Data-driven hypothesis generation among junior clinical researchers: A comparison of a secondary data analysis with visualization (VIADS) and other tools

Xia Jing[1], James J. Cimino[2], Vimla L. Patel[3], Yuchun Zhou[4], Jay H. Shubrook[5], Sonsoles De Lacalle[6], Brooke N. Draghi[1], Mytchell A. Ernst[1], Aneesa Weaver[1], Shriram Sekar[7], Chang Liu[8]

[1]Department of Public Health Sciences, Clemson University, Clemson, SC; [2]Informatics Institute, School of Medicine, University of Alabama, Birmingham, Birmingham, AL; [3]Cognitive Studies in Medicine and Public Health, The New York Academy of Medicine, New York City, NY; [4]Patton College of Education, Ohio University, Athens, OH; [5]College of Osteopathic Medicine, Touro University, Vallejo, CA; [6]Department of Health Science, California State University Channel Islands, Camarillo, CA; [7]Schoole of Computing, Clemson University, Clemson, SC; [8]Russ College of Engineering and Technology, Ohio University, Athens, OH;

Abstract:

Objectives: To compare how junior clinical researchers generate data-driven hypotheses with a visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies (VIADS) or other analytical tools routinely used by participants on the same datasets.

Methods: We recruited clinical researchers from all over the United States of America and separated them into "experienced" and "inexperienced" groups using predetermined criteria. Within the groups, participants were randomly assigned to a VIADS or non-VIADS groups (control) group. We recruited two participants for the pilot study and 18 for the main study. Fifteen (out of 18) were junior clinical researchers, including seven in the control group and eight in the VIADS group. All participants used the same datasets and study scripts. Each participant conducted a remote 2-hour study session for hypothesis generation. The VIADS groups also had a 1-hour training session. The same researcher coordinated the study session. Two participants in the pilot study were one experienced and one inexperienced clinical researcher. During the session, all participants followed a think-aloud protocol to verbalize their thoughts and actions during data analysis and hypothesis generation. Follow-up surveys were administered to all participants after each study session. All screen activities and audio were recorded, transcribed, coded, and analyzed. Every ten randomly selected hypotheses were included in one Qualtrics survey for quality evaluation. Seven expert panel members rated each hypothesis on validity, significance, and feasibility.

Results: Eighteen participants generated 227 hypotheses, of which 147 (65%) were valid based on our criteria. Each participant generated between one and 19 valid hypotheses during the 2-hour session. The VIADS and control groups generated a similar number of hypotheses on average. It took the VIADS group participants approximately 258 seconds to generate one valid hypothesis; for the control group— it took 379 seconds; however, the difference was not statistically significant. Furthermore, the validity and significance of the hypotheses were slightly lower in the VIADS group, though not statistically significant. The feasibility of the hypotheses was statistically significantly lower in the VIADS group than in the control group. The average quality rating of hypotheses per participant ranged from 7.04 to 10.55 (out of 15). In the follow-up surveys, VIADS users provided overwhelmingly positive feedback on VIADS, and they all agreed (100%) that VIADS offered new perspectives on the datasets.

Conclusion: The role of VIADS in hypothesis generation trended favorably with respect to the assessment of hypotheses generated; however, a statistically significant difference was not reached, possibly related to sample size or the 2-hour study session being inadequate. Further characterization of hypotheses, including specifics on how they might be improved, could guide future tool development. Larger-scale studies may help to reveal more conclusive hypothesis generation mechanisms.

Keywords: scientific hypothesis generation; clinical research; VIADS; utility study; secondary data analysis tools

Highlights of the paper:

- Identified the scientific hypothesis generation process from other parts of scientific or medical reasoning.
- Conducted a human subject study to generate data-driven hypotheses among clinical researchers, recorded the process, and analyzed the results.
- Established baseline data for junior clinical researchers: the number, the quality, the validity rate, and the time needed to generate data-driven hypotheses within 2 hours.
- VIADS might stimulate users' new ways of thinking during hypothesis generation.

Introduction

A scientific hypothesis is an educated guess regarding the relationships among several variables [1,2]. A hypothesis is a fundamental component of a research question [3], which typically can be answered by testing one or several hypotheses [4]. A hypothesis is critical for any research project; it dictates its direction and determines its impact. Many cognitive studies focusing on scientific research have made significant progress in scientific [5,6] and medical reasoning[7-11], problem-solving, analogy, working memory, and learning and thinking in educational contexts [12]. However, most of these studies begin with a question. Many studies on convergent and divergent thinking [13], scientific reasoning [14], medical diagnosis, or differential diagnosis [10,15,16] work on closed and open-ended questions [17] or medical symptoms. Exploring the reasoning mechanisms and processes used in solving an existing puzzle is critical to understanding human cognitive behavior in conducting advanced intellectual tasks. However, the current literature provides limited information about the scientific hypothesis generation process [4-6], which is to identify the focused area to start with, not the hypotheses generated to solve existing problems.

There have been attempts to generate hypotheses automatically using, for example, text mining, literature mining, knowledge discovery, natural language processing techniques, semantic web technology, or machine learning methods to reveal new relationships among diseases, genes, proteins, and other conditions [18-22]. Many of these efforts were based on Swanson's ABC Model [23-25]. Other researchers proposed a human-AI hybrid form to create an automated system for scientific discovery, such as hypothesis generation [26]. Several research teams explored automatic literature systems for generating [27,28] and validating [29] or enriching hypotheses [30]. However, the researchers realized the complexity of the hypothesis generation process; it does not seem feasible to generate hypotheses completely automatically [18-20,23,31]. In addition, hypothesis generation is not just identifying new relationships based on synthesizing and integrating large-scale data. New connections are critical to hypothesis generation; however, a new connection is a critical component of hypothesis generation to the maximum extent, but not identical to hypothesis generation. Other literature-related efforts include adding temporal dimensions to machine learning models to predict connections between terms [32,33] or evaluating hypotheses using knowledge bases and Semantic Web technology [31,34]. Most studies used existing literature to verify the system's validity, which is the state-of-the-art practice. However, to understand how humans use such systems to generate hypotheses in practice may provide additional insights into our understanding of scientific hypothesis generation. The new information can help system developers to better automate systems to facilitate the hypothesis generation process in future.

Many researchers believe that their secondary data analytical tools (such as a visual interactive analytic tool for filtering and summarizing large health data sets coded with hierarchical terminologies —VIADS [35-38]) can provide novel perspectives on underlying datasets, facilitating hypothesis generation [39,40]. Whether these tools work as expected, and how, has not been systematically investigated. Data-driven hypothesis generation is critical in clinical and translational research [41]. Therefore, we conducted a study to determine if and how VIADS, a secondary data analysis and visualization tool, can facilitate generating data-driven hypotheses. The study was conducted with clinical researchers. We recorded their data-driven hypothesis generation process and compared the results with and without

using VIADS. Hypothesis quality evaluation is usually part of a larger work, for example, a scientific paper or a research grant proposal. It is rare to evaluate a hypothesis alone independently and explicitly. Therefore, there are no existing metrics for hypothesis quality evaluation. We developed the metrics based on a literature review [1,3,4,42-48] and our own research and review experiences via iterative internal and external validation [42,43]. The metrics were used to evaluate the quality of hypotheses generated in this study by an expert panel. In this paper, we mainly introduce the human study of scientific hypothesis generation by clinical researchers with or without VIADS and the corresponding results: the quality evaluation of hypotheses, the quantitative measures of the hypotheses, and the results of the follow-up questions. The cognitive processes during the hypothesis generation are still under analysis, and the results will be published separately.

Methods

Research question and hypothesis

The research question for this study was:

- Can secondary data analytic tools, e.g., VIADS, facilitate the hypothesis-generation process?

We hypothesize there will be group differences between clinical researchers who use this tool in generating hypotheses and those who do not.

Rationale of the research question

Many researchers believed the new analytical tools can provide new opportunities to reveal new patterns, insights from existing data, furthermore, to facilitate users in hypothesis generation while using the tools [1,27,40,41,49]. We developed the underlying algorithms for VIADS [37,38] and VIADS [36,50,51], which can provide new ways of summarizing, comparing, and visualizing datasets. We believe VIADS can provide new perspectives for users to understand the datasets, therefore, we attempted to explore the role of VIADS in hypothesis generation process in this study.

