

Original Paper

Refining Established Practices for Research Question Definition to Foster Interdisciplinary Research Skills in a Digital Age: Consensus Study With Nominal Group Technique

Jana Sedlakova^{1,2}, PhD; Mina Stanikić^{1,2,3}, MD, PhD; Felix Gille^{1,2}, PhD; Jürgen Bernard^{1,4}, Prof Dr; Andrea B Horn^{1,5,6}, Dr rer nat; Markus Wolf^{1,6}, Dr phil; Christina Haag^{1,2,3}, PhD; Joel Floris^{1,7}, Dr; Gabriela Morgenshtern^{1,4}, MSc; Gerold Schneider^{1,8}, Prof Dr; Aleksandra Zumbrunn Wojczyńska^{1,9}, Dr; Corine Mouton Dorey^{1,10}, Dr med, Dr phil; Dominik Alois Ettlin^{1,9}, PD, Dr med; Daniel Gero^{1,11}, PD, MD, PhD; Thomas Friemel^{1,12}, Prof Dr; Ziyuan Lu^{1,7}, Dr med; Kimon Papadopoulos^{1,13}, MPH; Sonja Schläpfer^{1,14}, MSc; Ning Wang¹, PhD; Viktor von Wyl^{1,2,3}, Prof Dr

¹Digital Society Initiative, University of Zurich, Zurich, Switzerland

²Institute of Implementation Science in Healthcare, Faculty of Medicine, University of Zurich, Zurich, Switzerland

³Epidemiology, Biostatistics and Prevention Institute, Faculty of Medicine, University of Zurich, Zurich, Switzerland

⁴Department of Informatics, Faculty of Business, Economics and Informatics, University of Zurich, Zurich, Switzerland

⁵Center for Gerontology, University of Zurich, Zurich, Switzerland

⁶Department of Psychology, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

⁷Institute of Evolutionary Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland

⁸Department of Computational Linguistics, Faculty of Business, Economics and Informatics, University of Zurich, Zurich, Switzerland

⁹Center of Dental Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland

¹⁰Institute of Biomedical Ethics and History of Medicine, Faculty of Medicine, University of Zurich, Zurich, Switzerland

¹¹Department of Surgery and Transplantation, University Hospital of Zurich, Zurich, Switzerland

¹²Department of Communication and Media Research, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

¹³Institute of Implementation Science in Healthcare, Faculty of Medicine, University of Zurich, Zurich, Switzerland

¹⁴Institute for Complementary and Integrative Medicine, University Hospital of Zurich, Zurich, Switzerland

Corresponding Author:

Jana Sedlakova, PhD
Digital Society Initiative
University of Zurich
Raemistrasse 69
Zurich, 8001
Switzerland
Phone: 41 0786753991
Email: sedlakova@ifi.uzh.ch

Abstract

Background: The increased use of digital data in health research demands interdisciplinary collaborations to address its methodological complexities and challenges. This often entails merging the linear deductive approach of health research with the explorative iterative approach of data science. However, there is a lack of structured teaching courses and guidance on how to effectively and constructively bridge different disciplines and research approaches.

Objective: This study aimed to provide a set of tools and recommendations designed to facilitate interdisciplinary education and collaboration. Target groups are lecturers who can use these tools to design interdisciplinary courses, supervisors who guide PhD and master's students in their interdisciplinary projects, and principal investigators who design and organize workshops to initiate and guide interdisciplinary projects.

Methods: Our study was conducted in 3 steps: (1) developing a common terminology, (2) identifying established workflows for research question formulation, and (3) examining adaptations of existing study workflows combining methods from health research and data science. We also formulated recommendations for a pragmatic implementation of our findings. We conducted

a literature search and organized 3 interdisciplinary expert workshops with researchers at the University of Zurich. For the workshops and the subsequent manuscript writing process, we adopted a consensus study methodology.

Results: We developed a set of tools to facilitate interdisciplinary education and collaboration. These tools focused on 2 key dimensions— content and curriculum and methods and teaching style—and can be applied in various educational and research settings. We developed a glossary to establish a shared understanding of common terminologies and concepts. We delineated the established study workflow for research question formulation, emphasizing the “what” and the “how,” while summarizing the necessary tools to facilitate the process. We propose 3 clusters of contextual and methodological adaptations to this workflow to better integrate data science practices: (1) acknowledging real-life constraints and limitations in research scope; (2) allowing more iterative, data-driven approaches to research question formulation; and (3) strengthening research quality through reproducibility principles and adherence to the findable, accessible, interoperable, and reusable (FAIR) data principles.

Conclusions: Research question formulation remains a relevant and useful research step in projects using digital data. We recommend initiating new interdisciplinary collaborations by establishing terminologies as well as using the concepts of research tasks to foster a shared understanding. Our tools and recommendations can support academic educators in training health professionals and researchers for interdisciplinary digital health projects.

(*JMIR Med Educ* 2025;11:e56369) doi: [10.2196/56369](https://doi.org/10.2196/56369)

KEYWORDS

research question; digitalization; digital data; data science; health research; interdisciplinary

Introduction

Background

Health research increasingly leverages the abundance of data from our “digital lives,” including mobility data, social media data, or data from wearables [1,2]. Such digital data are commonly “unstructured” because it may not conform to a tabular format (eg, images, videos, sound, and free text) and often require specific expertise for harvesting; transforming; preprocessing; and creating meaningful insights into health, disease, and treatment [1,3-5]. Moreover, such digital data are often originally generated for nonresearch purposes and without addressing a specific research question [6]. In turn, they may lack standard quality attributes found in digital data collected for specific research purposes, such as depth, completeness, or consistency, which present methodological complexities to meaningfully use these data [1]. Therefore, reusing these digital unstructured data for health research requires diverse expertise, skills, and interdisciplinary collaboration between health domain experts (eg, clinicians and health scientists) and data scientists as method experts (eg, from data science, computer science, or statistics) [5,7,8].

Such interdisciplinary collaborations are often faced with challenges due to the seemingly conflicting research approaches between the disciplines. In addition to differences in terminologies and concept definitions, the prevailing emphasis of linear deductive approaches in health research contrasts with the often more explorative and iterative approaches used in data science [7]. In health research, it is customary to predefine key elements of the scientific process, including a research question and related hypothesis, in a protocol and scientific report (eg, STROBE [Strengthening the Reporting of Observational Studies in Epidemiology] or PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] guidelines) [9-12]. These standard practices are deeply influenced by the tradition of clinical trials and treatment development, which place a strong emphasis on measurement validity, robustness, scientific rigor, and safety [13], as errors in study conduct or treatment

could place study participants at risk. By contrast, data science generally tends to emphasize exploration, pattern discovery, or hypothesis generation as well as more iterative and inductive analysis approaches [1,14]. Some health researchers may perceive this greater emphasis on iterative approaches as lacking scientific rigor or focus on specific research questions.

For young researchers, interdisciplinary digital health collaborations might be particularly challenging because they need to balance traditional scientific methods with more iterative data-driven techniques. This dual demand highlights the importance of fostering interdisciplinary skills in education, enabling students to balance the rigorous demands of hypothesis-driven research with the iterative and inductive approaches of data science. Addressing these complexities represents an educational challenge for both established and young researchers.

Despite broad recognition of their importance, both practical and teaching or educational guidance on how to manage and overcome the challenges of interdisciplinary digital health collaborations are scarce. Such guidance is also important for educational purposes to foster skills for interdisciplinary collaboration among both young and established researchers as well as health professionals. Continuous education for experienced researchers is equally important to keep them updated with evolving methods and foster effective collaboration across disciplines.

This Study

To address this need, our study focuses on skill development to successfully navigate interdisciplinary collaborations and education in health-related research fields. We reviewed established workflows for research question formulation and investigated whether and how established workflows in health research may require adaptations to accommodate inductive and exploratory data science practices and novel analysis techniques. The study findings were translated into a set of tools and recommendations designed to facilitate interdisciplinary education and collaboration. These tools focus on 2 key

dimensions—*content and curriculum* and *methods and teaching style*—and can be applied in various educational and research settings. Lecturers can use them to design interdisciplinary courses, supervisors can guide PhD and master’s students in their interdisciplinary projects, and principal investigators can design and organize workshops to initiate and guide interdisciplinary projects. By implementing these tools, educators and researchers can create more cohesive and productive educational resources for interdisciplinary collaborations. In the following sections, we offer our insights and a more detailed outline of how our study findings can inform both the *content* and *methods* dimensions, using an existing interdisciplinary course as an example.

