

Perceptions of AI-Assisted Voice Explanations for Medical Diagnostic Reports in Multilingual India

Kabir Gupta

kabirgupta@woodstock.ac.in

ABSTRACT

Advances in medical diagnostic technologies have resulted in an increasing number of diagnostic reports being available on digital channels; however, technical jargon, numerical results and reporting in a foreign language (English) can still be hard for participants to understand. This becomes an important question in a multilingual country like India in the sense that language and health-literacy requirements may play a role in the way a participant makes sense of the diagnostic data. The current study was based on a qualitative-dominant mixed-methods design to investigate participants' awareness of medical reports and their opinions on a proposed AI-based voice explanation system in their preferred language. MediTrack was described to participants as a proposed service rather than deployed, and participants reported their perceptions of it rather than a tested or compared experience.

Descriptive survey procedures and conservative thematic analysis of Hindi/Hinglish interviews has been used in the research. The meaning units were translated to categories and themes and translation and transcription limitations were addressed with caution. Empirical evidence is limited to the Analysis. AI-assisted explanation as a form of communication aid that should be monitored clinically.

Keywords: artificial intelligence, medical reports, voice explanations, health literacy, participant comprehension, code-switching, India.

1. INTRODUCTION AND BACKGROUND

Medical diagnostic reports aid in diagnosis, treatment and follow-up; but the language, abbreviations used, reference ranges and numerical values can make them difficult to interpret. Moreover, understanding, decision and recall is also influenced by presentation (Zhang et al., 2020; van der Mee et al., 2024). There is therefore a need for a participant's access to be joined by explanations which allow for discussions with clinicians and enable appropriate action.

An important context in which to view this problem is the concept of health literacy. It comprises the capacity to access, comprehend, evaluate and apply health information, not just the capacity to read

(Berkman et al., 2011; Sørensen et al., 2012). In reality, specialist abbreviations and diagnostic terms may continue to be challenging even for individuals who are comfortable with reading everyday terms. This separation can facilitate analysis of patient understanding of the vocabulary and numerical data presented and where clarification is needed.

In India, there is an additional level – language. Multilingual communication challenges have been documented in Indian healthcare and it has been found that communication problems can be influenced by language, use of health literacy, workload and clinical hierarchy (Narayan, 2013; Douglass et al., 2021). While language concordance may help communication, it does not eliminate communication barriers (Diamond et al., 2019). These results give background information when studying preferred-language explanation and code switching (Hindi-English) as communication preference, and not as an effective or less effective choice.

Although AI-Assisted communication can help make communication easier in the clinical setting, evidence also identifies critical limitations. In the study of conversational agents and generative Artificial Intelligence, improvements in accessibility or understandability have been noted, alongside variation in the quality of evaluation (Laranjo et al., 2018), and, in more recent generative-AI studies, risk of the emergence of hallucination and potential for inaccuracies or harmful simplifications (Park et al., 2024; Tepe & Emekli, 2024). As such, this paper will address comprehension, trust and language and voice issues as communication problems and consider AI as an aid that needs to be managed appropriately (World Health Organization, 2021).

2. RESEARCH PROBLEM, RESEARCH GAP, AND RATIONALE

2.1 RESEARCH PROBLEM

What we are seeing is that getting a medical report does not mean people can confidently understand key words, values or implications of what they are reading, and this has been shown in health-literacy and laboratory-result research (Berkman et al., 2011; Zhang et al., 2020; van der Mee et al., 2024).

2.2 RESEARCH GAP

Patient-friendly AI research has focused on readability, perceived understandability and accuracy (Park et al., 2024; Tepe & Emekli, 2024), while voice-assistant studies have focused on accessibility, usability and readiness (Masina et al., 2020; Sezgin et al., 2020). While there is empirical evidence on language barriers from India (Narayan, 2013; Douglass et al., 2021), its findings on participants' perception of AI-assisted, preferred-language voice support for medical reports are limited. This study focuses on this intersection with comprehension, language, trust and preferences for communication modes and privacy.

2.3 RESEARCH RATIONALE

Limited access to care due to difficulties in being able to understand the information provided on a diagnostic tool can also be a source of confusion and/or anxiety (Zhang et al., 2020; van der Mee et al., 2024). To explore the communication needs and forms of explanation preferred, the study therefore adopted an exploratory patient-centred design with the empirical material being held for the Analysis section.