Study design

We conducted a 2 × 2 study. We divided participants into four groups: inexperienced clinical researchers (1) without VIADS (participants were free to use any other analytical tools they were familiar with), and (2) with VIADS. Experienced clinical researchers (3) without VIADS, and (4) with VIADS. The main differences between experienced and inexperienced clinical researchers were years of experience in conducting clinical research and the number of publications in which they were significant contributors. The detailed criteria of experienced and inexperienced clinical researchers have been published in our protocol [52].

A pilot study, involving two participants, a test dataset and four study sessions, was conducted before we finalized the study datasets (Appendix A), training material (Appendix B), study scripts (Appendix C), follow-up surveys (Appendices D and E), and study session flow. Afterwards, we recruited study participants and conducted the formal study sessions.

Recruitment

We recruited study participants through multiple local and national platforms. The platforms included American Medical Informatics Association (AMIA) mailing lists for working groups (including clinical research informatics, clinical information system, implementation, clinical decision support, and Women in AMIA), N3C [53] study network Slack channels, South Carolina Clinical and Translational Research Institute newsletter, guest lectures and invited presentations in peer-reviewed conferences (e.g., MIE 2022), and several more internal and local research related newsletters (e.g., PRISMA Health Research Updates). All collaborators shared the recruitment invitations to clinical research colleagues. The recruitment invitations linked to a screening survey. Based on the experience level and our block randomization list, the participants were randomly assigned to the VIADS or non-VIADS groups. After scheduling, the study script and IRB-approved consent forms were shared with participants beforehand. The datasets were shared on the study date. All participants received compensation based on the time they spent.

Study flow

Every study participant used the same datasets and followed the same study scripts. The same researcher conducted all study sessions. For the two groups using VIADS, we scheduled a training session (one hour) and all groups had a study session lasting a maximum of 2 hours. During the training session, the researcher demonstrated how to use VIADS and

then the participants demonstrated the use of the tool. During the study session, each participant used the same datasets, analyzed them with VIADS or their routinely used analytic tools (i.e., the control group without using VIADS), and developed hypotheses based on the analytical results, their prior experience and knowledge. The participants followed think-aloud protocol during study sessions. During the study sessions, the researcher asked questions, provided reminders, and acted as a colleague to the participant, being careful not to interrupt the participant's thinking too often. All training and study sessions were conducted remotely via WebEx meetings. Figure 1 shows the study flow.

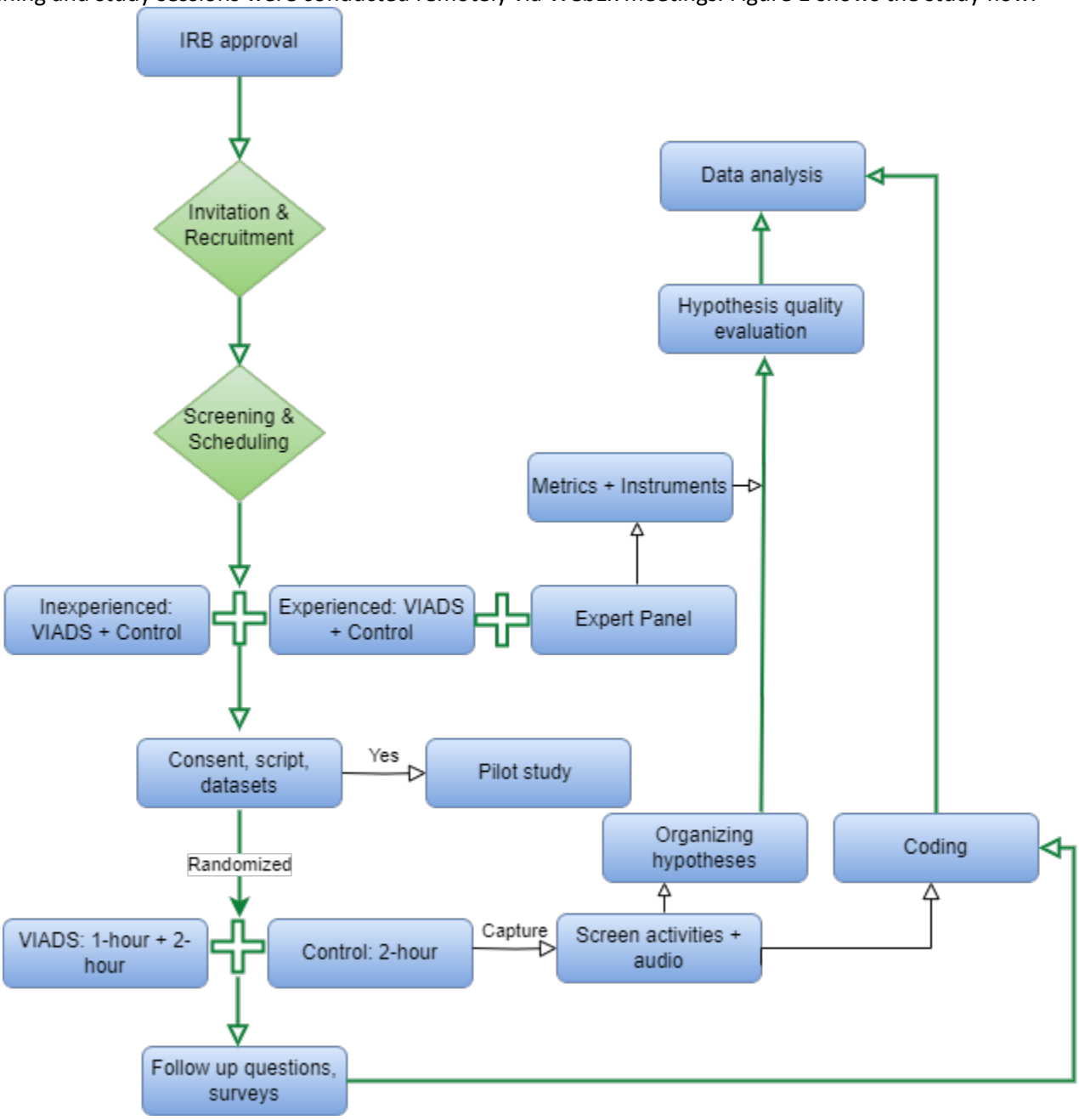

Figure 1 Study flow for the data-driven hypothesis generation

During the study session, all the screen activities and conversations between the participant and the researcher were recorded via FlashBB and converted to audio files for professional transcription. At the end of each study session, the researcher asked follow-up questions about the participants' experience related to creating and capturing new research ideas. The participants in the VIADS groups also completed two follow-up surveys after their study sessions; one was about the participant and questions on how to facilitate the hypothesis generation process better (Appendix D), and the other evaluated VIADS usability with a modified version of the System Usability Scale (Appendix E). The participants in the non-VIADS groups received one follow-up survey about hypothesis generation (Appendix D).

<u>Hypothesis evaluation</u>

We developed a complete and brief hypothesis quality evaluation instruments (Appendix F) based on hypothesis quality evaluation metrics. We recruited a clinical research expert panel with four external members, who helped us to validate the metrics. Their detailed eligible criteria were published [52]. Three senior project advisors from our investigation team with clinical research backgrounds joined the panel. The seven-member expert panel evaluated the quality of the study participants generated hypotheses during the study sessions. In Phase 1, the full version of the instrument was used to evaluate 30 hypotheses, and the evaluation results enabled us to develop a brief version of the instrument [42] (Appendix G); Phase 2 used the brief instrument to evaluate the remaining hypotheses. The brief version of the instrument included three dimensions: validity, significance, and feasibility. Each dimension used a 5-point scale, from 1 (the lowest) to 5 (the highest). Therefore, for each hypothesis, the total score will be 15.
All hypotheses were coded based on the participant's identity and the hypothesis number generated by that participant. We generated a random list of all hypotheses. Then, based on the random list, we put ten randomly selected hypotheses into one Qualtrics survey for evaluation. We initiated the evaluation process when we completed all the study sessions. Therefore, all hypotheses can be included in generating the random list.