The aims and findings of our study are intended to be globally relevant and applicable to all researchers using digital data in the context of health research and health care. Importantly, they also provide academic educators with a clear workflow and practical recommendations for discussing and addressing the challenges of interdisciplinary collaboration. As the focus is on research question formulation, a fundamental aspect of the research process, these recommendations are especially valuable for educational purposes, helping educators guide researchers and students through this essential phase of research projects.

To achieve our aims, we chose a consensus study approach that is appropriate to harmonize and bridge insights from experts from diverse research disciplines. Moreover, we focused our

effort on the different approaches of research question formulation as the guiding example for this study because it represents a central step in guiding the research process and subsequent study design decisions. This process also served as an illustrative example to highlight the differences in research approaches between health research and data science.

Methods

Consensus Methods

We used the nominal group technique with expert groups to gather insights from a diverse range of experts. This approach aimed to foster interdisciplinary skills and knowledge and achieve consensus on adapting research question development.

This study was structured by the following three high-level steps (Figure 1):

1. To create a common terminology to facilitate interdisciplinary and transdisciplinary collaborations that are required for research projects reusing digital data (ie, repurposing data originally generated for nonresearch purposes)
2. To describe the “established workflow” for research question formulation in health research on the basis of existing literature
3. To formulate suggestions and recommendations for adapting the “established workflow”

Figure 1. Study flow.



To inform steps 1 and 2, a rapid literature review was performed to identify established concepts for defining a research question in health research and data science as well as in other fields (refer to the Preparatory Research: Literature Search for the “Established Workflow” and an Example Scenario section). Expert inputs were gathered in a series of three 1.5-hour expert

workshops. To foster a focused discussion in the workshops, participants were asked to complete preworkshop tasks. These inputs were summarized by JS and VvW and presented at the beginning of each workshop to discuss potential disagreements and allow participants to explain or comment on their and others’ inputs. The consensus and agreement of each objective

were reached by an iterative, deliberative process. This included expert inputs before workshops, discussions during the workshop, and finally, expert feedback on and approval of the consolidated findings. These findings were synthesized, formulated, and shared by JS and VvW after each workshop. Furthermore, each participant was actively involved in the manuscript writing. These methods facilitated the systematic collection of input from participants in group and individual settings, enabling a comprehensive understanding of experts' knowledge and consolidating diverse perspectives. Workshops were recorded after receiving consent from the team. Workshop minutes, including the results of the preworkshop tasks, were sent for approval to the expert group. When necessary, individual researchers were contacted after workshops for clarification on specific issues raised during the workshops. The documentation and reporting of the workshop and the Accurate Consensus Reporting Document (ACCORD) checklist [15] for the consensus methodology are available in [Multimedia Appendices 1 and 2](#).

The first 2 steps were accomplished in workshops 1 and 2. Building on these results, a third workshop was dedicated to identifying the need for the adaptation of established research practices in the health field to streamline collaboration with data scientists and to better integrate and communicate the need for research principles and standards, including open science and reproducibility.

Participants

The consensus meetings in the form of expert groups were led by JS and VvW, who led a previous project focusing on challenges and best practices of digital data, which inspired this study. Furthermore, JS is a scientific manager of the scientific community whose members were recruited for the consensus exercise. VvW's expertise lies in epidemiology and digital health research, and JS's expertise is mainly in digital ethics considering health research and health care. The workshop participants were recruited by JS and VvW among the diverse members of the Digital Society Initiative (DSI) Health Community at the University of Zurich. The members from the DSI were selected because it is a competence center for digital transformation that fosters interdisciplinary collaborations and projects studying the interplay and implications of digital transformations in society. Participants were included if they had experience with projects using digital data or planned to be involved in such projects. The workshop was promoted on the DSI website, through newsletters, and via word-of-mouth within the community. A total of 21 researchers from different disciplines and from all career stages participated in the workshops. This number of participants enabled to have an expert group with sufficient diversity to foster discussions and include insights from diverse disciplines. Of the 21 researchers, 13 (62%) represented health research, 3 (14%) represented data science, and 7 (33%) represented the social sciences and humanities.

Preparatory Research: Literature Search for the “Established Workflow” and an Example Scenario

A rapid literature search was conducted to inform the planning of the workshops and to develop a project roadmap (by JS and

VvW). To gather information on the established workflow for research question formulation (steps 1 and 2), we searched the literature for publications, reviews, and course guidelines written either in English or German in PubMed and Google Scholar databases (search terms are provided in [Multimedia Appendix 1](#)). The search was further complemented by retrieval of guidelines from universities in Switzerland, Germany, the United States, and the United Kingdom, for which we searched on selected university websites. In addition, coauthors contributed materials they were familiar with or had previously used for teaching or research purposes. On the basis of this literature, we proposed the initial model for the “established workflow” that combines existing well-established frameworks and practices. To guide the discussions of our workshops, we developed an example scenario of digital data reuse for health research, which was communicated to participants before the workshops ([Multimedia Appendix 3](#)).

Ethical Considerations

The study followed the recommended procedures of the ethics committee of the Medical Faculty of the University of Zurich by completing the Data Protection/Ethics Self-Assessment Tool and received an exempt status. Participants were informed about the study's scope and goals as well as the nature of their involvement. They provided consent before the workshop and were informed that they could withdraw from the study at any time without providing a reason. The participants did not receive any compensation for their participation. The only personal information collected for the study was sociodemographic data, which were anonymized.

Results

Establishing a Common Terminology

Anticipating that a lack of harmonization concerning terminologies and concepts may hinder an effective interdisciplinary workshop collaboration, we aimed to establish a shared understanding of common terminologies. To this end, the workshop leaders (JS and VvW) developed a glossary before the first workshop, which was discussed and further refined by collecting written feedback from the participants after workshop 1 ([Table 1](#)).

The workshop discussions concerning the glossary centered around discipline-specific interpretations of concepts such as “research task,” “research objectives,” “research aims,” and “research goals,” whose interpretations were dependent on the embedding in different research methodologies, such as qualitative or quantitative research approaches. A central discussion centered around the recognition of different “research tasks,” that is, high-level research aims from a methodological viewpoint, including, for example, exploration, confirmation, prediction, methods development, or theory development. For prediction and classification tasks, participants mentioned 2 subcategories of analyses, which are supervised learning methods that rely on labeled data and outcomes and include the broad class of (multivariable) regression models. By contrast, unsupervised methods (eg, neural networks) aim to find new data structures and features without the need for prior labeling and are often developed in a less linear, inductive manner.

Table 1. Common terminology for interdisciplinary research projects using digital data.

Terms	Definitions
Confirmatory research	<ul style="list-style-type: none"> Hypothesis-driven research, experimental research, or research aiming at testing and confirming a hypothesis in a broader context of a theory. This research is also referred to as hypothetico-deductive research in some disciplines.
Exploratory research	<ul style="list-style-type: none"> Data-driven research that aims at exploring new patterns and associations to formulate hypotheses.
Hypothesis	<ul style="list-style-type: none"> A tentative, hypothetical prediction of the nature and direction of relationships between sets of data, phrased as a declarative statement. It is an assumption about scientific laws, causation, or empirical regularities. A hypothesis should be testable or falsifiable. This refers to quantitative evidence-based health research.
Unstructured data	<ul style="list-style-type: none"> Raw data that are not in a predefined structure (eg, tables) or data that may be structured but still require substantial preprocessing or feature extraction (eg, continuous sensor data).
Principles and criteria of good research and research practice	<ul style="list-style-type: none"> A set of values and norms for good conduct of research, including validity, scientific integrity, objectivity, and ethical study conduct.
Research aim	<ul style="list-style-type: none"> The research aim is the overall, general, and long-term intention of a research project. The research aim describes the “what” of the research—where we aspire to be at the end.
Research design	<ul style="list-style-type: none"> Research design describes the general outline of data collection (eg, cross-sectional and longitudinal studies) and analytical methods (eg, randomization, observational, and with or without control group) to answer the RQ^a. It describes the “how” of research.
Research objective	<ul style="list-style-type: none"> The specific goal linked to a RQ [16].
Research problem	<ul style="list-style-type: none"> The research problem describes the rationale for a study, for example, by highlighting the societal or medical needs. It describes the “why”—the specific needs a study wants to address.
RQ	<ul style="list-style-type: none"> A clear and concise question determining the research aim, objective, design, methodology, data collection, and analysis. The RQ narrows the aim and objective of the research. The process of defining a good RQ is dynamic and iterative. The RQ is refined through the different steps of the research cycle. We define the RQ in the context of quantitative evidence-based health research.
Research task	<ul style="list-style-type: none"> A research task describes a high-level classification of aims or tasks in research, including descriptive research, exploratory research, confirmatory research, prediction and classification, theory development, or methods development.
Reuse of digital data	<ul style="list-style-type: none"> The process of harvesting, transforming, and using structured or unstructured digital information that was initially generated for purposes other than research.
Theory or model	<ul style="list-style-type: none"> A systematic, structured explanation or representation of facts, phenomena, or processes that sets the ground for research design, formulation of hypotheses, and predictions.
Tools to specify RQ	<ul style="list-style-type: none"> Frameworks and tools that facilitate the development of specific aspects of defining the RQ or study design.
Types of RQ	<ul style="list-style-type: none"> The type of RQ determines the main approach for achieving the research aim. Usually, there is a difference between quantitative and qualitative RQs that reflect quantitative and qualitative approaches. Quantitative approaches use statistical and mathematical methods to address precise questions, typically using a deductive approach with a strong emphasis on the framework and structure. By contrast, qualitative approaches use, for example, open-ended responses, focus groups, and interview-based techniques and focus on individual experiences and singularities. It seeks to determine or discover a process or define experiences. RQs tend to be inductive, flexible, adaptable, and nondirectional [17].