3. LITERATURE AND CONCEPTUAL BACKGROUND

3.1 LANGUAGE BARRIERS IN INDIAN HEALTHCARE

In a healthcare context, language can't be an aside in India, and its relevance is not just to the extent of sharing or not sharing clinical information, but to sharing at all, or even sharing safely. In the case of linguistic diversity, Narayan (2013) proposed that because healthcare may not be available in a language that participants understand or because formal medical communication may be carried out in English, this diversity could be disadvantageous to the participants. A further piece of evidence highlighting the issue can be found in emergency departments from India, as carried out by Douglass et al. (2021). Their study found that there was lots of multi-lingual communication from clinicians and that communication issues, in addition to language differences, were linked to low health literacy and workload and hierarchy issues. When viewed together these studies demonstrate that the issue is a larger one than simply translation. When the participant is a subject of consultation, explanation or a report, understanding of common day to day language like Hindi or Marathi or some other language vocabulary instead of medical English language vocabulary may suffice.

International evidence also helps to make the point on the role of language. Research in Diamond et al., 2019 revealed that in a systematic review, patient-physician language concordance had been found to be associated with better communication and care processes and outcomes. In multilingual settings, however, other factors, such as time pressures and jargon and hierarchical relationships within professional communities can also hinder explanation (Douglass et al., 2021). Language matching can be helpful, but is insufficient on its own.

3.2 PARTICIPANT UNDERSTANDING OF MEDICAL INFORMATION

The growing body of health literacy studies clearly differentiate between accessing information and understanding it. In a large systematic review by Berkman et al. (2011) low-health literacy was linked with lower health knowledge and less effective health care information seeking. Sørensen et al. (2012) also saw health literacy as a problem on how to find information but also to comprehend, evaluate and utilise it. This distinction is important with diagnostic reports in particular as recognizing that a result is 'high' or 'low' is not the same as understanding what the result means, how certain this result is or what to do next. In India, Dsouza et al. (2021) demonstrated the need to adapt and validate health-literacy measures for local language contexts, while Mathias et al. (2023) found that comparatively few

community-based health-literacy interventions in India had been formally evaluated. Together, these studies suggest that participant understanding is shaped by literacy, language, numeracy and the way information is presented, rather than by education level alone.

Studies focused directly on test results make the comprehension problem more concrete. Zhang et al. (2020) found that participants often had difficulty interpreting laboratory results and identifying what those results meant for their health and subsequent decisions. Van der Mee et al. (2024), reviewing presentation formats for laboratory results, likewise showed that design and explanation influence perception, decision-making, action and memory. The two studies converge on the importance of explanation, but they also expose a measurement problem: information that feels clear may still be misunderstood. Talevski et al. (2020) found that teach-back can support confirmation of understanding by asking participants to explain information in their own words. Teach-back therefore remains an important option for future comparative evaluation of patient-facing explanations.

3.3 AI AND PARTICIPANT COMMUNICATION

Artificial intelligence has increasingly been studied as a way to simplify technical health information, but the evidence remains mixed. Laranjo et al. (2018), in a systematic review of conversational agents in healthcare, found promising uses for patient support and information delivery, while also noting wide variation in study quality and outcome measurement. More recent generative-AI studies have moved closer to the problem addressed here. Park et al. (2024) evaluated AI-generated patient-friendly radiology reports and found substantially higher perceived comprehension scores for the simplified versions than for original reports. Tepe and Emekli (2024) similarly reported that large language models can translate or explain radiological jargon in more accessible terms. These findings make AI-assisted explanation plausible, but they do not establish that participants can safely interpret the clinical meaning of a report or choose an appropriate next step.

But this is a critical reading; making the difference between readability and comprehension; convenience and safety. Park et al. (2024) found hallucinations and potentially harmful translations, which illustrates that fluent output may change or add clinical meaning. Where there is translation, simplification or verbalisation of content, the risk is further increased, meaning that caution is needed where there is human operation. (World Health Organization, 2021).

3.4 VOICE COMMUNICATION AND CODE-SWITCHING

Some users may find it helpful to have a voice communication system to support access to an electronic communication system. Some individual skills however are determining factors, the design of the interface, readiness and acceptability (Masina et al., 2020; Sezgin et al., 2020). The results support the consideration of voice preference alone from the objectively proven comprehension.

Furthermore, in the case of multilingual contexts, the use of spoken language is an added dimension to the difficulties of communication. Switching languages/dialects/tone or using different words can be appropriate to establish rapport, but, as noted by Wood (2019) when not it can be patronising and/or

culturally insensitive. These are complemented by more empirical data from bilingual medical consultations – 586 recorded consultations of Galician use in medical practice in Galicia – that show that bilingualism in medical contexts is associated with meaning negotiation, emphasising information and dealing with interpersonal aspects of the medical consultation (Rodríguez Tembrás, 2024). This evidence, although external to India, is relevant as it shows that language of bilingual medical talk is not a simple issue of translating from one language to another – but still bilingual speakers shift between the codes for practical and/or relational purposes. In one sense, code switching in Hindi-English can be communicative in a similar way but this needs to be quality-tested and cannot be presumed to be the same.

Based upon the literature examined, it may be presumed that there is some real utility in the use of explanations in both languages (spoken and mixed) and that their effectiveness has not been finally decided. These considerations will make it easier to consider 'usability', 'code switching', 'understanding' and 'clinical protection'.