Hypothesis evaluation data analysis plan

Our data analysis focuses on the quality of hypotheses generated by participants in the different groups. We conducted independent t-test in MPlus 7 to compare the VIADS group and the control group to see if there were significant mean differences in the quality of the hypotheses. We also examined the correlations between the quality ratings and the participant's self-perceived creativity.
We first analyzed all hypotheses to explore the aggregated results. A second analysis was conducted by using only valid hypotheses after removing any hypothesis that was scored at "1" (the lowest rating) for validity by three or more experts. We include both sets of results in this paper. The usability results of VIADS were published separately [35].
The average number of hypotheses generated by participants and the time needed per hypothesis per participant were also compared between the VIADS group and the control group via independent t-test.
All hypotheses were coded by two research assistants who worked separately and independently. They coded the time needed for each hypothesis, coded the hypothesis generation process, and counted the number of hypotheses generated by each participant. The coding principles (Appendix H) were developed as the two research assistants worked. Whenever there was a discrepancy, a third researcher joined the discussion to reach a consensus by refining the coding principles.

Ethical statement

Our study was approved by the Institutional Review Boards (IRB) of Clemson University, South Carolina (IRB2020-056) and Ohio University (18-X-192).

Results

Participant demographics

We screened 39 researchers, among whom 20 participated, of which 2 were in the pilot study. Participants were from different locations and institutions in the United States. Among the 18 study participants, 15 were inexperienced clinical researchers and three were experienced. The experienced clinical researchers were underrepresented, and their results were mainly for informational purposes, without further comparison. Table 1 presents the background information of the inexperienced participants.

Table 1: Participants profile (15 inexperienced clinical researchers)

Gender	**VIADS Group**	**Control Group**
Female	5	4
Male	3	3
Hypothesis generation experience (years)		
< 2	6	3
2–5	2	3
> 5		1
Collaborator role	5	5
Leading role	1	1
Other	2	1
Design experience (years)		
< 2	6	3
2–5	2	4
5–10		
Collaborator role	6	5
Leading role		1
Other	2	1
Data analysis experience (years)		
< 2	7	3
2–5	1	4
5–10		
Number of main publications		
< 5	7	7
≥ 5 and < 10	1	
5 – 10		
Specialties/Domain		
Pediatrics		1
Primary care provider	1	2
Health science	2	2
Clinical research coordinator	2	1
Pharmacy	2	
Nursing		1
Internal medicine	1	
Routine data analytic tools		
Excel	6	4
R	3	2
SPSS		1
SAS	2	2
Stata		1

The expert panel composition and intraclass correlation coefficient (ICC)

Seven expert panel members validated the metrics and instrument and evaluated hypotheses using the instrument. All seven experts from seven different institutions in the United States. Five of them have medical backgrounds; three are in clinical practice, and two have methodology backgrounds. They all have at least ten years of intense clinical research experience. The development and validation of the metrics and instrument were published separately [42,43]. The ICC for the seven experts' ratings is moderate at 0.49 for the hypothesis evaluation.

Hypothesis quality and quantity evaluation results

The 18 participants generated 227 hypotheses during the study sessions. After removing the invalid ones (rated at "1", the lowest validity score, by three or more experts), 147 (65%) hypotheses were left for further analysis and comparison. Of these, 121 were generated by inexperienced clinical researchers (n = 15) in the VIADS (n = 8) and control (n = 7) groups.

Table 2 shows the main comparison of the hypothesis evaluation results between the two groups: the VIADS and the control groups. We analyzed and reported the results in separate categories: valid hypotheses by inexperienced clinical researchers (n = 121), valid hypotheses by inexperienced and experienced clinical researchers (n = 147), all hypotheses by inexperienced clinical researchers (n = 192), and all hypotheses by inexperienced and experienced clinical researchers (n = 227). The results of the four strategies have generated similar trends. That is, the VIADS group received slightly lower validity and significance scores, but the differences were statistically insignificant ($p > 0.05$). In contrast, the VIADS group received statistically significantly lower scores in the feasibility dimension and the overall evaluation of hypotheses ($p < 0.001$).

Table 2 Expert panel rating results for hypotheses generated by VIADS and control groups

Analytic strategies (n = number of hypotheses)	**Seven expert rating hypothesis group mean**													
	Overall				**Validity**			**Significance**			**Feasibility**			
	VIADS	Control	95% CI of diff.	*P* value	VIADS	Control	*P* value	VIADS	Control	*P* value	VIADS	Control	95% CI of diff.	*P* value
Inexperienced + valid hypotheses only (n = 121)	2.97	3.17	[−0.36, −0.04]	0.014	2.79	2.85	0.57	3.14	3.15	0.934	2.98	3.53	[−0.82, −0.27]	< 0.001
Inexperienced + experienced + valid hypotheses (n = 147)	2.93	3.19	[−0.4, −0.12]	<0.001	2.75	2.86	0.20	3.17	3.19	0.769	2.88	3.52	[−0.88, −0.40]	< 0.001
Inexperienced (n = 192)	2.86	3.06	[−0.35, −0.04]	0.016	2.45	2.57	0.18	3.09	3.14	0.532	2.88	3.33	[−0.69, −0.25]	<0.001
Inexperienced + experienced (n = 227)	2.84	3.09	[−0.40, −0.11]	0.001	2.43	2.61	0.03[&]	3.10	3.18	0.289	2.80	3.34	[−0.76, −0.32]	<0.001

Note: 95% CI of diff is the 95% confidence interval of the mean difference between the two groups. Each hypothesis was rated on a 5-point scale, from 1 (the lowest) to 5 (the highest) in each dimension.

Table 3 shows the number of hypotheses and the time needed to generate a hypothesis per participant based on valid hypotheses only by inexperienced clinical researchers. The VIADS group and the control group generated a similar number of valid hypotheses. The VIADS group took less time on average to generate a valid hypothesis. However, the group differences were not statistically significant in either mean numbers or mean time ($p > 0.05$). The results were consistent with the results when analyzing all (valid and invalid) hypotheses [54].

Table 3 Valid hypotheses generated by inexperienced clinical researchers in 2 hours and average time/hypothesis

	Min #	**Max #**	**Mean #**	**Min T (s)**	**Max T (s)**	**Mean T (s)**
VIADS (n = 8)	1	19	8.43	102	610	278.6
Non-VIADS (n = 7)	2	13	7.63	250	566	358.2

Note, #, number; T, average time (second)/person/hypothesis; s, second; 5 participants in the control group had their durations coded.

Evaluation of hypothesis quality and quantity by individual participants

Tables 4 and Figures 2 and 3 present the individual participants (i.e., inexperienced clinical researchers) and their average score per valid hypothesis. Two participants with the highest scores are in the control group, and two with the lowest are in the VIADS group.

Table 4 The average rating score of valid hypotheses generated by each participant

Participant - Control	Mean	SD	Participant-VIADS	Mean	SD
PA (n = 5)	8.75	1.54	P1 (n = 8)	8.7	1.18
PB (n = 9)	10.55	0.98	P2 (n = 6)	8.30	1.43
PC (n = 13)	8.92	1.34	P3 (n = 5)	8.92	0.58
PD (n = 2)	9.30	0.64	P4 (n = 1)	8.50	
PE (n = 9)	9.07	1.57	P5(n = 7)	8.94	1.44
PF (n = 9)	9.30	0.93	P6(n = 8)	9.81	0.99
PG (n = 12)	10.25	1.27	P7(n = 19)	8.82	1.61
			P8 (n = 8)	8.94	1.60
Group average (n = 59)	9.45	0.68	Group average-V (n = 62)	8.87	0.45

Note: Control – Control group; VIADS – VIADS group; n represents the number of valid hypotheses generated. The two highest scoring participants are highlighted in yellow, and the two lowest in green.