^aRQ: research question.

The group further discussed the central role of hypotheses and linear, highly structured research approaches in health research, for example, in confirmatory research tasks (confirming a hypothesis, eg, by use of randomized controlled experiments or trials) and, to some extent, research focusing on predictions tasks (developing prediction models or classifiers to predict future events or out-of-sample attributes). In health research, it

is generally recommended that the development of prediction methods follows a protocol that includes careful selection of predictors and (external) validation of the final model [18]. At the same time, it was also pointed out that some quantitative research tasks, such as methods development (ie, the development and validation of data analysis methods) or exploratory research (ie, detection of patterns and associations

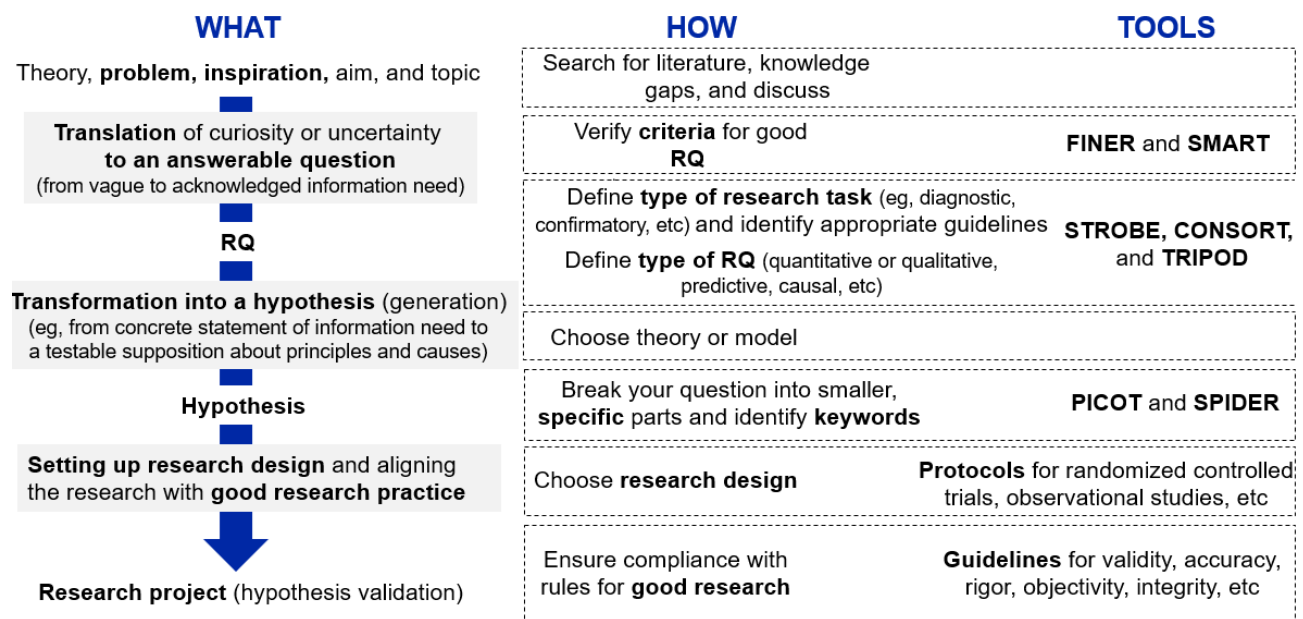
to generate new hypotheses), as well as qualitative research, generally depend much less on the specification of hypotheses. Workshop participants with a qualitative background argued that hypotheses can be implicitly involved in the research project. In qualitative research, it is common for the research question to evolve due to the necessity to critically reflect and adjust the study focus in each research step. As a result, the overall research process in qualitative research and some quantitative tasks, such as methods development, can be more iterative and dialectical when compared to deductive or confirmatory health research.

These discussions led to a key insight that interdisciplinary collaborations may be streamlined through the identification and discussion of the most appropriate “research task” early on, which can help guide subsequent discussions about the research question and the role of hypotheses in a common direction.

Summarizing Established Workflows for Research Question Formulation

The first 2 workshops were dedicated to better understanding how different disciplines approach the initial steps of a research project, including research question definition and study design choices. Informed by our rapid literature review, Figure 2 illustrates a summary workflow for established research design practices in health research. The vertical axis of Figure 2 illustrates the recommended steps for defining a research question (the “what”), starting from finding inspiration to developing a hypothesis, designing an appropriate study, and validating the hypothesis. Aligned with these definition steps, Figure 2 displays established practices (the “how”) to execute the recommended steps. The third column references various frameworks and checklists aiding the implementation of each recommended step (the “tools”).

Figure 2. Workflow of the recommended practices for defining a good research question. CONSORT: Consolidated Standards of Reporting Trials; FINER: Feasible, Interesting, Novel, Ethical, and Relevant; PICOT: Population, Intervention, Comparison, Outcome, and Time; RQ: research question; SMART: Specific, Measurable, Achievable, Realistic, and Timely; SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, Research Type; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.



The workflow of established practices generally starts with identifying a meaningful problem or question to be addressed in a study. The inspiration often emerges from a real-world challenge or knowledge gaps, but it can also be derived from existing theories or be triggered by discussions among colleagues. Workshop members also mentioned the influential role of funding criteria (ie, to increase chances for funding success) or topic-specific funding calls. This inspiration, curiosity, or uncertainty then needs to be translated into an answerable question [19-21]. Although we found little guidance in the literature on how to operationalize this step, it is often recommended to check the research question against the FINER (feasible, interesting, novel, ethical, and relevant) and SMART (specific, measurable, achievable, realistic, and timely) quality attributes to ensure its suitability for testing in a research study [16,20,22,23].

The wording of the research question itself may already imply a specific research task (eg, exploratory, confirmatory, or qualitative research) [21]. We differentiate between research aim and research objective. Research aim is the overall goal of the research, whereas research objective is the specific goal linked to a research question [16]. Having clarity on the research task will also facilitate the identification of appropriate reporting guidelines, such as STROBE (for observational studies), CONSORT (Consolidated Standards of Reporting Trials; for randomized controlled studies), or TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis; for the development of prediction models). These reporting guidelines primarily intend to guide the communication of study results but can also be useful in converting the research question into a study design.

Ultimately, decisions regarding the study design should be guided by the research question, while also considering practical

limitations and available means and resources [23]. Frameworks such as PICOT (population, intervention, comparison, outcome, and time) and similar tools (eg, SPIDER [sample, phenomenon of interest, design, evaluation, research type] for qualitative research) [9,20,24] provide useful starting points for defining the study design. We include both the FINER and PICOT tools and their equivalents to ensure the best possible quality of the research question. Some research studies have shown that using only PICOT might be suboptimal [23]. The PICOT framework is frequently applied in health research, as PICOT is already defined above [21]. Further high-level study design decisions concern the study duration and measurement frequency (longitudinal vs cross-sectional studies), the allocation of study participants into comparator groups (randomization vs “as is” in observational research), as well as numerous practical aspects concerning study execution (eg, sample size and methods of data collection and analysis) [21,24]. Study design and study execution also need to adhere to the principles and criteria of good research and research practice to achieve valid, reliable, and accurate results [25]. Moreover, the research must comply with the standards of objectivity, reproducibility, and research integrity [26,27].