4. AIM, OBJECTIVES AND RESEARCH QUESTIONS

4.1 RESEARCH AIM

The study examined how adults in India interpreted medical reports and perceived AI-assisted, preferred-language voice support, including issues of trust, privacy, accessibility and clinical confirmation.

4.2 RESEARCH OBJECTIVES

The study was guided by the following objectives:

1. To examine the difficulties participants reported when interpreting medical terminology, numerical values and report meaning.
2. To explore how language, health literacy, education, digital access, anxiety and doctor-patient communication were related to participants' reported understanding.
3. To describe participant preferences and perceptions regarding preferred-language explanation, voice communication and Hindi-English code-switching.
4. To examine participant views of trust, privacy, safety and the appropriate role of clinical confirmation in AI-assisted explanation.

4.3 MAIN RESEARCH QUESTION

How did participants experience and interpret medical reports, and what did they report about the potential role, acceptability and safeguards of AI-assisted preferred-language voice explanations?

4.4 SUPPORTING RESEARCH QUESTIONS

RQ1: What difficulties did participants report when interpreting medical terminology, values and the meaning of diagnostic reports?

RQ2: How did language, health literacy, education, digital access, anxiety and doctor-patient communication shape reported understanding?

RQ3: What preferences and accessibility needs did participants report regarding preferred language, voice communication and Hindi-English code-switching?

RQ4: What concerns and expectations did participants report regarding trust, privacy, safety and clinical confirmation in AI-assisted explanations?

5. RESEARCH METHODOLOGY

5.1 RESEARCH DESIGN

The study used a qualitative-dominant mixed-methods design informed by phenomenology. Phenomenology is a qualitative tradition concerned with understanding how people experience, interpret and give meaning to events in everyday life, with particular attention to participants' accounts of lived experience (Smith et al., 2022). This orientation was appropriate for examining how participants made sense of medical terminology, language difficulty, anxiety, trust and AI-assisted explanation. The quantitative component was designed to provide descriptive context, while the qualitative component explored the meanings attached to those experiences. Integration occurred during interpretation.

5.2 PARTICIPANTS AND SAMPLING

Purposive sampling was used to recruit adults who had previously received a diagnostic report. Participants were approached at a health camp organised by Rajyvansh Hospital and Institute of Medical Science in Muzaffarnagar, Uttar Pradesh, where those who had recently received a diagnostic report were invited to take part. Recruitment sought variation in age, gender, education, preferred language, English proficiency and digital confidence. The qualitative component was treated as exploratory lived-experience inquiry rather than as a basis for claiming exhaustive thematic coverage (Malterud et al., 2016).

5.3 SURVEY AND INTERVIEW DATA COLLECTION

The survey recorded demographic characteristics and structured items concerning understanding of report results, hesitation in repeatedly asking a doctor, anxiety about results, language barriers, perceived usefulness of Meditrack and preference for voice chat over text. The full survey instrument is provided in Appendix A. The interviews were conducted in Hindi/Hinglish and addressed report difficulty, unfamiliar medical terms, help-seeking, hesitation, language preference, confidence and the perceived usefulness of WhatsApp-based explanation. MediTrack was described to participants as a proposed AI-based

voice-explanation service; it was not deployed, and participants were not asked to use it or compare it against an actual system. The study also did not include a controlled comparison between conventional reports, AI-simplified text and AI-assisted voice.

5.4 MEASURES AND DATA PREPARATION

Survey items were prepared for descriptive analysis without assigning interpretive labels that were not supported by the source materials. Interview files were checked for duplication, assigned pseudonymous participant identifiers and reviewed for automated-transcription problems, truncated passages and limited speaker separation. Only clearly attributable responses were eligible for quotation or frequency-based interpretation. All resulting counts, percentages and source-specific observations are reported in the Analysis section.

5.5 ANALYTICAL PROCEDURE

The quantitative component was analysed descriptively using frequencies and percentages; no inferential comparison of communication conditions was performed. Qualitative analysis progressed from transcript meaning units to initial codes, categories, themes and interpretation. Iterative reading and code clustering were followed by comparison with corresponding survey domains. Hindi/Hinglish quotations were translated into English and lightly edited for punctuation and intelligibility without changing meaning. Ambiguous, interviewer-led or truncated fragments were excluded from quotation-level claims. This audit trail linked source responses to codes and themes (Braun & Clarke, 2019; Nowell et al., 2017).

5.6 ETHICS AND TRUSTWORTHINESS

Formal institutional ethics review was not obtained, however verbal informed consent was obtained from each participant before the interview and survey. Interview introductions informed participants that personal health details were not required and requested permission to record. Identifiers were removed from the analytical presentation and replaced with pseudonymous labels. Trustworthiness was strengthened through duplicate-file checking, transcript-by-transcript coding, a source-to-theme audit trail, inclusion of contrasting responses and conservative exclusion of unclear automated-transcription fragments. Reflexive interpretation distinguished participant responses from leading prompts and avoided claims of experimental effectiveness or population-wide generalisability (Nowell et al., 2017).