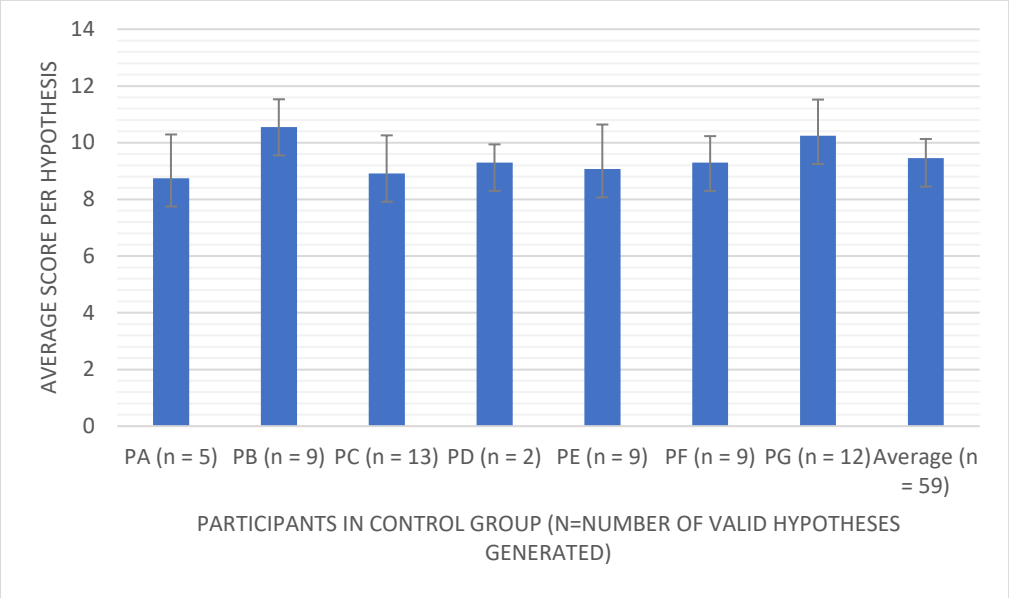

Figure 2 Study participant's average score per valid hypothesis generated during the study sessions in the control group (the maximum score is 15 per hypothesis)

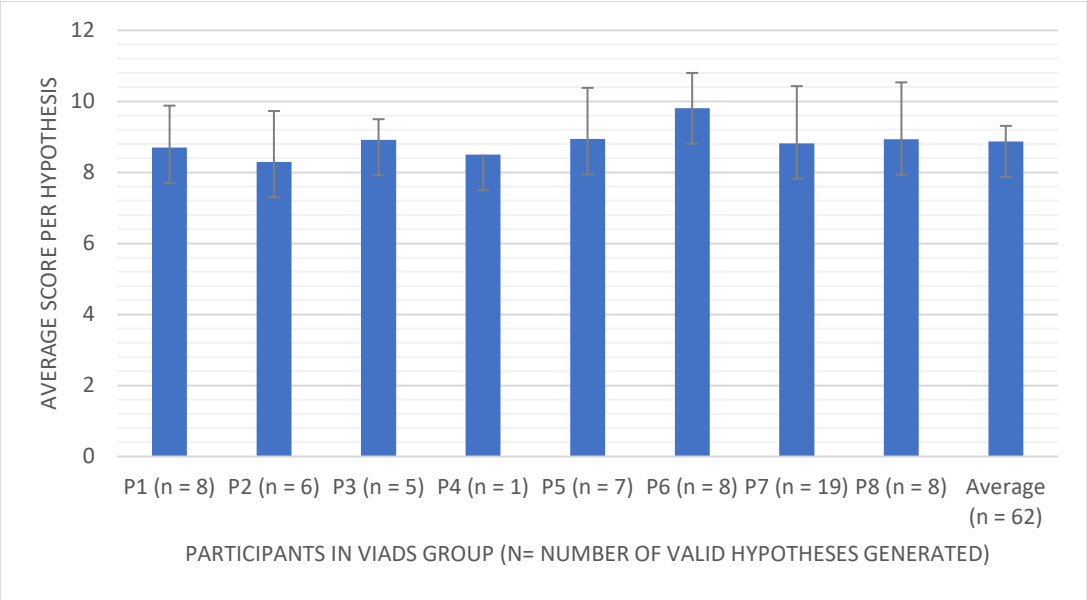

Figure 3 Study participant's average score per valid hypothesis generated in the VIADS group (the maximum score is 15 per hypothesis)

Figures 4 and 5 demonstrate the individual participant, their corresponding number of valid hypotheses, and the average time needed per valid hypothesis. Individual variations can be observed in both groups.

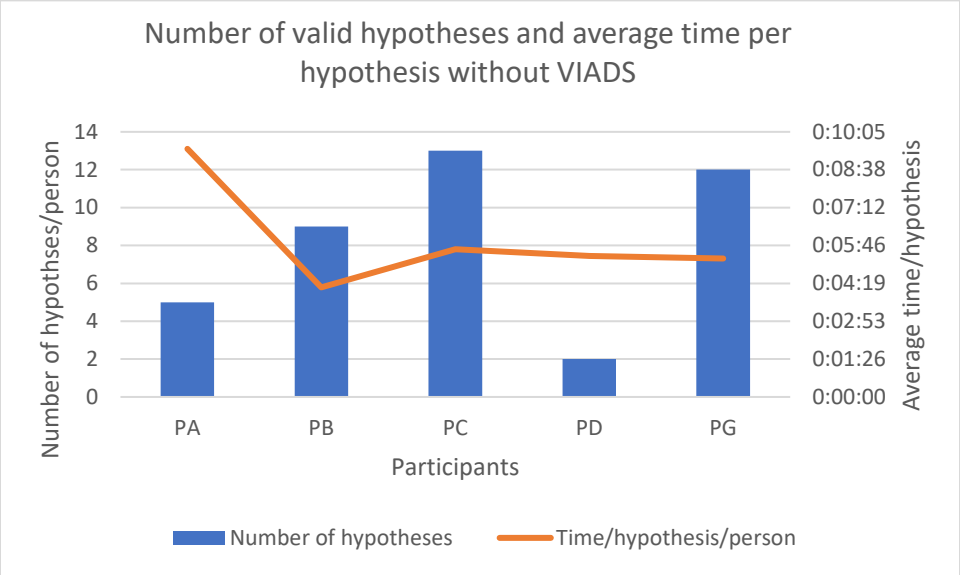

Figure 4 The number of valid hypotheses generated and the average duration per hypothesis for each participant among inexperienced clinical researchers without VIADS

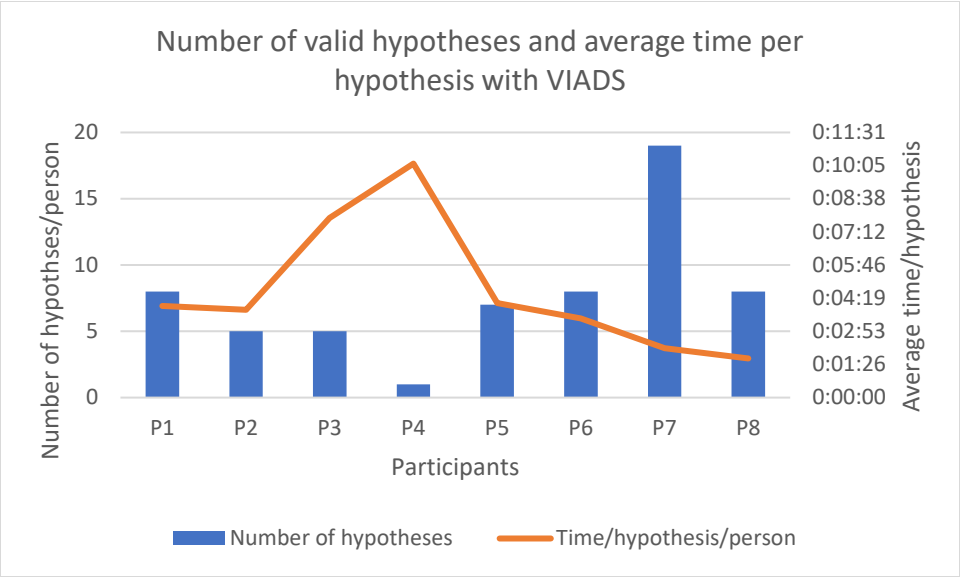

Figure 5 The number of valid hypotheses generated and the average duration per hypothesis for each participant among inexperienced clinical researchers with VIADS

<u>Experienced clinical researchers</u>

There were three experienced clinical researchers among the participants, two in the VIADS group and one in the control group. Table 5 lists their basic descriptive statistics and their hypothesis generation results.