Overall, the workshop discussions confirmed that the workflow summary (Figure 2) represents a useful starting point for interdisciplinary collaborations to illustrate the established practices and to explore conceptual differences between health research, data science, and other scientific disciplines.

Developing Recommendations for an Adapted Workflow

Building on the proposed example scenario for using digital data in health research (Multimedia Appendix 3) and the established workflow description from step 1, the participants then discussed 2 types of workflow adaptations to better reflect practices and approaches from data science (Figure 2). These included the following: (1) structural modifications by changing the sequence of workflow steps (ie, introducing additional steps that should become standard in a novel workflow—the “what”) and (2) the need for introducing additional contextual constraints

or novel quality criteria (ie, modifications that do not change the workflow but may impact their execution—the “how”).

Modifications to the “What”: The Steps in Research Questions Workflow

Overall, the workshop participants perceived that the general sequence of the established workflow (the “what”) still applies to studies using (structured and unstructured) digital data. Complementary steps with their potential pitfalls were proposed to better reflect the additional challenges of working with digital unstructured data (Multimedia Appendix 4).

First, for unstructured data, preprocessing and feature extraction should be allocated a distinct workflow step to emphasize the need for thorough consideration during study planning and execution, to ensure that the data are usable, credible, and useful for the research question at hand [28-31]. On one hand, the assessment of data quality and validity is more challenging. On the other hand, preprocessing and feature extraction through machine learning require additional assumptions and may lead to predictions and derived parameters with uncertain distributional characteristics (eg, normal ranges) or propagation of algorithmic errors and biases.

Second, the selection of appropriate analysis methods to address the research question as a new workflow step would underscore the importance of scientific rigor [1,31-38]. For example, deciding between pretrained deep learning models requires preliminary investigations about the model features and the training database, which goes beyond the choice of more standard statistical techniques (eg, regression models) [39,40].

Finally, the general importance of efforts to render science reproducible and transparent was identified as a new step in the workflow.

Modifications of the “How” of the Research Question Workflow

The workshop group identified potential contextual and methodological changes to research practices (the “how”; Table 2). These proposed changes can be grouped into 3 clusters.

Table 2. Proposed changes to the “how” parts of the research question formulation workflow.

Change number	Contextual constraints and quality criteria	Description	Steps this is applicable to
I	Consider real-life incentives and constraints in defining research problems	<ul style="list-style-type: none"> The decision about RQ^a can be strongly influenced by other nonacademic factors such as the availability of funding or data. 	RQ
II	Acknowledge feasibility and resource constraints	<ul style="list-style-type: none"> The choice of research design and data analysis tools involves costs that must be considered, particularly to ensure compliance with scientific integrity. 	Research design
III	Declare limitations in RQ scope	<ul style="list-style-type: none"> Each RQ has limitations; it is important to define what RQ can and cannot answer. 	RQ
IV	Allow for and document iterations in RQ development and analysis	<ul style="list-style-type: none"> Proper documentation is important for ensuring transparency and helps with evaluating and tracking the decisions regarding the iterations in RQ. 	RQ
V	Acknowledge and respond to the increasing need for interdisciplinary expertise	<ul style="list-style-type: none"> For the feasibility of the RQ, it is important to consider the needed expertise and skills. This becomes particularly important in research involving digital unstructured data as it requires an interdisciplinary set of skills. 	All steps
VI	Enhance reproducibility	<ul style="list-style-type: none"> Reproducibility in data science means obtaining consistent results using the same input data and methods. On a higher level, reproducibility in science also refers to the ability to duplicate findings if the same methods are used [41]. Reproducibility in science also refers to the concept of making data; computational steps, methods, and codes; and conditions of analysis transparent and available, so that others can verify the findings. 	All steps
VII	Enhance replicability	<ul style="list-style-type: none"> Replicability refers to applying the same methods from a different study on different data. Observed differences in findings should be explicable by data-specific differences between studies. 	All steps
VIII	Enhance robustness	<ul style="list-style-type: none"> Robustness refers to analyses that apply the same database but use different methods. Observed differences in findings should be explicable by method-specific differences between studies. Within the same study, robustness is often evaluated by sensitivity analyses that use the same data but vary methods (eg, by applying different model parameters). 	All steps
IX	Critically assess generalizability	<ul style="list-style-type: none"> Generalizability means that the study results or outcomes are also applicable in other study settings and samples. 	All steps

^aRQ: research question.

The first cluster includes the acknowledgment of what we have labeled as real-life constraints (change numbers I and II) and limitations in the scope of research questions (change number III). Appropriately addressing such real-life constraints can be fostered by greater transparency and experience exchange.

The second cluster of proposed contextual changes pertains to enabling more interdisciplinary and iterative workflows (change numbers IV and V). Reasons for iterative approaches include more complex choices of analysis methods, the need for verifying the validity and robustness of model results, or the need to manually search for the best model parametrization. These challenges also require a greater emphasis on interdisciplinary collaborations that combine subject-domain knowledge and data science expertise.

The final cluster reflects the need for strengthening research quality criteria to foster open science, better reproducibility,

and greater transparency (change numbers VI-IX). As analytical methods and databases become more complex, there is also an increasing need for transparency; adequate documentation; as well as publicly available analysis protocols, software codes, data, and analysis files. Studies should critically examine their findings under changing data or method combinations, thus exploring reproducibility, robustness, replicability, or generalizability (the 3RG criteria) and enhancing the overall quality of research. An important means to achieve these goals are open science and Findable, Accessible, Interoperable, and Reusable (FAIR) data principles [42].

Recommendations Toward a Pragmatic Approach of Teaching and Conducting Research Question Formulation

The workshop discussions produced a set of specific recommendations to promote approaches for defining good

research questions for reusing digital data. These recommendations are also well suited for educational use, helping to navigate the challenges of interdisciplinary collaboration and to foster interdisciplinary skills.

Iterative Research Question Formulation

As a principle, data collection, preprocessing, and analysis methods should follow the research question; researchers should not lose sight of the research aim, objective, and question. Defining a good research question is a fundamental and universal first step of science, which ideally should not be preceded by the choice of data or methods. However, the linear process of defining a research question common to health research may need several iterations to ensure that the complexity and feasibility of reusing and integrating digital data are accounted for.

The lecture instructors, supervisors, and principal investigators of interdisciplinary projects can apply this recommendation by emphasizing the importance of research question formulation in interdisciplinary projects. Furthermore, they can facilitate a discussion or create exercises for students to practice how the linear process of research question definition changes into a more iterative process when collaborating with other disciplines.

Reconciling Linear And Iterative Approaches: Continuum of Research Tasks

To reconcile the apparent conceptual differences between health research and data science approach research projects, we propose to reframe the scientific process as a continuum of knowledge accumulation over the course of multiple studies. Such a continuum can consist of several different research tasks (projects) combining deductive and inductive research approaches. Not all research tasks will involve explicit research questions or hypotheses. However, systematic reflections on how study results can inform new hypotheses and research questions and how they could be tested in future studies could become an integral part of a study, for example, as a last step in exploratory analyses.

Lecture instructors, supervisors, and principal investigators of interdisciplinary projects can use this recommendation to emphasize the continuum of research tasks. They can create exercises consisting of different research tasks where students practice combining deductive and inductive approaches in research design. These exercises can guide students to recognize that research is not always a straightforward process of hypothesis testing but may involve exploratory tasks that inform future studies. Instructors can also encourage students to reflect systematically on their research results, guiding them to think about how current findings can shape future hypotheses and research directions. This reflection can be incorporated into project work, where students work on iterative research tasks, examining how knowledge accumulates across studies and how inductive and deductive methods interact throughout this process. This practice prepares students to handle the nonlinear nature of interdisciplinary research, especially when bridging health research and data science.