6. ANALYSIS AND THEMATIC FINDINGS

6.1 RESPONDENT PROFILE AND SURVEY FINDINGS

After duplicate workbooks and transcript files were excluded, the final analytical corpus contained the survey records and unique interviews summarised in Table 1. Survey ages ranged from 18 to 88 years, with a mean of 48.0 years.

Oct 2026

Vol 12, No 1.

Table 1. Respondent Profile and Data Sources

Measure	Value	Percentage
Survey respondents	103	100.0%
Age range	18–88 years	—
Male	52	50.5%
Female	51	49.5%
Unique interview transcripts	9	-

Table 2. Descriptive Summary of Verified Survey Indicators

Survey indicator	Reported category grouping	n	%
Understanding report results	Categories 1–2	63	61.2
Hesitation in repeatedly asking doctor	Categories 1-2	86	83.5
Worry/anxiety about results	Categories 4-5	48	46.6
Language difficulty	Categories 4–5	84	81.6
Meditrack usefulness	Categories 4–5	103	100.0
Voice rather than text	Categories 4–5	99	96.1

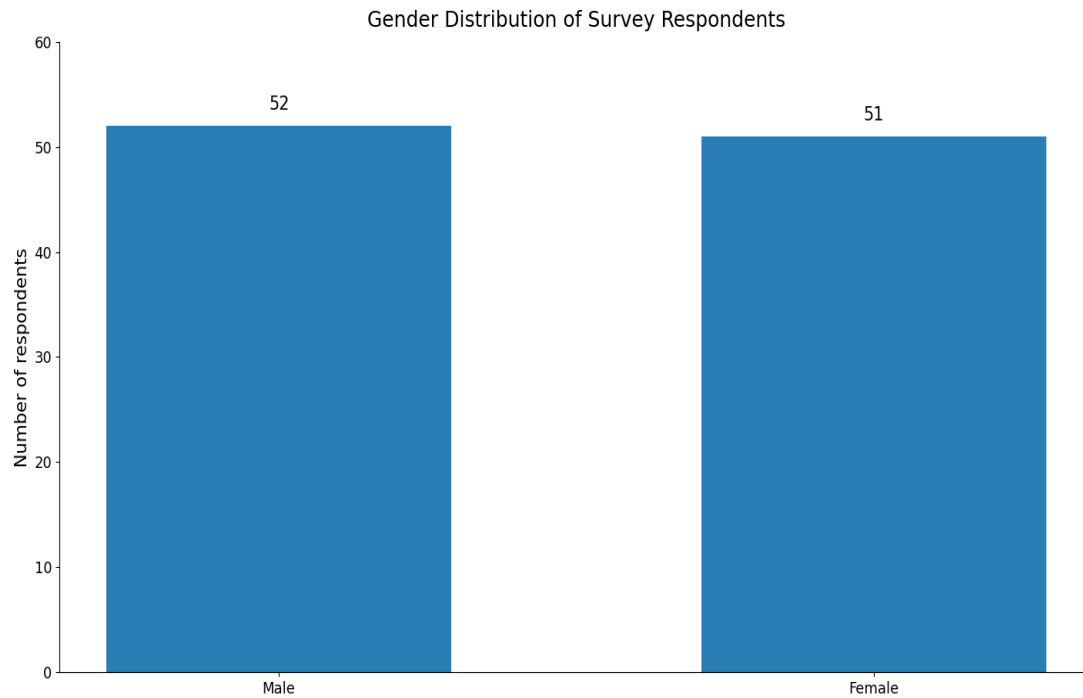


Figure 1. Gender Distribution of Survey Respondents

Figure 1 shows an almost even gender split, limiting simple numerical dominance by either group. Because indicator-by-gender cross-tabulations were unavailable, subgroup response differences could not be assessed.