Table 5 Results of the experienced clinical researchers in VIADS and the control groups

	VIADS	**Control/Non-VIADS**
Male	2	1
Hypothesis generation experience (years)		
≥ 10	1	
≥ 5 and < 10	1	1
Collaborating role	1	
Leading role	1	1
Study design experience (years)		
≥ 5 and < 10	2	1
Collaborating role	1	
Leading role	1	1
Data analysis experience (years)		
≥ 10	1	
≥ 5 and < 10	1	1
Number of main publications		
≥ 10	1	1
5 – 10		1
Routine data analytic tools		
SPSS	2	
MS Excel	1	1
R		1
Number of valid hypotheses	12	12
	3	
Average time needed/hypothesis	03:35	06:53
	06:47	
Average score ± SD	8.51 ± 1.18	9.84 ± 0.81

	7.04 ± 0.30	

Follow-up questions

There were three parts to the follow-up questions: researcher questions at the end of the study session for all participants, a follow-up survey for all participants, and a SUS usability survey for the VIADS group participants. The SUS results have been published separately as a VIADS usability study [35]. The results from the first two parts are summarized below.

The verbal questions asked each participant after the study session and the summary answers are presented in Table 6. Reading and interactions with others were the most used activities to generate new research ideas. Attending conferences, seminars, educational events, and conducting clinical practice were important in generating hypotheses. There were no specific tools used to initially capture hypotheses or research ideas. Most participants used text documents in Microsoft Word, text messages, emails, or sticky notes to summarize their initial ideas.

Table 6 Follow-up questions (verbal) and answers after the study sessions (all study participants)

#	Follow-up questions	Summary answers
1	What activities/events (e.g., reading papers/books, discussing with colleagues/students/family, presentation Q & A, daily work, witnessing something completely irrelevant) provoked new research ideas in the past?	Reading (14/18) Conversation/interaction (13/18) Conference/seminar (10/18) Clinical practice/observation (6/18)
2	How do you capture research ideas in the first place usually?	A text document (13/18) Bouncing ideas to obtain feedback (10/18)
3	Do you describe yourself as a creative person versus someone who generally follows instructions carefully? I.e., a creative person may always try to find new ways to do things or to look at things. (Likert scale of 1 to 5, the lowest to the highest)	"2" points (1/18) "3" or 3.5 points (5/18) "4" points (9/18) "5" points (3/18)
4	Do you know if others, e.g., your family, close friends, and colleagues, perceive you as creative? (yes, no, not sure)	Yes (14/18) Not sure (4/18)
5	What will facilitate the generation of new research ideas in your view?	Conversation/interaction (11/18) Reading (6/18) Time (4/18) Conference (3/18)

Figure 6 is a scientific hypothesis generation framework we developed based on a literature review [1,3,4,44,55,56], follow-up questions and answers after study sessions, and self-reflection on our research project trajectories. The external environment, cognitive capacity, and interactions between the individual and the external world, especially the tools used, are categories of critical factors that significantly contribute to hypothesis generation. Cognitive capacity takes a long time to change, and the external environment can be unpredictable. The tools that can interact with existing datasets are one of the modifiable factors in the hypothesis generation framework and this is what we aimed to test in this study.

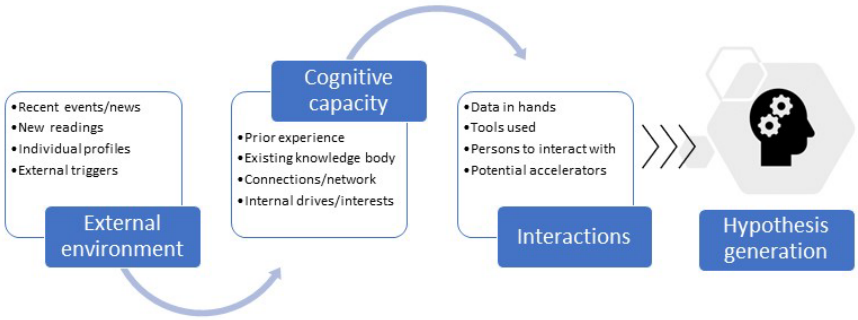

Figure 6 Scientific hypothesis generation framework: contributing factors

The average hypothesis quality rating score per participant did not correlate with the self-perceived creativity (p = 0.616, 2-tailed Pearson correlation test) or the number of valid hypotheses generated (p = 0.683, 2-tailed Pearson correlation test) by inexperienced clinical researchers. There was no correlation between the highest and lowest 10 ratings and the individual's self-perceived creativity, in either group of inexperienced clinical researchers.

In our follow-up survey, the questions were mainly about participants' current roles and affiliations, their experience in clinical research, their preference for analytic tools, and their rating of the importance of different factors considered routinely in clinical research study design (Figure 7). Most of the results have been included in Table 1. Figure 7 shows the ratings of the study design factors by the two groups of participants. The VIADS group rated almost every factor slightly higher than the control group.

In our follow-up survey, one question was, "If you were provided with more detailed information about research design (such as focused population) during your hypothesis generation process, do you think the information would help formulate your hypothesis overall?" All 20 participants, (including 2 in the pilot study), selected Yes. This demonstrates the recognition and need for assistance during hypothesis generation. In the follow-up surveys, VIADS users provided overwhelmingly positive feedback on VIADS, and they all agreed (100%) that VIADS offered new perspectives on the datasets [35].

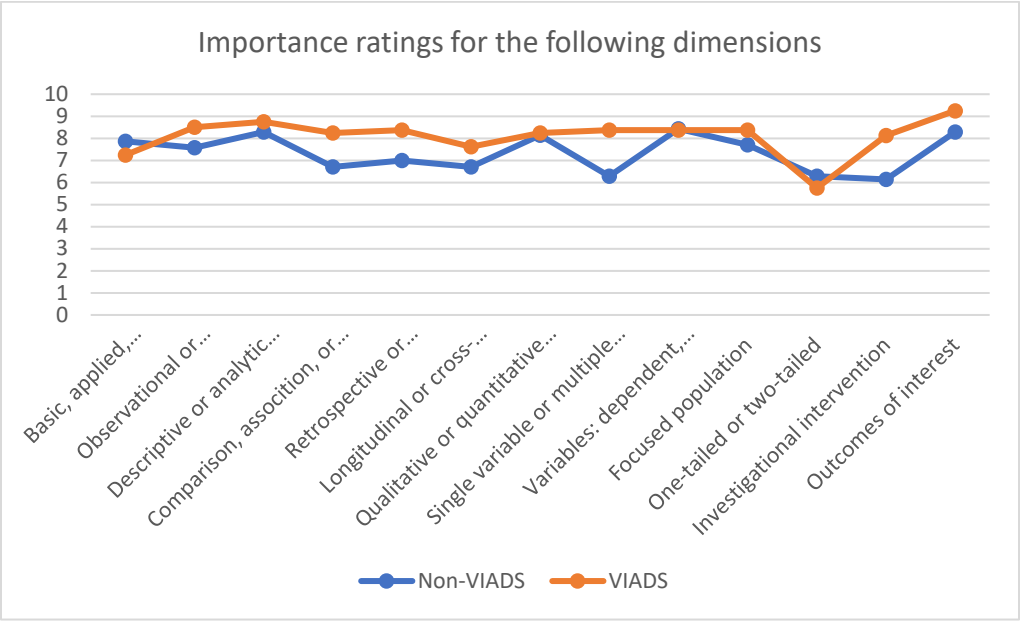

Figure 7 Participants' (inexperienced clinical researchers) perceived importance ratings of factors in clinical research study design

Discussion

Interpretation of the results

This study had one objective: to discover the role of secondary data analytic tools, such as VIADS, in generating scientific hypotheses in clinical research. We also use this process to evaluate the utility and usability of VIADS. The usability of VIADS has been published separately [35]. Regarding the role and utility of VIADS in hypothesis generation, we measured the number of hypotheses generated, the time needed to generate each hypothesis, the quality evaluation of the hypotheses, and the feedback on VIADS. Participants in the VIADS and control groups generated similar numbers of valid and total hypotheses among junior clinical researchers. The VIADS group a) needed a shorter time to generate each hypothesis on average; however, this was statistically not significant; b) received slightly lower ratings on quality measures in validity and significance than the control group; however it was statistically not significant; c) received statistically significantly lower quality ratings in feasibility; d) provided very positive feedback on VIADS [35] with 75% agreed that VIADS facilitates understanding, presentation, and interpretation of the underlying datasets; e) agreed (100%) that VIADS provided new perspectives on the datasets after 1-hour training and 2-hour intense use of VIADS. The feasibility results were consistent regardless of the analytical strategies implemented: using valid hypotheses only, examining inexperienced clinical researchers only, using all hypotheses, or looking at both inexperienced and experienced clinical researchers.