Research Quality Criteria

The complexities involved in digital data preprocessing and analysis require careful design decisions and thorough reporting to ensure adherence to research quality standards. The reuse of existing, digital unstructured data and the need for extensive preprocessing may obfuscate or compound issues of external and internal validity [14]. Moreover, the use of machine learning techniques such as deep neural networks may generate “unexplainable” predictions or classifications that challenge the transparency and open science paradigms. The verification of “whether the data measure what they are supposed to measure (in the context of the research question)” [14] remains crucial and deserves appropriate attention, but it may become more difficult to achieve. We recommend that researchers systematically scrutinize interim results to ensure that they are “on the good track.” Such checks can, for example, include the replication of results from different studies. Furthermore, transparency in reporting and reproducibility are key to scientific rigor.

Lecture instructors and supervisors can emphasize the importance of maintaining research quality in interdisciplinary projects. They can design exercises where students practice making careful design decisions in their research projects, ensuring that issues of validity, transparency, and reproducibility are addressed throughout the process. One approach could be to guide students in developing protocols for systematic checks of their interim results. Instructors can also promote transparency by teaching students how to document their research processes thoroughly, facilitating reproducibility and open science principles. By applying these exercises, students learn to critically evaluate the quality of their research.

Take Active Measures to Foster Interdisciplinarity

We recommend reflecting these aspects appropriately in teaching and training of next-generation researchers as well as in establishing new interdisciplinary research groups or collaborations. Therefore, in teaching, it is important to also convey a realistic view of how research works in practice. Students should be sensitized to real-world challenges and the need for pragmatic decision-making, while still striving for the basic principles of “good research practices.” The literature review and our own experiences suggest that students are mostly taught the “ideal model,” and thus, they are often not well prepared for the realities of research. It seems preferable to discuss challenges openly and to expose students to ethical and practical dilemmas early on.

Lecture instructors and supervisors can sensitize students toward real-world challenges. They can prepare specific exercises where students can reflect on the problems that might arise from real-life constraints.

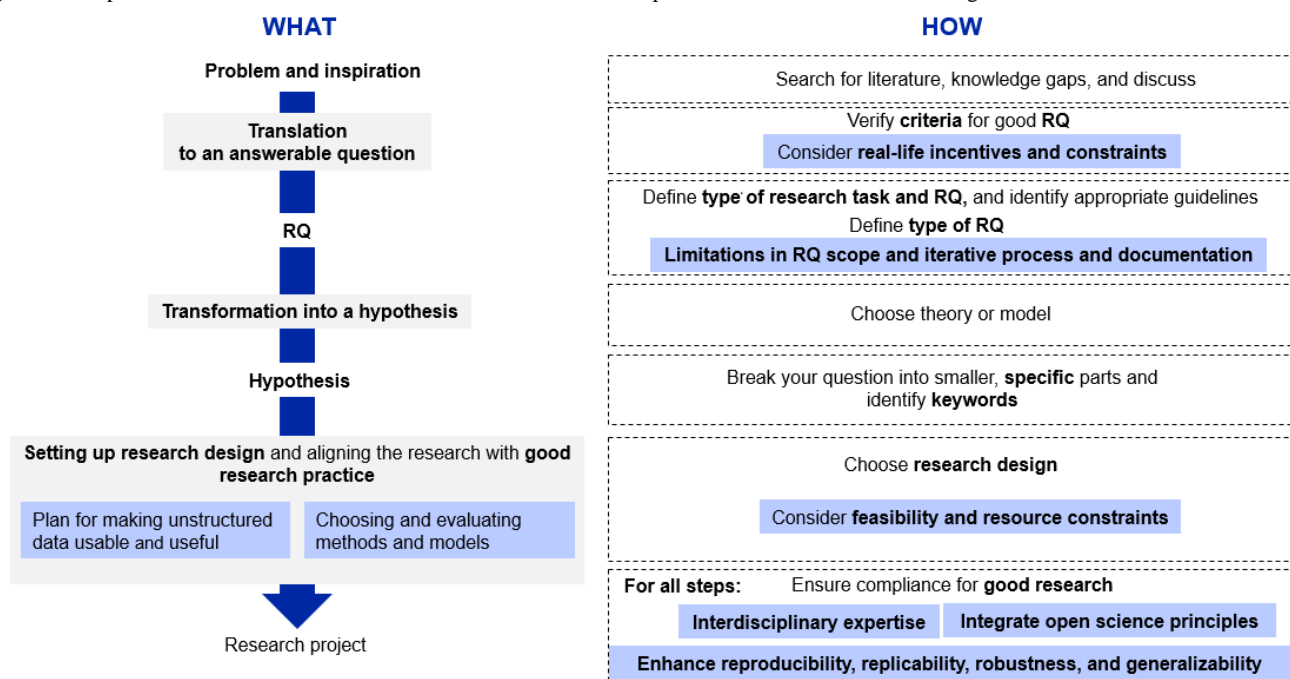
The added complexity and additional financial needs for education for interdisciplinary collaboration and open science should be acknowledged by funding agencies.

Specific Tools to Inform the Teaching of Interdisciplinary Courses on Real-World Data Analyses

Our study provides practical tools to guide the content and curricula of courses focused on interdisciplinary projects and collaborations. A more detailed description of the application of the study results to teaching is provided in [Multimedia Appendix 5](#). The structure of our workshops ([Figure 1](#)) and the results of each workshop can be directly translated into the tools focusing on both dimensions of *content and curriculum* as well as *methods and teaching styles*. In terms of *content and curriculum*, the glossary with key concepts and terminology can be used to introduce students to interdisciplinary work. The workflow ([Figure 2](#)) combined with the glossary can serve as an interdiction to research practices for students with a different

background, for example, humanities. Finally, our adapted workflow ([Figure 3](#)) sensitizes students for additional topics of transparency, FAIR data, reproducibility, and open science. Regarding *methods and teaching styles*, the sequence of workshops ([Figure 1](#)) and their results as outlined in the *content and curriculum section* can be directly translated into teaching phases, which build on top of each other. As illustrated by the example described in [Multimedia Appendix 5](#), the structure of 3 teaching phases is useful and effective for teaching interdisciplinary research collaborations. A key insight from our workshop (modifications to the “how”—cluster 1) consisted of the need to acknowledge and address real-world challenges in study planning and execution. In our experience, case studies and illustrations of the scientific process of real-world examples are greatly appreciated by students.

Figure 3. Adapted workflow. The blue-colored boxes contain the added aspects in the “what” and “how” categories.



Discussion

Principal Findings

This study examined how interdisciplinary research collaborations between health research and data science can be streamlined by creating a shared conceptual understanding of terminologies and best practice workflows and by acknowledging or merging approaches from other disciplines. In a series of interactive workshops, our interdisciplinary group of coauthors concluded that the workflow of established practices for formulating a research question, generating hypotheses, and defining research designs remain valid. We argue that the reuse of digital data does not substantially change scientific activity, particularly the fundamental step of defining a good research question [1,43]. Achieving clarity on the research question benefits data analysis and interpretation by providing structure and informing the study design workflow. Moreover, a shared understanding of the research question and study workflow facilitates the inclusion of diverse domain

knowledge to ensure research quality and result quality [6,14,44]. In line with this, the group noted general tendencies in research toward more open, transparent, and reproducible research, which are influenced by recommended data science practices. Along those lines, health research should increasingly foster good scientific practices that help to align the reuse of digital data with principles of reproducibility, robustness, generalizability, validity [1,32], transparency, and open science.

Our set of tools and recommendations can also be integrated into medical education by providing academic educators with a structured approach to teaching research question formulation in the context of using digital data in health research. By emphasizing the importance of both hypothesis-driven and data-driven research methods, educators can guide researchers in navigating the interdisciplinary challenges of health research and data science. The importance of creating a common terminology and discussion about scientific principles can further increase awareness about the challenges of interdisciplinary collaboration between health researchers and

data scientists. The proposed workflow and recommendations equip researchers with the tools to address the challenges of research question definition for interdisciplinary projects. The clear and practical steps provided by the workflow ensure that students not only grasp theoretical concepts but also apply them effectively in real-world scenarios, preparing them for collaborative, data-driven environments in health care and research.

For implementation, the set of tools and recommendations could be integrated into medical curricula and PhD programs through dedicated courses, workshops, or modules focusing on research methods and interdisciplinary collaboration for young researchers. Medical educators can adopt these recommendations to structure class discussions, assignments, and group projects, ensuring that students are exposed to both research approaches. In [Multimedia Appendix 5](#), we provide an example of an interdisciplinary course implementing this set of tools. To evaluate the effectiveness of this implementation, a combination of qualitative and quantitative assessments can be used. Surveys and feedback from both students and educators can measure how well the workflow improves understanding and application of interdisciplinary research question formulation.