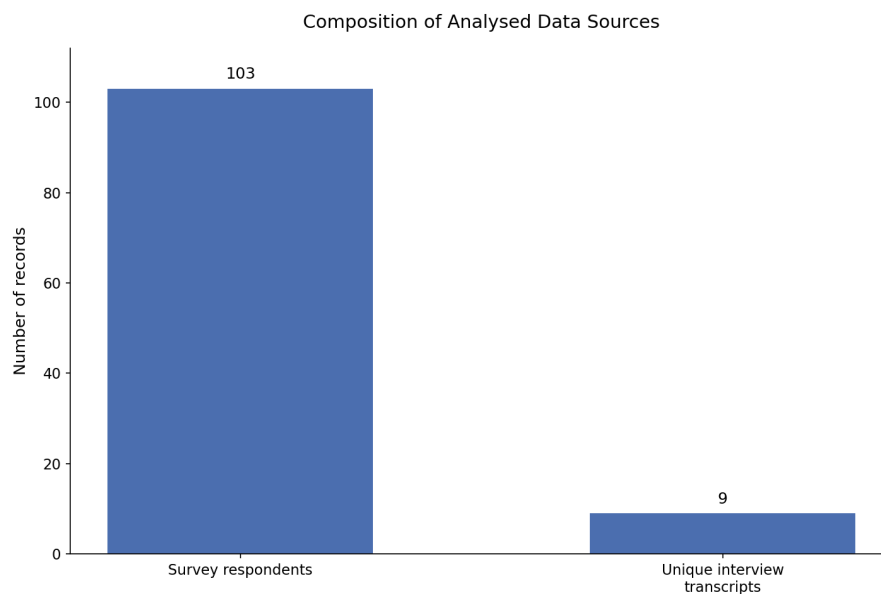


Figure 2. Composition of Analysed Data Sources

Figure 2 illustrates the larger survey component within the combined corpus. The interviews supplied context for difficulties with technical English, information-seeking through Google, AI or clinicians, and the importance of familiar language and privacy. They were used to qualify the survey patterns, not as a separate experimental condition.

The response anchors were not provided so that for Tables 2 and the following figures the original numerical category response are retained without being labeled as “agree” or “disagree.”

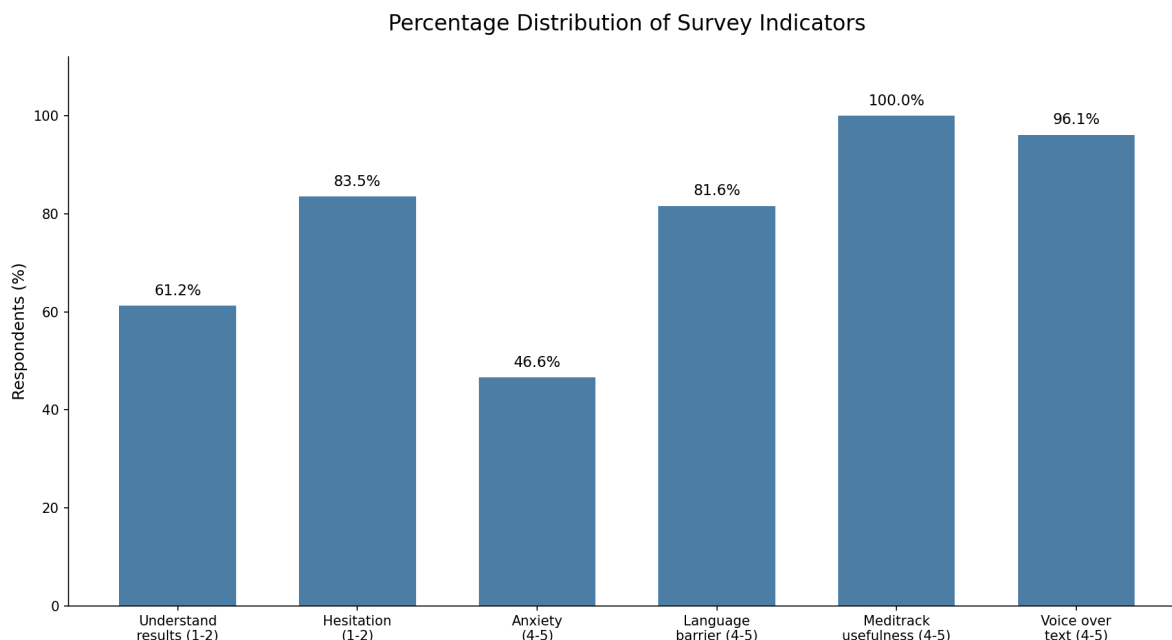


Figure 3. Percentage Distribution of Survey Indicators

Figure 3 presents the grouped counts as percentages – a proportional restatement of the same grouped counts, not independent evidence . It should therefore be interpreted with the same caution concerning missing response anchors and causal inference.

Accordingly, the survey results are used only as descriptive context for the interview themes.

6.2 PARTICIPANT-LEVEL INTERVIEW EVIDENCE AND DATA QUALITY

Nine unique interview transcripts were reviewed after duplicate copies were removed. Five transcripts (P1, P3, P4, P5 and P7) contained clearly attributable substantive participant responses. P2 and P9 contained some usable participant wording but had merged speaker turns, while P6 and P8 were heavily interviewer-led or truncated. Table 3 therefore records the source location and evidential status for every

participant; interviewer prompts and uncertain speaker turns were not counted as confirmed participant views. Quotations are English translations of the Hindi/Hinglish transcript wording.