The current results were inconclusive in answering our research question. The direct measurements did not demonstrate statistically significant differences between the VIADS and the control groups, except for feasibility rating (Table 2). This may indicate that VIADS stimulated participants' new ways of thinking causing generation of some unfeasible hypotheses. This outcome was unexpected and the study design does not explain it, as measuring changes of new ways of thinking or creativity would be difficult [17]; however, new ways of thinking could be a possible reason to explain the systematically lower feasibility ratings in the VIADS group. Especially considering the relatively small sample size in this study, the differences in feasibility indicate a relatively large effect size. These results also align with VIADS group participants' feedback on VIADS [35].

Although Dumas's and Dunbar's study [17] demonstrated that divergent thinking could be enhanced or diminished, the authors emphasized that creativity is not likely a stable individual trait but is related to context and perspectives. Therefore, the feasibility evaluation of each hypothesis in our study is good practice. However, the latent semantic analysis used to measure creativity [17] is not necessarily a suitable measurement in our context. Another study also measured new ideas and creativity [13]. However, the measurements were based on answers to information discovery questions, including open-ended ones. That setting differs from ours, which does not have a question to start with but comes up with questions based on the datasets. An analogy would be that their research had an anchor to begin with, but ours had to find the anchor. The relationship between creativity and feasibility of scientific hypotheses generated by participants needs a larger scale study to explore further, which can be one potential future direction.

Evidence in the literature [57] suggests that learning a complex tool and doing the task simultaneously presents extra cognitive load on the participants. This is likely the case in this study; the results are reflected in Table 2. VIADS group participants needed to learn how to use the tool and then analyze the datasets with it to come up with the hypotheses. The cognitive overload may not have been conscious. Therefore, they perceived VIADS as helpful in understanding datasets. However, the quality evaluation results after 2 hours of use did not support the participants' perceptions of VIADS.

The role of VIADS in the hypothesis generation process may not be linear. The 2-hour use of VIADS did not generate statistically higher average hypothesis quality ratings; however, all participants (100%) agreed that VIADS provided new perspectives on the datasets. The true role of VIADS in hypothesis generation might be more complicated than we initially thought. Either two hours were insufficient to generate higher average quality ratings, or our evaluation was not adequately granular to capture the value of a tool, like VIADS. A more natural use environment might be necessary instead of a simulated environment in 2 hours to demonstrate detectable differences. In addition, the cognitive

processes during hypothesis generation are under analysis, and the results will be published separately. The cognitive process results may shed additional light on our understanding of VIADS and its role in hypothesis generation.

Figure 7 shows the VIADS group rated almost all factors relevant to clinical research study design slightly higher. VIADS group participants showed higher awareness of hypothesis generation and the importance of additional assistance in facilitating the process.

Researchers have long been keen to understand where good research ideas come from [58,59]. Participants' answers to our follow-up questions have provided more anecdotal information about possible events or activities contributing to the hypothesis generation process. From these insights, and a literature review [1,3,4,44,55,56], we formulated a hypothesis generation framework (Figure 6). All the following activities and events were identified by participants as associated to generating hypothesis: reading, interactions with others to obtain feedback and refine the ideas, observations during clinical practice, teaching, learning, and listening to presentations. Individuals think to connect these ideas, facts, and phenomena, and formulate them into research questions and hypotheses to test. Although these events or activities did not answer the question of where good research ideas come from directly, they were identified by the participants as associated with hypothesis generation in the past. Identifying an impactful research question is necessary but does not guarantee a successful research project. The study should be well-designed, rigorously conducted, and the data should be analyzed thoroughly. Further, the results should be communicated effectively. All these take more knowledge, logical thinking, reasoning skills, diligence, passion, and persistence along the way. Serious researchers deserve more recognition and credit for successful research projects, big or small. Even with all these merits, research results are almost always unpredictable. This is partially the beauty of scientific exploration and discovery; to verify something known is never the goal.

Three participants with the highest average ratings (i.e., 10.55, 10.25, and 9.84 out of 15) were all in the non-VIADS group. The two participants with above ten average score ratings were inexperienced clinical researchers, and the third was an experienced clinical researcher. They all practice medicine. Based on the conversations during the study sessions and follow-up questions after study sessions, they all put much thought into research and connecting observations in medical practice and daily life; their clinical practice experience, education, observation, thinking, and making connections between clinical practice and research ideas contributed to their higher ratings on their hypotheses. This observation verifies the belief that good research ideas come from these three pillars: 1) a deep understanding of the domain knowledge and the supporting fundamental mechanisms, 2) connection of knowledge and practice observations (problems or phenomena), 3) putting the observations (problems or phenomena) into the appropriate research contexts. These three pillars were formulated based on synthetization and summarization the answers to follow-up questions, literature review [3,4,58,60-62], and our reflections of research project development processes.

One participant had the lowest average score rating on hypotheses (i.e., the only one with a below eight average rating score), and the participant was as an experienced clinical researcher in the VIADS group. The participant also practices medicine. There were discrepancies between the self-perceived experience level or professional title and the performance during the study session based on the average score rating. However, reasons irrelevant to one's research or medical practice experience might explain the results. For example, not everyone performs (or thinks) naturally or well with time constraints and a stranger present. Sometimes a self-perceived correlation between the total number of publications and one's research capacity can be mismatched, depending on actual contributions to the design, conduct, and analysis of a study that a researcher has.

The three participants with the highest average score ratings were in the non-VIADS group, and the participant with the lowest average score rating was in the VIADS group. Regardless of our randomization, individual variations may play an amplified role when the sample size is relatively small. This might be why the current results did not demonstrate a statistically significant difference in the VIADS group, except for feasibility.

<u>Significance of the work</u>

We conducted the first human subject study using the same datasets to investigate data-driven hypothesis generation and compare VIADS and other routinely used analytic tools by participants. The significance of the study can be demonstrated in the following aspects: ***firstly***, this study demonstrated the feasibility of remotely conducting a data-driven hypothesis generation study via the think-aloud protocol. ***Secondly***, we established hypothesis quality evaluation metrics and instruments, which may be useful for clinical researchers to assess others' research ideas during peer review or prioritize their research ideas before investing too many resources. ***Thirdly***, this study measured the baseline data for number of hypotheses, time needed, and quality of hypotheses generated by clinical researchers. The baseline provides a reference for other researchers interested in diving deeper into this field. ***Fourthly***, hypothesis generation is complicated, and our current measurements may be inadequate to deal with such complexity. Our experience and all the VIADS group participants agreed that VIADS is helpful in generating hypotheses; however, the quantitative and qualitative measures used did not show a statistically significant difference. ***Fifthly***, among junior clinical researchers, we identified that more assistance is needed in the hypothesis generation process. ***Sixthly***, our study indicates that VIADS might stimulate the creativity of users while analyzing the datasets. This work set the foundation in achieving more structured and organized clinical research projects, starting from a more explicit hypothesis generation process. However, we believe that this is only the tip of the hypothesis generation iceberg.

Strengths and limitations of the work

The study participants were from all over the country, not a single health system or institution. Although the sample of participants may be more representative, individual variations may have played a more significant role than estimated in the hypothesis quality measurements.

We implemented several strategies to create comparable groups. For example, we used the same data sets, study scripts, and platform (WebEx) for all the study sessions. The same researcher conducted all study sessions. Two research assistants examined the time measurements of the hypothesis generation process independently and compared their results later, which made the coding much more consistent and robust.