Our interdisciplinary effort recognized and discussed several potential obstacles toward bridging the approaches of established health research and data science. In the following sections, we repeat 4 key insights from our workshop interactions on how such obstacles can be overcome. First, we noted substantial differences in the use of terminologies across disciplines. For interdisciplinary collaborations, it is important to clarify key terms and concepts early on and to develop a shared understanding of the research aim and research question.

Second, in the early stages of the project, workshop participants expressed confusion about different types of analysis methods and their relationship with specific research tasks and high-level aims, such as prediction and classification, confirmatory research, or exploratory research. Agreeing on the high-level conceptual framework of “research tasks” helped structure the workshop discussions effectively. The discussion around the concepts of “research task” also fostered insights about commonalities and overlaps between concepts of data and health research. For example, many data science tasks can be classified as exploratory or prediction or classification tasks, which have conceptual counterparts in health research methodologies, each with corresponding reporting quality guidelines. Referring to specific research tasks rather than making global statements about data science or health research resonated well with the workshop participants and facilitated the discussions considerably.

Third, by introducing the concept of a “research task,” the group was also better able to examine the relationships among research aims, objectives, and tasks and how they are reflected in the workflow of established practices. Participants believed that exploratory or prediction or classification tasks, in particular, did not fit well into the workflow because such work is often not strictly hypothesis driven. However, 2 insights helped to align the workflow framework with the task concept: answering a research question may involve multiple research tasks in the

same analysis, such as using prediction and classification tasks for data preprocessing, and later using these predictions in a confirmatory analysis, for example, as an exposure variable. Moreover, the scientific process can be viewed as a continuum of studies. From this perspective, the workflow of established practices can also be seen as a higher-level discovery cycle that spans across multiple studies. For example, an initial study may explore initial exploratory hypotheses or generate a first iteration of a prediction model, thus leading to new hypotheses. Indeed, exploratory and inductive methods can be useful to keep an open mind and become inspired by empirical data. In this way, the research tasks can be seen as a continuum—where data-driven research ends, hypothesis-driven research can start. Follow-up studies could then explore the hypotheses or validate the prediction model (whose structure can also be considered a hypothesis) in new data or in a confirmatory analysis. In combination, these multiple research tasks or study sequences are still likely to conform to the proposed workflow of recommended practices.

Finally, reusing unstructured and structured digital data brings new ethical challenges, such as privacy and consent issues, and problems with (public) trust and data diversity [45-49]. Traditional ethical assessments for data use in research and ethics review committees might not be well suited to address the challenges of digital data and might need adaptations [45,50]. Weighing the potential benefits and risks of using digital data becomes more complex. This problem is accentuated because the availability and production of digital data are often not based on a scientific decision, and rather, other factors such as political or social phenomena play a role [1]. While the need for novel ethical mechanisms to guide researchers is to be found in recently developed self-assessment tools for ethical data use [51,52], these new ethical mechanisms need further refinement to be widely adopted.

Strengths and Limitations

The strength of the expert groups was that participants represented a diverse group in terms of disciplines and career stages. However, it is possible that not all potentially relevant viewpoints were represented. A further strength was that the inputs from experts were collected systematically via different channels (eg, discussions, preworkshop tasks, and commenting on documents) throughout the consensus process. This allowed to harmonize and synthesize knowledge and insights from diverse disciplines. Experts also had several opportunities to review discussion outcomes and final summaries through workshop protocol and involvement in manuscript writing.

There are limitations regarding our proposed workflow. First, it represents an idealized process for defining a good research question, which is often challenged by funding and resource constraints or established norms. Some parts of the workflow might not be explicitly applicable to all types of research. The example scenario used to develop the workflow was based on hypothesis-driven deductive research, which often uses relational and causal research questions. We did not explicitly include inductive, qualitative approaches in health research, but we see the deductive and inductive research on a spectrum [53]. This limitation does not prevent the overall concept of the workflow

from being applied to other types of research, such as inductive, data-driven, or exploratory research.

Finally, although the literature review was conducted with great care and the expert group included several experienced researchers and faculty from different scientific disciplines, it was not possible to conduct a fully systematic search across all research disciplines due to resource constraints. Therefore, it is possible that some potentially relevant concepts and guidelines were not included.

Conclusions

In an age of digital transformation, established scientific practices with a strong focus on formulating research question design remain relevant and useful for gaining clarity about research aims. We recommend initiating new collaborations in

the health domain with a review of terminologies and concepts to avoid misconceptions and problems further downstream in the research process. Our terminology and workflow may serve as tools to be used in medical education to support young and established researchers in interdisciplinary health research projects. To this end, we found the concept of “research tasks” particularly useful to foster a shared understanding among our collaborators. In addition, we recommend adapting the way the established workflow is taught to prospective researchers in health research and other disciplines, incorporating concepts from open science, the 3RG criteria, and the “science as a continuum” paradigm. We also call for funding agencies and publishers to incentivize and acknowledge investments in defining good research questions for complex novel data and analysis methods.

Data Availability

All data are available in the manuscript and multimedia appendices.

Authors' Contributions

JS contributed to conceptualization, data curation, investigation, methodology, project administration, resources, visualization, and writing the original draft. VvW contributed to the conceptualization, investigation, methodology, and writing the original draft. All the other authors contributed to writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study and workshop protocols and preparatory tasks.

[\[PDF File \(Adobe PDF File\), 1792 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Accurate Consensus Reporting Document (ACCORD) checklist.

[\[PDF File \(Adobe PDF File\), 127 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Example scenario of digital data reuse to structure workshop discussions.

[\[PDF File \(Adobe PDF File\), 20 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Proposed modifications to the “what”: additional steps to the workflow.

[\[PDF File \(Adobe PDF File\), 118 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Application of teaching tools: example from the course Interactive Data Science in Digital Health.

[\[DOCX File , 36 KB-Multimedia Appendix 5\]](#)

References

1. Caliebe A, Leverkus F, Antes G, Krawczak M. Does big data require a methodological change in medical research? *BMC Med Res Methodol*. Jun 17, 2019;19(1):125. [\[FREE Full text\]](#) [doi: [10.1186/s12874-019-0774-0](https://doi.org/10.1186/s12874-019-0774-0)] [Medline: [31208367](https://pubmed.ncbi.nlm.nih.gov/31208367/)]
2. Chiavilli M, Campagnini S, Baretta T, Castagnoli C, Paperini A, Politi AM, et al. Design and implementation of a Stroke Rehabilitation Registry for the systematic assessment of processes and outcomes and the development of data-driven prediction models: the STRATEGY study protocol. *Front Neurol*. Oct 10, 2022;13:919353. [\[FREE Full text\]](#) [doi: [10.3389/fneur.2022.919353](https://doi.org/10.3389/fneur.2022.919353)] [Medline: [36299268](https://pubmed.ncbi.nlm.nih.gov/36299268/)]