Table 3. Participant-Level Interview Evidence and Evidential Status

ID	Data-grounded extract and source location	Initial code	Evidential status
P1	At 24.9–42.8 s, the participant described the report as difficult and said AI was used before understanding was achieved.	Difficulty; AI-assisted clarification	Clear participant response
P2	A short response at 95.0–102.7 s indicated greater confidence when information was provided in Hindi.	Hindi; confidence	Limited: short phrase within a speaker-merged turn
P3	At 65.8–75.3 s, the participant acknowledged hesitation but said asking remained necessary; at 123.7–138.8 s, familiar-language explanation was linked to confidence and follow-up.	Hesitation; confidence; follow-up	Clear participant responses
P4	Some medical words were difficult to pronounce and also “thyroid deficiency” was not understood (36.9–56.8 s); Google was searched even prior to seeking a doctor's advice (76.8–81.8 s).	Terminology; Google; clinician confirmation	Clear participant responses
P5	At 22.0–28.5 s, the participant described the report as somewhat difficult because it was in English.	English-language difficulty	Clear but brief participant response
P6	The merged 31.1–60.2 s segment contains “Google and doctor,” but the speaker boundary is not reliable.	Possible external clarification	Not used for quotation-level or frequency claims
P7	It feels like it is really challenging, it is in English and there are also doctors words” (19.9–49.9 s); Google was first attempted, then if there was still uncertainty, then a doctor was called (58.9–88.9 s).	English; technical terms; staged information-seeking	Clear participant responses
P8	The 61.5–95.5 s segment includes “English is the main problem,” but the turn is truncated and interviewer/participant speech is merged.	Possible English-language barrier	Not used for quotation-level or frequency claims

ID	Data-grounded extract and source location	Initial code	Evidential status
P9	"Complex words" caused difficulty "many times" with some medical terms (94.5–100.6 s); examples in the merged transcript included bradycardia and tachycardia (30.7–60.9 s).	Complex terminology	Limited: participant wording present, speaker boundary partly unclear

6.3 THEME 1: PARTICIPANT COMPREHENSION AND INFORMATION-SEEKING

The first theme concerned the gap between receiving a report and converting it into usable understanding. In the survey, 63 of 103 respondents (61.2%) selected categories 1–2 for understanding report results. Because the response anchors were unavailable, this numerical grouping is reported without assigning an agreement label. The interviews supplied clearer explanations of the difficulty: P1, P4, P5 and P7 described technical or English-language barriers, while P9 referred to recurring difficulty with complex medical terms in a partly speaker-merged transcript.

"It was a little difficult ... I used AI to find out about it; only then did I understand." (P1, 24.9–42.8 s; translated from Hindi)

"Some medical words are difficult even to pronounce ... I did not understand thyroid deficiency." (P4, 36.9–56.8 s; translated from Hindi)

"It was a little difficult ... because of English." (P5, 22.0–28.5 s; translated from Hindi)

"It feels like it is really challenging, it is in English and there are also doctors' words." (P7, 19.9–49.9 s; translated from Hindi)

"Complex words ... caused difficulty many times ... bradycardia, tachycardia [were unclear]." (P9, 94.5–100.6 s and 30.7–60.9 s; translated from Hindi; speaker boundary partly unclear)

P1 (AI), P3 (Google), P4 (Google then medical advice) and P7 (Google, then doctors in case of any doubt) directly described information seeking.

These accounts suggest an interpretation that is twofold—in the first phase the digital tools provided clarity; in the second, where the report was still unclear, clinician input was important. P6 was not included in this claim because the "Google and doctor" wording appeared to be in a merged speaker turn.

6.4 THEME 2: ANXIETY, CONFIDENCE AND CONDITIONAL TRUST

The results showed that on the worry/concern item 48 out of 103 participants (46.6%) were on the "worry" end of the continuum (scores of 4 to 5) and on the hesitation item 86 (83.5%) were on the "hesitate" end of the continuum (scores of 1 - 2). These numbers correspond to concentration of categories with anchors from scale not completed. During the interviewing the P3 expressed his own personal feeling of ambivalence regarding asking again and P4 reported the recollection of a case that was balanced regarding his feeling about asking again.

"Perhaps there's a degree of hesitation, but . . . it's necessary to ask." (P3, 65.8–75.3 s; translated from Hindi). By contrast, P4 reported feeling comfortable asking: "I was comfortable asking." (P4, 56.8–72.3 s; translated from Hindi)

Once and for all, the explanation wasn't digital. Within its abbreviated natural time zone (141.0–146.0 s), P4 expressed a willingness to support only with the promise that there was no sharing of private information and privacy was assured. "Round the clock if there is a problem, they can call for a physician using WhatsApp which is really helpful," added P3. All of these studies as a whole suggest an alternative paradigm that builds on a conceptual framework that explanations can be transmitted digitally while still retaining the need for maintaining privacy protection and clinical transparency.