The study has a robust design and a consistent coding with carefully conducted analytic process. We implemented randomization in multiple levels to reduce potential bias. The study participants were separated into experienced or inexperienced groups based on predetermined criteria during screening [52] and then assigned randomly to VIADS or non-VIADS groups. During the hypothesis quality evaluation, every ten randomly selected hypotheses were organized into one Qualtrics survey. This practice provides fair opportunity for each hypothesis during evaluation, reduce the potential bias related to the order of hypotheses. The hypothesis quality measurement metrics and instruments were validated iteratively, internally and externally [42]. Two research assistants verified the coding system. Two research assistants examined the time measurements independently, then compared and consolidated the results. When we conducted the data analysis, we implemented multiple strategies during data analysis to provide a comprehensive view on the data collected. We examined a) valid hypotheses only, b) inexperienced clinical researchers only, c) all hypotheses, d) all clinical researchers (Table 2). We believe these practices contributed to the robustness of the study.

One major limitation of the study was the sample size. The power calculation used during the study design was overly optimistic about the potential effect size mainly due to our confidence that VIADS can provide more new aspects during data analysis than other tools. The VIADS groups participants verified our confidence on VIADS via the follow-up surveys. However, the quantitative measures did not show statistically significant difference to support. One possible explanation is that hypothesis generation is highly complicated. A tool like VIADS may be helpful, but the effects might not be easily measured after 2–3 hours of use. The discrepancy between the perceived usefulness of VIADS by participants and the quantitative measures may be due to aspects that were not captured by our current methods.

Another limitation of the study is that we used a simulated environment. In the study session, there is a time constraint for each participant and the researcher who conducted the study sessions could have been a stressor for participants. Time pressure can reduce hypothesis generation [63] . However, this influence would be similar in both groups. The hypothesis generation process in real life is usually lengthy, with discussions, revisions, feedback, and refinement. We do not know whether a simulated environment can reflect the true natural environment of hypothesis generation.

Our current methods and measurements could not accurately capture all the process details. The cognitive processes of hypothesis generation recorded through the think-aloud protocol during study sessions is still under analysis, which may shed more lights on the comparison between groups.

Challenges and future directions

In this study, we faced several challenges beyond our control that may have affected the study results. The current process can only capture the conscious and verbalized processes and may, for example, have failed to capture unconscious cognitive processes. Therefore, continue to analyze the recorded think-aloud sessions, especially the cognitive processes during hypothesis generation might help us to understand the process and the differences among groups better. To have a large-scale study with more participants for a study like this one can be another possible future direction.

VIADS group participants provided very positive feedback on VIADS and its role in hypothesis generation, however, the limitations of VIADS are obvious. By nature, VIADS is a data analysis and visualization tool. It can accept specific dataset format and only supports certain types of hypotheses. Therefore, more powerful and comprehensive tools designed to assist hypothesis generation particularly are needed [59]. In addition, a longer duration of use and use of the tool in a more natural environment instead of a simulated experiment environment might be necessary to demonstrate the tools' effectiveness.

Recruitment is always challenging in human subject studies or clinical trials [46-48]. It is particularly challenging to recruit experienced clinical researchers, even though we made similar efforts and used similar platforms to recruit inexperienced and experienced clinical researchers. The different recruitment outcomes may be due to 1) Hypothesis generation is not a high priority for experienced clinical researchers; they have sharpened their research skills over time, including hypothesis generation. 2) They are overwhelmed by existing responsibilities and could not take on additional tasks as needed by our study. 3) They may have ongoing research projects and did not want to explore other research possibilities. Overall, for experienced clinical researchers, hypothesis generation does not seem to be an area that needs urgent assistance.

Conclusion

The role of VIADS in hypothesis generation trended favorably with respect to the assessment of hypotheses generated; however, a statistically significant difference was not reached, possibly related to sample size or the 2-hour study session being inadequate. Further characterization of hypotheses, including specifics on how they might be improved, could guide future tool development. Larger-scale studies may help to reveal more conclusive hypothesis generation mechanisms.

Acknowledgment

The project is supported by a grant from the National Library of Medicine of the United States National Institutes of Health (R15LM012941). It is partially supported by the National Institute of General Medical Sciences of the National Institutes of Health (P20 GM121342). This work has also benefited from research training resources and the intellectual environment enabled by the NIH/NLM T15 South Carolina Biomedical Informatics and Data Science for Health Equity (SC BIDS4Health) research training program (T15LM013977). The content is solely the authors' responsibility and does not necessarily represent the official views of the National Institutes of Health.

Appendices

Appendix A: The data sets used during the hypothesis generation study

Appendix B: Training materials used for participants in VIADS groups

Appendix C: Study scripts for participants in VIADS and non-VIADS groups

Appendix D: Follow-up survey after the hypothesis generation study

Appendix E: Modified version of System Usability Scale

Appendix F: Hypothesis quality evaluation instrument for clinical research—full version

Appendix G: Hypothesis quality evaluation instrument for clinical research—brief version

Appendix H: Coding principles on timing of hypothesis generation process

References

1. Supino P, Borer J. Principles of research methodology: A guide for clinical investigators. 2012
2. Parahoo A. Nursing research: Principles, Process & issues. 1997
3. Farrugia P, Petrisor B, Farrokhyar F, Bhandari M. Research questions, hypotheses and objectives. *J Can Chir* 2010;**50**
4. Pruzan P. *Research Methodology: The Aims, Practices and Ethics of Science*: Springer International Publishing Switzerland, 2016.
5. *The Oxford handbook of thinking and reasoning*. New York, NY, US: Oxford University Press, 2012.
6. *The Cambridge Handbook of Thinking and Reasoning*. New York: Cambridge University Press, 2005.
7. Patel VL, Arocha JF, Zhang J. Chapter 30: Thinking and Reasoning in Medicine. In: Holyoak KJ, Morrison RG, eds. The Cambridge Handbook of Thinking and Reasoning. New York: Cambridge University Press, 2005:727-50.
8. Kushniruk A, Patel V, Marley A. Small worlds and medical expertise: implications for medical cognition and knowledge engineering. *Int J Med Inform* 1998;**49**:255-71.
9. Patel V, Groen G, Patel Y. Cognitive aspects of clinical performance during patient workup: The role of medical expertise. *Advances in Health Sciences Education* 1997;**2**:95-114.
10. PATEL VL, GROEN GJ, AROCHA JF. Medical expertise as a function of task difficulty. *Memory & cognition* 1990;**18**(4):394-406.
11. Patel V, Groen G. Knowledge Based Solution Strategies in Medical Reasoning. *Cognitive Sci* 1986;**10**:91-116. doi: 10.1207/s15516709cog1001_4
12. Moseley D, Baumfield V, Elliott J, et al. *Frameworks for Thinking: A Handbook for Teaching and Learning*. Cambridge: Cambridge University Press, 2005.
13. Kerne A, Smith S, Koh E, Choi H, Graeber R. An Experimental Method for Measuring the Emergence of New Ideas in Information Discovery. *International Journal of Human-Computer Interaction* 2008;**24**(5):460-77. doi: 10.1080/10447310802142243
14. Dunbar K, Fugelsang J. Causal thinking in science: How scientists and students interpret the unexpected. In: Gorman M, Kincannon A, Gooding D, Tweney R, eds. New directions in scientific and technical thinking. Mahway, NJ: Erlbaum, 2004:57-59.
15. Arocha J, Patel V, Patel Y. Hypothesis generation and the coordiantion of theory and evidence in novice diagnostic reasoning. *Medical Decision Making* 1993;**13**:198-211.
16. Joseph G-M, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making* 1990;**10**:31-46.
17. Dumas D, Dunbar K. The Creative Stereotype Effect. *PLoS ONE* 2016;**11**(2):e0142567. doi: doi:10.1371/journal.pone.0142567
18. Spangler S, Wilkins AD, Bachman BJ, et al. Automated hypothesis generation based on mining scientific literature. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, New York, USA: Association for Computing Machinery, 2014:1877–86.
19. Soldatova LN, Rzhetsky A. Representation of research hypotheses. *J Biomed Semantics* 2011;**2 Suppl 2**(Suppl 2):S9. doi: 10.1186/2041-1480-2-s2-s9
20. Wittkop T, TerAvest E, Evani US, et al. STOP using just GO: a multi-ontology hypothesis generation tool for high throughput experimentation. *BMC Bioinformatics* 2013;**14**:53. doi: 10.1186/1471-2105-14-53
21. Sang S, Yang Z, Li Z, Lin H. Supervised Learning Based Hypothesis Generation from Biomedical Literature. *Biomed Res Int* 2015;**2015**:698527. doi: 10.1155/2015/698527
22. Petric I, Ligeti B, Gyorffy B, Pongor S. Biomedical hypothesis generation by text mining and gene prioritization. *Protein Pept Lett* 2014;**21**(8):847-57. doi: 10.2174/0929866511320999006 3

