traditional Methods:**
   - How do you think AI-assisted analysis will compare with traditional manual coding?
   - In what ways could AI enhance or detract from qualitative data analysis?

Tasks:
1. **Import a sample text dataset into ATLAS.ti.**
   - Choose a dataset relevant to your field of study.
2. **Read through a few excerpts and manually code them as 'positive', 'negative', or 'neutral'.**
   - Select excerpts that clearly convey an emotion.
3. **Run the AI sentiment analysis tool on the same excerpts.**
   - Compare the AI-generated codes with your manual codes.

4. **Reflect on the results of the sentiment analysis.**
    - Consider the accuracy and nuances captured or missed by the AI.
5. **Experiment with line-by-line AI suggested codes.**
    - Analyze how the AI interprets and codes each line.
6. **Apply AI coding to the entire document.**
    - Observe the patterns and themes identified by the AI.
7. **Engage with the conversational AI feature, if available, to ask questions about the data,**
    - Evaluate the usefulness of conversational AI in exploring qualitative data.

Reflection and Discussion:
- After completing the tasks, write a brief reflection on your findings, particularly focusing on the thought questions. Be prepared to discuss your insights in the next class session.

Approach 3: Utilizing ChatGPT for Qualitative Analysis

Objective:
- Explore text analysis capabilities in ChatGPT and discuss ethical issues using AI tools.

Duration:
- 25 minutes

Tools Required:
- Access to ChatGPT and excerpts from an HPV vaccination study.

Pre-Activity Thought Questions:
Before you begin, consider:
1. **ChatGPT's Analytical Approach:**
    - How might ChatGPT interpret and code data differently from a human researcher, especially in a sensitive topic like HPV vaccination?
2. **Accuracy and Depth:**
    - What level of accuracy and depth do you expect from ChatGPT in identifying themes and sentiments in the data?
3. **Ethical Considerations:**
    - What ethical considerations should be kept in mind when using AI for analyzing health-related qualitative data?

Tasks:
1. **Prepare Excerpts from the HPV Vaccination Study.**
    - Excerpts will be provided.
2. **Perform Emotion/Sentiment Coding Using ChatGPT.**
    - Input the excerpts into ChatGPT.
    - Ask ChatGPT to code each excerpt as positive, negative, or neutral based on the sentiment conveyed.
3. **Conduct Open Coding with ChatGPT.**
    - Ask ChatGPT to identify themes or categories in the data without preconceived labels.

4. **Create a Thematic Table with ChatGPT.**
   - Request ChatGPT to construct a table summarizing the main themes, associated codes, and representative quotes from the data.
5. **Evaluate ChatGPT's Performance.**
   - Assess the appropriateness of the sentiments and themes identified by ChatGPT.
   - Note how well ChatGPT captures the nuances and complexities of the data.
6. **Reflect on the Use of ChatGPT.**
   - Consider how the AI's analysis aligns with or deviates from your understanding of the data.
   - Think about the potential uses and limitations of AI in qualitative health research.

Reflection and Discussion:

- After completing the tasks, write a brief reflection on your experience and findings, focusing on your initial thought questions and the effectiveness of ChatGPT in this context. Be prepared to discuss your insights and any surprises or challenges you encountered in the next class session.