6.5 THEME 3: LANGUAGE, VOICE AND ACCESSIBILITY

The third theme was related to the known language and mode of communication. In the survey, 84 of 103 respondents (81.6%) selected categories 4–5 for the language-barrier item, all 103 selected categories 4–5 for Meditrack usefulness, and 99 (96.1%) selected categories 4–5 for preferring voice over text. These are descriptive groupings, not evidence that voice caused better comprehension. Interview evidence was strongest for familiar-language confidence: P2 offered a short Hindi-confidence response, while P3 and P4 provided fuller accounts.

"More confidence in Hindi." (P2, 95.0–102.7 s; short response in a speaker-merged turn; translated from Hindi)

"If it is explained in my normal language, my confidence level will definitely increase." (P4, 126.5–132.0 s; translated from Hindi)

P3 similarly said that familiar-language explanation would increase confidence and support follow-up (123.7–138.8 s). P7 described Hindi explanation as helpful, although the speaker boundary was less distinct. The interviews therefore support perceived usefulness of familiar-language explanation, while the survey alone supplies the explicit voice-over-text preference. Leading questions and unclear responses in P6 and P8 were not treated as evidence of benefit.

7. DISCUSSION AND CONTRIBUTION

The paper's main contribution is an integrated framework linking comprehension, information-seeking and clinical clarification. It distinguishes information availability from usable understanding and situates report design within participant interpretation and decision-making (Zhang et al., 2020; van der Mee et al., 2024).

A second contribution is separating language choice from communication mode. Indian studies foreground multilingual barriers (Narayan, 2013; Douglass et al., 2021), whereas voice-assistant research concentrates on accessibility, usability and readiness (Masina et al., 2020; Sezgin et al., 2020). Evaluation needs to therefore look at the clarity, accuracy and inclusiveness instead of 'gist' as proof of comprehension.

Lastly, AI-powered explanations can be most effectively viewed as an adjunct to clinical interpretation. Inaccuracies, omissions of context or unwarranted certainty, can occur when "generated content" enhances readability. It is crucial to include essential safeguards, such as clinician review, privacy protection, clear scope and limits, and a direct check for understanding (Park et al., 2024; Tepe & Emekli, 2024; World Health Organization, 2021).

8. CONCLUSION, LIMITATIONS AND FUTURE SCOPE

Medical-report explanations need to be assessed in four areas: accessibility, readability, participant preference, and showing understanding of the explanation. Language choice and voice interfaces can be alternate supports within a structured and supervised facilitative communication environment that values privacy.

If one is claiming superiority, then one must have the controlled comparison with information in the array of the communication modes, proven outcomes and objective comprehension or teaching back process.

8.1 LIMITATIONS

Several methodological restrictions limit the scope of the study. Because only five of nine interview transcripts (P1, P3, P4, P5 and P7) yielded clearly attributable participant responses, Themes 1–3 should be read as illustrative patterns from a handful of participants rather than as representative of Indian participants as a group. The research is qualitatively exploratory, and the use of purposive sampling limits generalisability. There are potential risks of interviewers' interpretation errors and response bias when using automated transcription and asking interviewer-led questions. Validated instruments to measure demographic cross-tabulation, conventional reports, simplified text, voice explanation, health literacy, anxiety, confidence and trust were not included in the survey instrument, and a controlled comparison between conventional report and simplified text/simplified voice explanation was not designed. These methodologic constraints make causal, longitudinal and population level interpretation problematic.

8.2 FUTURE SCOPE

Using improved/cleaner speaker-labelled transcripts and recruiting a larger, more diverse qualitative sample, representing regional languages and healthcare settings, should be addressed in future research. Equivalent text and voice explanations should be tested on objective comprehension/teach-back measures, along with measures of literacy, digital confidence, anxiety, privacy, pace, pronunciation/replayability.

APPENDIX A: SURVEY INSTRUMENT

1. Think about the last time a doctor gave you a report or wrote you a prescription. How easy or hard was it to understand it ?
2. Were there any words or terms the doctor used that you didn't understand? What did you do at that moment?
3. Did you feel comfortable asking the doctor to explain something again, or did you feel hesitant?
4. If you don't fully understand something in your medical report, who or what do you turn to for help – family, a pharmacist, the internet, or someone else?
5. Do you think your education, social class, or how comfortable you are with English affects how much you understand or how easily you can ask questions?
6. Has not fully understanding a report or prescription ever affected how you took your medicines or whether you went back for a follow-up visit?
7. Do you think having someone explain your medical reports to you in your own mother tongue will make you more relieved or confident about the diagnosis?
8. Are you comfortable using WhatsApp? What do you feel about having someone explain your diagnosis using voice notes?