23. Swanson DR, Smalheiser NR. Implicit Text Linkages between Medline Records: Using Arrowsmith as an Aid to Scientific Discovery. Library Trends, 1999:48.
24. Swanson DR. Undiscovered Public Knowledge. *The Library Quarterly: Information, Community, Policy* 1986;**56**(2):103-18.
25. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in biology and medicine* 1986;**30**(1):7-18. doi: 10.1353/pbm.1986.0087
26. Kitano H. Nobel Turing Challenge: creating the engine for scientific discovery. *npj Systems Biology and Applications* 2021;**7**(1):29. doi: 10.1038/s41540-021-00189-3
27. Sybrandt J, Shtutman M, Safro I. Moliere: Automatic biomedical hypothesis generation system: ACM, 2017.
28. Liekens AM, De Knijf J, Daelemans W, Goethals B, De Rijk P, Del-Favero J. BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. *Genome Biol* 2011;**12**(6):R57. doi: 10.1186/gb-2011-12-6-r57
29. Sybrandt J, Shtutman M, Safro I. Large-Scale Validation of Hypothesis Generation Systems via Candidate Ranking. *Proc IEEE Int Conf Big Data* 2018;**2018**:1494-503. doi: 10.1109/bigdata.2018.8622637
30. Baek SH, Lee D, Kim M, Lee JH, Song M. Enriching plausible new hypothesis generation in PubMed. *PLOS ONE* 2017;**12**(7):e0180539. doi: 10.1371/journal.pone.0180539
31. Callahan A, Dumontier M, Shah NH. HyQue: evaluating hypotheses using Semantic Web technologies. *Journal of Biomedical Semantics* 2011;**2**:NA.
32. Akujuobi U, Spranger M, Palaniappan SK, Zhang X. T-PAIR: Temporal Node-Pair Embedding for Automatic Biomedical Hypothesis Generation. *IEEE Transactions on Knowledge and Data Engineering* 2022;**34**(6):2988-3001. doi: 10.1109/TKDE.2020.3017687
33. Akujuobi U, Chen J, Elhoseiny M, Spranger M, Zhang X. Temporal Positive-unlabeled Learning for Biomedical Hypothesis Generation via Risk Estimation. 34th Conference on Neural Information Processing Systems (NeurIPS 2020); 2020; Vancouver, Canada.
34. Whelan K, Ray O, King RD. Representation, simulation, and hypothesis generation in graph and logical models of biological networks. *Methods Mol Biol* 2011;**759**:465-82. doi: 10.1007/978-1-61779-173-4_26
35. Jing X, Patel VL, Cimino JJ, et al. A Visual Analytic Tool (VIADS) to Assist the Hypothesis Generation Process in Clinical Research: Mixed Methods Usability Study. *JMIR Human Factors* 2023;**10**:e44644. doi: doi: 10.2196/44644
36. Jing X, Emerson M, Masters D, et al. A visual interactive analysis tool for filtering and summarizing large data sets coded with hierarchical terminologies (VIADS). *BMC Med Inform Decis Mak* 2019;**19**(31) doi: https://doi.org/10.1186/s12911-019-0750-y
37. Jing X, Cimino JJ. Graphical methods for reducing, visualizing and analyzing large data sets using hierarchical terminologies. AMIA 2011. Washington DC, 2011:635-43.
38. Jing X, Cimino JJ. A complementary graphical method for reducing and analyzing large data sets: Case studies demonstrating thresholds setting and selection. *Methods of Information in Medicine* 2014;**53** doi: 10.3414/ME13-01-0075
39. Spangler S. Accelerating discovery : mining unstructured information for hypothesis generation. 2016
40. Cheng H, Phillips M. Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry* 2014;**26**:371-75.
41. Biesecker L. Hypothesis-generating research and predictive medicine. *Genome Res* 2013;**23**:1051-53.
42. Jing X, Zhou Y, Cimino J, et al. Development, validation, and usage of metrics to evaluate clinical research hypothesis quality. *Journal of Clinical and Translational Science, under review* 2023 doi: https://www.medrxiv.org/content/10.1101/2023.01.17.23284666v2
43. Jing X, Zhou YC, Cimino JJ, et al. Development and preliminary validation of metrics to evaluate data-driven clinical research hypotheses. AMIA 2022; 2022 Nov 5-9, 2022; Washington DC.
44. Hicks CM. Research methods for clinical therapists: Applied project design and analysis. 1999
45. Hulley S, Cummings S, Browner W, Grady D, Newman T. Designing clinical research. 2013
46. Glasser SP. Essentials of clinical research. 2014
47. Portney LG. *Foundations of Clinical Research: Applications to Evidence-based Practice*: F.A. Davis, 2020.
48. Gallin JI, Ognibene FP, Ognibene FP. *Principles and Practice of Clinical Research*. Burlington, UNITED STATES: Elsevier Science & Technology, 2007.
49. Spangler S. Accelerating discovery: Mining unstructured informaiton for hypothesis generation. 2016

50. Emerson M, Brooks M, Masters D, et al. Improved visualization of hierarchical datasets with VIADS. AMIA Annual Symposium. San Francisco, 2018:1956.
51. Jing X, Emerson M, Gunderson D, et al. Architecture of a visual interactive analysis tool for filtering and summarizing large data sets coded with hierarchical terminologies (VIADS). AMIA Summits Transl Sci Proc 2018:444-45.
52. Jing X, Patel VL, Cimino JJ, et al. The Roles of a Secondary Data Analytics Tool and Experience in Scientific Hypothesis Generation in Clinical Research: Protocol for a Mixed Methods Study. *JMIR Res Protoc* 2022;**11**(7):e39414. doi: 10.2196/39414
53. Haendel M, Chute C, Gersing K. The National COVID Cohort Collaborative (N3C): Rationale, Design, Infrastructure, and Deployment. *J Am Med Inform Assoc* 2020 doi: https://doi.org/10.1093/jamia/ocaa196
54. Draghi B, Ernst M, Patel V, et al. Number of scientific hypotheses and time needed in a 2-hour study session among inexperienced clinical researchers—preliminary results. AMIA Summit 2023; 2023 Mar 13-16, 2023; Seattle, Washington.
55. Misra DP, Gasparyan AY, Zimba O, Yessirkepov M, Agarwal V, Kitas GD. Formulating Hypotheses for Different Study Designs. *J Korean Med Sci* 2021;**36**(50):e338. doi: https://doi.org/10.3346/jkms.2021.36.e338
56. Foster JG, Rzhetsky A, Evans JA. Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review* 2015;**80**(5):875-908.
57. Sprenger AM, Dougherty MR, Atkins SM, et al. Implications of cognitive load for hypothesis generation and probability judgment. *Front Psychol* 2011;**2**:129. doi: 10.3389/fpsyg.2011.00129
58. Johnson S. *Where good ideas come from: the natural history of innovation*. New York: Riverhead Books, 2010.
59. Jing X, Patel V, Cimino J, Shubrook J. Hypothesis generation in clinical research: challenges, opportunities, and role of AI. MIE 2022; 2022 May 27-30, 2022; Nice, France. IOS.
60. Lipowski EE. Developing great research questions. *Am J Health Syst Pharm* 2008;**65**:1667-70. doi: https://doi.org/10.2146/ajhp070276
61. Haynes R. Forming research questions. *J Clin Epidemiol* 2006;**59**:881-86.
62. Browner W, Newman T, Cummings S, et al. *Designing Clinical Research*. 5th ed. Philadelphia, PA: Wolters Kluwer, 2023.
63. Alison L, Doran B, Long ML, Power N, Humphrey A. The effects of subjective time pressure and individual differences on hypotheses generation and action prioritization in police investigations. *Journal of experimental psychology. Applied* 2013;**19**(1):83-93. doi: 10.1037/a0032148