3. Sim I. Mobile devices and health. *N Engl J Med*. Sep 05, 2019;381(10):956-968. [doi: [10.1056/NEJMr1806949](https://doi.org/10.1056/NEJMr1806949)] [Medline: [31483966](https://pubmed.ncbi.nlm.nih.gov/31483966/)]
4. Batko K, Ślęzak A. The use of big data analytics in healthcare. *J Big Data*. 2022;9(1):3. [FREE Full text] [doi: [10.1186/s40537-021-00553-4](https://doi.org/10.1186/s40537-021-00553-4)] [Medline: [35013701](https://pubmed.ncbi.nlm.nih.gov/35013701/)]
5. Sedlakova J, Daniore P, Horn Wintsch A, Wolf M, Stanikic M, Haag C, et al. Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review. *PLOS Digit Health*. Oct 11, 2023;2(10):e0000347. [FREE Full text] [doi: [10.1371/journal.pdig.0000347](https://doi.org/10.1371/journal.pdig.0000347)] [Medline: [37819910](https://pubmed.ncbi.nlm.nih.gov/37819910/)]
6. Kitchin R. Big data, new epistemologies and paradigm shifts. *Big Data Soc*. Apr 01, 2014;1(1). [doi: [10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481)]
7. Goldsmith J, Sun Y, Fried LP, Wing J, Miller GW, Berhane K. The emergence and future of public health data science. *Public Health Rev*. Apr 26, 2021;42:1604023. [FREE Full text] [doi: [10.3389/phrs.2021.1604023](https://doi.org/10.3389/phrs.2021.1604023)] [Medline: [34692178](https://pubmed.ncbi.nlm.nih.gov/34692178/)]
8. Mirin N, Mattie H, Jackson L, Samad Z, Chunara R. Data science in public health: building next generation capacity. *Harv Data Sci Rev*. Oct 27, 2022;4(4). [doi: [10.1162/99608f92.18da72db](https://doi.org/10.1162/99608f92.18da72db)]
9. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qual Health Res*. Oct 2012;22(10):1435-1443. [doi: [10.1177/1049732312452938](https://doi.org/10.1177/1049732312452938)] [Medline: [22829486](https://pubmed.ncbi.nlm.nih.gov/22829486/)]
10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. Apr 2008;61(4):344-349. [doi: [10.1016/j.jclinepi.2007.11.008](https://doi.org/10.1016/j.jclinepi.2007.11.008)] [Medline: [18313558](https://pubmed.ncbi.nlm.nih.gov/18313558/)]
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Mar 29, 2021;372:n71. [FREE Full text] [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
12. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med*. Oct 16, 2007;4(10):e297. [FREE Full text] [doi: [10.1371/journal.pmed.0040297](https://doi.org/10.1371/journal.pmed.0040297)] [Medline: [17941715](https://pubmed.ncbi.nlm.nih.gov/17941715/)]
13. Donaldson L, Ricciardi W, Sheridan S, Tartaglia R. *Textbook of Patient Safety and Clinical Risk Management*. Cham, Switzerland. Springer International Publishing; 2020.
14. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. *NPJ Digit Med*. Jun 3, 2019;2:45. [FREE Full text] [doi: [10.1038/s41746-019-0121-1](https://doi.org/10.1038/s41746-019-0121-1)] [Medline: [31304391](https://pubmed.ncbi.nlm.nih.gov/31304391/)]
15. Gattrell WT, Logullo P, van Zuuren EJ, Price A, Hughes EL, Blazey P, et al. ACCORD (ACCurate CONsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLoS Med*. Jan 23, 2024;21(1):e1004326. [FREE Full text] [doi: [10.1371/journal.pmed.1004326](https://doi.org/10.1371/journal.pmed.1004326)] [Medline: [38261576](https://pubmed.ncbi.nlm.nih.gov/38261576/)]
16. Doody O, Bailey ME. Setting a research question, aim and objective. *Nurse Res*. Mar 2016;23(4):19-23. [doi: [10.7748/nr.23.4.19.s5](https://doi.org/10.7748/nr.23.4.19.s5)] [Medline: [26997231](https://pubmed.ncbi.nlm.nih.gov/26997231/)]
17. Creswell JW. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 3rd ed. Thousand Oaks, CA. SAGE Publications; 2012.
18. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. Mar 31, 2009;338:b604. [doi: [10.1136/bmj.b604](https://doi.org/10.1136/bmj.b604)] [Medline: [19336487](https://pubmed.ncbi.nlm.nih.gov/19336487/)]
19. Lipowski EE. Developing great research questions. *Am J Health Syst Pharm*. Sep 01, 2008;65(17):1667-1670. [doi: [10.2146/ajhp070276](https://doi.org/10.2146/ajhp070276)] [Medline: [18714115](https://pubmed.ncbi.nlm.nih.gov/18714115/)]
20. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. *Indian J Sex Transm Dis AIDS*. Jan 2010;31(1):47-50. [FREE Full text] [doi: [10.4103/0253-7184.69003](https://doi.org/10.4103/0253-7184.69003)] [Medline: [21808439](https://pubmed.ncbi.nlm.nih.gov/21808439/)]
21. Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. *Can J Hosp Pharm*. Jan 2014;67(1):31-34. [FREE Full text] [doi: [10.4212/cjhp.v67i1.1320](https://doi.org/10.4212/cjhp.v67i1.1320)] [Medline: [24634524](https://pubmed.ncbi.nlm.nih.gov/24634524/)]
22. Fandino W. Formulating a good research question: pearls and pitfalls. *Indian J Anaesth*. Aug 2019;63(8):611-616. [FREE Full text] [doi: [10.4103/ija.IJA_198_19](https://doi.org/10.4103/ija.IJA_198_19)] [Medline: [31462805](https://pubmed.ncbi.nlm.nih.gov/31462805/)]
23. Abbade LP, Wang M, Sriganesh K, Jin Y, Mbuagbaw L, Thabane L. The framing of research questions using the PICOT format in randomized controlled trials of venous ulcer disease is suboptimal: a systematic survey. *Wound Repair Regen*. Sep 2017;25(5):892-900. [doi: [10.1111/wrr.12592](https://doi.org/10.1111/wrr.12592)] [Medline: [29080311](https://pubmed.ncbi.nlm.nih.gov/29080311/)]
24. Booth A. Clear and present questions: formulating questions for evidence based practice. *Libr Hi Tech*. 2006;24(3):355-368. [doi: [10.1108/07378830610692127](https://doi.org/10.1108/07378830610692127)]
25. Mehta A, Malley B, Walkey A. *Formulating the research question*. In: *Secondary Analysis of Electronic Health Records*. Cham, Switzerland. Springer; 2016.
26. Sørensen MP, Ravn T, Marušić A, Elizondo AR, Kavouras P, Tjeldink J, et al. Strengthening research integrity: which topic areas should organisations focus on? *Humanit Soc Sci Commun*. Aug 12, 2021;8(1):1-15. [doi: [10.1057/s41599-021-00874-y](https://doi.org/10.1057/s41599-021-00874-y)]
27. Quiroga Gutierrez AC, Lindegger DJ, Taji Heravi A, Stojanov T, Sykora M, Elayan S, et al. Reproducibility and scientific integrity of big data research in urban public health and digital epidemiology: a call to action. *Int J Environ Res Public Health*. Jan 13, 2023;20(2):1473. [FREE Full text] [doi: [10.3390/ijerph20021473](https://doi.org/10.3390/ijerph20021473)] [Medline: [36674225](https://pubmed.ncbi.nlm.nih.gov/36674225/)]