REFERENCES

1. Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, 155(2), 97–107. <https://doi.org/10.7326/0003-4819-155-2-201107190-00005>
2. Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
3. Diamond, L., Izquierdo, K., Canfield, D., Matsoukas, K., & Gany, F. (2019). A systematic review of the impact of patient–physician non-English language concordance on quality of care and outcomes. *Journal of General Internal Medicine*, 34(8), 1591–1606. <https://doi.org/10.1007/s11606-019-04847-5>
4. Douglass, K., Narayan, L., Allen, R., Pandya, J., & Talib, Z. (2021). Language diversity and challenges to communication in Indian emergency departments. *International Journal of Emergency Medicine*, 14, Article 57. <https://doi.org/10.1186/s12245-021-00380-7>
5. Laranjo, L., Dunn, A. G., Tong, H. L., Kocaballi, A. B., Chen, J., Bashir, R., Surian, D., Gallego, B., Magrabi, F., Lau, A. Y. S., & Coiera, E. (2018). Conversational agents in healthcare: A systematic review. *Journal of the American Medical Informatics Association*, 25(9), 1248–1258. <https://doi.org/10.1093/jamia/ocy072>
6. Masina, F., Orso, V., Pluchino, P., Dainese, G., Volpato, S., Nelini, C., Mapelli, D., Spagnoli, A., & Gamberini, L. (2020). Investigating the accessibility of voice assistants with impaired users: Mixed methods study. *Journal of Medical Internet Research*, 22(9), e18431. <https://doi.org/10.2196/18431>
7. Narayan, L. (2013). Addressing language barriers to healthcare in India. *The National Medical Journal of India*, 26(4), 236–238. <https://pubmed.ncbi.nlm.nih.gov/24758452/>
8. Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16, 1–13. <https://doi.org/10.1177/1609406917733847>
9. Park, J., Oh, K., Han, K., & Lee, Y. H. (2024). Patient-centered radiology reports with generative artificial intelligence: Adding value to radiology reporting. *Scientific Reports*, 14, Article 13218. <https://doi.org/10.1038/s41598-024-63824-z>
10. Sezgin, E., Huang, Y., Ramtekkar, U., & Lin, S. (2020). Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *npj Digital Medicine*, 3, Article 122. <https://doi.org/10.1038/s41746-020-00332-0>
11. Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12, Article 80. <https://doi.org/10.1186/1471-2458-12-80>
12. Talevski, J., Wong Shee, A., Rasmussen, B., Kemp, G., & Beauchamp, A. (2020). Teach-back: A systematic review of implementation and impacts. *PLOS ONE*, 15(4), e0231350. <https://doi.org/10.1371/journal.pone.0231350>

13. Tepe, M., & Emekli, E. (2024). Decoding medical jargon: The use of AI language models (ChatGPT-4, BARD, Microsoft Copilot) in radiology reports. *Patient Education and Counseling*, 126, 108307. <https://doi.org/10.1016/j.pec.2024.108307>
14. World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance*. <https://www.who.int/publications/i/item/9789240029200>
15. Zhang, Z., Citardi, D., Xing, A., Luo, X., Lu, Y., & He, Z. (2020). Patient challenges and needs in comprehending laboratory test results: Mixed methods study. *Journal of Medical Internet Research*, 22(12), e18725. <https://doi.org/10.2196/18725>
16. van der Mee, F. A. M., Schaper, F., Jansen, J., Bons, J. A. P., Meex, S. J. R., & Cals, J. W. L. (2024). Enhancing patient understanding of laboratory test results: Systematic review of presentation formats and their impact on perception, decision, action, and memory. *Journal of Medical Internet Research*, 26, e53993. <https://doi.org/10.2196/53993>
17. Dsouza, J. P., Van den Broucke, S., & Pattanshetty, S. (2021). Validity and reliability of the Indian version of the HLS-EU-Q16 questionnaire. *International Journal of Environmental Research and Public Health*, 18(2), 495. <https://doi.org/10.3390/ijerph18020495>
18. Mathias, E. G., Dhyani, V. S., Krishnan, J. B., Rani, U., Gudi, N., & Pattanshetty, S. (2023). Community based health literacy interventions in India: A scoping review. *Clinical Epidemiology and Global Health*, 22, 101310. <https://doi.org/10.1016/j.cegh.2023.101310>
19. Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretative phenomenological analysis: Theory, method and research* (2nd ed.). SAGE.
20. Wood, N. I. (2019). Departing from doctor-speak: A perspective on code-switching in the medical setting. *Journal of General Internal Medicine*, 34(3), 464-466. <https://doi.org/10.1007/s11606-018-4768-0>
21. Rodríguez Tembrás, V. (2024). Two languages, one goal: Code-switching in doctor-patient communication in the Galician healthcare system. *Languages*, 9(6), 209. <https://doi.org/10.3390/languages9060209>
22. Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753-1760. <https://doi.org/10.1177/1049732315617444>