28. Kandel S, Heer J, Plaisant C, Kennedy J, van Ham F, Riche NH, et al. Research directions in data wrangling: visualizations and transformations for usable and credible data. *Inf Vis*. Sep 02, 2011;10(4):271-288. [doi: [10.1177/1473871611415994](https://doi.org/10.1177/1473871611415994)]
29. Munzner T. *Visualization Analysis and Design*. Boca Raton, FL. CRC Press; 2014.
30. Fekete JD. Visual analytics infrastructures: from data management to exploration. *Computer*. Jul 2013;46(7):22-29. [doi: [10.1109/MC.2013.120](https://doi.org/10.1109/MC.2013.120)]
31. Arbesser C, Spechtenhauser F, Muhlbacher T, Piringer H. Visplause: visual data quality assessment of many time series using plausibility checks. *IEEE Trans Visual Comput Graphics*. Jan 2017;23(1):641-650. [doi: [10.1109/tvcg.2016.2598592](https://doi.org/10.1109/tvcg.2016.2598592)]
32. Cerreta F, Ritzhaupt A, Metcalfe T, Askin S, Duarte J, Berntgen M, et al. Digital technologies for medicines: shaping a framework for success. *Nat Rev Drug Discov*. Sep 2020;19(9):573-574. [doi: [10.1038/d41573-020-00080-6](https://doi.org/10.1038/d41573-020-00080-6)] [Medline: [32398879](https://pubmed.ncbi.nlm.nih.gov/32398879/)]
33. Sedlmair M, Meyer M, Munzner T. Design study methodology: reflections from the trenches and the stacks. *IEEE Trans Visual Comput Graphics*. Dec 2012;18(12):2431-2440. [doi: [10.1109/tvcg.2012.213](https://doi.org/10.1109/tvcg.2012.213)]
34. Muhlbacher T, Piringer H, Gratzl S, Sedlmair M, Streit M. Opening the black box: strategies for increased user involvement in existing algorithm implementations. *IEEE Trans Visual Comput Graphics*. Dec 31, 2014;20(12):1643-1652. [doi: [10.1109/tvcg.2014.2346578](https://doi.org/10.1109/tvcg.2014.2346578)]
35. Schreck T, Bernard J, von Landesberger T, Kohlhammer J. Visual cluster analysis of trajectory data with interactive Kohonen maps. *Inf Vis*. Feb 12, 2009;8(1):14-29. [doi: [10.1057/ivs.2008.29](https://doi.org/10.1057/ivs.2008.29)]
36. Lipton ZC. The doctor just won't accept that! arXiv. Preprint posted online on November 20, 2017. [FREE Full text] [doi: [10.48550/arXiv.1711.08037](https://doi.org/10.48550/arXiv.1711.08037)]
37. Bernard J, von Landesberger T, Bremm S, Schreck T. Multi-scale visual quality assessment for cluster analysis with self-organizing maps. In: *Proceedings of SPIE - The International Society for Optical Engineering*. 2011. Presented at: SPIE 2011; 2011; Bellingham, WA. [doi: [10.1117/12.872545](https://doi.org/10.1117/12.872545)]
38. Munzner T. A nested model for visualization design and validation. *IEEE Trans Vis Comput Graph*. Nov 2009;15(6):921-928. [doi: [10.1109/TVCG.2009.111](https://doi.org/10.1109/TVCG.2009.111)] [Medline: [19834155](https://pubmed.ncbi.nlm.nih.gov/19834155/)]
39. de Hond AA, Leeuwenberg AM, Hooft L, Kant IM, Nijman SW, van Os HJ, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *NPJ Digit Med*. Jan 10, 2022;5(1):2. [FREE Full text] [doi: [10.1038/s41746-021-00549-7](https://doi.org/10.1038/s41746-021-00549-7)] [Medline: [35013569](https://pubmed.ncbi.nlm.nih.gov/35013569/)]
40. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ*. May 18, 2022;377:e070904. [FREE Full text] [doi: [10.1136/bmj-2022-070904](https://doi.org/10.1136/bmj-2022-070904)] [Medline: [35584845](https://pubmed.ncbi.nlm.nih.gov/35584845/)]
41. Schwab S, Janiaud P, Dayan M, Amrhein V, Panczak R, Palagi PM, et al. Ten simple rules for good research practice. *PLoS Comput Biol*. Jun 23, 2022;18(6):e1010139. [FREE Full text] [doi: [10.1371/journal.pcbi.1010139](https://doi.org/10.1371/journal.pcbi.1010139)] [Medline: [35737655](https://pubmed.ncbi.nlm.nih.gov/35737655/)]
42. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. Mar 15, 2016;3(1):160018. [FREE Full text] [doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)] [Medline: [26978244](https://pubmed.ncbi.nlm.nih.gov/26978244/)]
43. Mazzocchi F. Could big data be the end of theory in science? A few remarks on the epistemology of data-driven science. *EMBO Rep*. Oct 10, 2015;16(10):1250-1255. [FREE Full text] [doi: [10.15252/embr.201541001](https://doi.org/10.15252/embr.201541001)] [Medline: [26358953](https://pubmed.ncbi.nlm.nih.gov/26358953/)]
44. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. *Front Commun*. Feb 27, 2020;5. [doi: [10.3389/fcomm.2020.00011](https://doi.org/10.3389/fcomm.2020.00011)]
45. Ferretti A, Ienca M, Sheehan M, Blasimme A, Dove ES, Farsides B, et al. Ethics review of big data research: what should stay and what should be reformed? *BMC Med Ethics*. Apr 30, 2021;22(1):51. [FREE Full text] [doi: [10.1186/s12910-021-00616-4](https://doi.org/10.1186/s12910-021-00616-4)] [Medline: [33931049](https://pubmed.ncbi.nlm.nih.gov/33931049/)]
46. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Med*. Jul 17, 2019;17(1):137. [FREE Full text] [doi: [10.1186/s12916-019-1377-7](https://doi.org/10.1186/s12916-019-1377-7)] [Medline: [31311535](https://pubmed.ncbi.nlm.nih.gov/31311535/)]
47. Facca D, Smith MJ, Shelley J, Lizotte D, Donelle L. Exploring the ethical issues in research using digital data collection strategies with minors: a scoping review. *PLoS One*. 2020;15(8):e0237875. [FREE Full text] [doi: [10.1371/journal.pone.0237875](https://doi.org/10.1371/journal.pone.0237875)] [Medline: [32853218](https://pubmed.ncbi.nlm.nih.gov/32853218/)]
48. Clark RA, Foote J, Versace VL, Brown A, Daniel M, Coffee NT, et al. The keeping on track study: exploring the activity levels and utilization of healthcare services of acute coronary syndrome (ACS) patients in the first 30-days after discharge from hospital. *Med Sci (Basel)*. Apr 19, 2019;7(4):61. [FREE Full text] [doi: [10.3390/medsci7040061](https://doi.org/10.3390/medsci7040061)] [Medline: [31010168](https://pubmed.ncbi.nlm.nih.gov/31010168/)]
49. Arora A, Alderman JE, Palmer J, Ganapathi S, Laws E, McCradden MD, et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med*. Nov 26, 2023;29(11):2929-2938. [FREE Full text] [doi: [10.1038/s41591-023-02608-w](https://doi.org/10.1038/s41591-023-02608-w)] [Medline: [37884627](https://pubmed.ncbi.nlm.nih.gov/37884627/)]
50. Viberg Johansson J, Bentzen HB, Mascalonzi D. What ethical approaches are used by scientists when sharing health data? An interview study. *BMC Med Ethics*. Apr 11, 2022;23(1):41. [FREE Full text] [doi: [10.1186/s12910-022-00779-8](https://doi.org/10.1186/s12910-022-00779-8)] [Medline: [35410285](https://pubmed.ncbi.nlm.nih.gov/35410285/)]
51. Wiltshire D, Alvanides S. Ensuring the ethical use of big data: lessons from secure data access. *Heliyon*. Feb 2022;8(2):e08981. [FREE Full text] [doi: [10.1016/j.heliyon.2022.e08981](https://doi.org/10.1016/j.heliyon.2022.e08981)] [Medline: [35243099](https://pubmed.ncbi.nlm.nih.gov/35243099/)]

52. Bigdata-ethics-HeALth framework - BEHALF. Eidgenössische Technische Hochschule Zürich. URL: <https://bioethics.ethz.ch/research/BEHALF.html> [accessed 2022-11-28]
53. Young M, Varpio L, Uijtdehaage S, Paradis E. The spectrum of inductive and deductive research approaches using quantitative and qualitative data. *Acad Med*. Jul 2020;95(7):1122. [doi: [10.1097/ACM.0000000000003101](https://doi.org/10.1097/ACM.0000000000003101)] [Medline: [31833855](https://pubmed.ncbi.nlm.nih.gov/31833855/)]

Abbreviations

3RG: reproducibility, robustness, replicability, or generalizability

ACCORD: Accurate Consensus Reporting Document

CONSORT: Consolidated Standards of Reporting Trials

DSI: Digital Society Initiative

FAIR: findable, accessible, interoperable, and reusable

FINER: feasible, interesting, novel, ethical, and relevant

PICOT: population, intervention, comparison, outcome, and time

SMART: specific, measurable, achievable, realistic, and timely

SPIDER: sample, phenomenon of interest, design, evaluation, research type

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by B Lesselroth; submitted 15.01.24; peer-reviewed by G Richter, M Mahmic Kaknjo; comments to author 12.02.24; revised version received 16.05.24; accepted 23.11.24; published 23.01.25

Please cite as:

Sedlakova J, Stanikić M, Gille F, Bernard J, Horn AB, Wolf M, Haag C, Floris J, Morgenshtern G, Schneider G, Zumbunn Wojczyńska A, Mouton Dorey C, Ettl DA, Gero D, Friemel T, Lu Z, Papadopoulos K, Schläpfer S, Wang N, von Wyl V

Refining Established Practices for Research Question Definition to Foster Interdisciplinary Research Skills in a Digital Age: Consensus Study With Nominal Group Technique

JMIR Med Educ 2025;11:e56369

URL: <https://mededu.jmir.org/2025/1/e56369>

doi: [10.2196/56369](https://doi.org/10.2196/56369)

PMID:

©Jana Sedlakova, Mina Stanikić, Felix Gille, Jürgen Bernard, Andrea B Horn, Markus Wolf, Christina Haag, Joel Floris, Gabriela Morgenshtern, Gerold Schneider, Aleksandra Zumbunn Wojczyńska, Corine Mouton Dorey, Dominik Alois Ettl, Daniel Gero, Thomas Friemel, Ziyuan Lu, Kimon Papadopoulos, Sonja Schläpfer, Ning Wang, Viktor von Wyl. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.