Title: The Observer Repository: Advancing Ambulatory Care Innovation Through Video-Based Clinical Ethnography

Authors

Kevin B. Johnson, MD, MS, FACMI, FAMIA, FAAP, FIAHSI, FAIMBE (corresponding author)
David L. Cohen University Professor
Division of Biomedical Informatics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine at the University of Pennsylvania
Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania
Kevin.johnson1@pennmedicine.upenn.edu

Basam Alasaly, MS
Division of Biomedical Informatics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania

Kuk Jin Jang, PhD
Division of Biomedical Informatics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine at the University of Pennsylvania
Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania

Eric Eaton, PhD
Department of Computer and Information Science, School of Engineering and Applied Science, University of Pennsylvania

Sriharsha Mopidevi, MS
Division of Biomedical Informatics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania

Ross Koppel, PhD, MA, FACMI, FIAHSI
Division of Biomedical Informatics, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine at the University of Pennsylvania
Professor of Sociology, School of Arts and Sciences, University of Pennsylvania
rkoppel@sas.upenn.edu

Word Count – Abstract 130/150 words. Manuscript 3125 words.

# Abstract

Conventional biomedical research often yields incremental improvements in healthcare technologies, yet these fall short in addressing broader systemic issues—such as clinician burnout, regulatory complexity, and electronic health record (EHR) inefficiencies. To foster transformative innovation, researchers must engage with the sociotechnical realities of modern care delivery.

We present the Observer Repository, a platform that captures real outpatient visits through high-resolution video, structured metadata, and collaborative annotations. This system enables detailed analysis of clinical routines, team interactions, and workflow dynamics, supporting co-design and reflection among clinicians, informaticians, and designers.

To date, Observer includes over 100 recorded visits from adult primary care practices at the University of Pennsylvania Health System, with associated demographic, diagnostic, and satisfaction data. It offers a unique resource for education, workflow analysis, and model development—providing structured, real-world clinical access to support innovation in ambulatory care delivery.

# Introduction

The U.S. healthcare system—once a global exemplar—is now criticized for its high costs and relatively poor health outcomes.[1] While biomedical researchers and industry leaders have proposed incremental improvements to existing technological infrastructure—once seen as key to reform[2–4]--these efforts have proven inadequate for addressing the deeper systemic issues. Johnson refers to this as the "Era of Entanglement," a period marked by the intertwined demands of clinical care, regulation, and finance that have produced a fragmented, burdensome system.[5] As with the leap from horse-drawn carriages to motorized vehicles, transformative, rather than iterative, innovation is needed to meaningfully advance healthcare delivery.

The National Research Council defines this sort of thinking as radical change, where healthcare exploits "qualitatively new ways of conceptualizing and solving health and healthcare problems and revolutionary ways of addressing these problems."[4] Radical change may be more likely when researchers from disciplines across academia are recruited to solve health care delivery and health problems. However, these researchers will need to understand how healthcare works–they need access to data and observations that spur new hypotheses and ideas. The conundrum they face is how challenging it is in our current environment for non-medical faculty to obtain access to patient-care settings due to hospital and privacy policies.

Video as a data collection tool has been under-utilized in healthcare due to inherent barriers, such as confidentiality and privacy.[6–8] Earlier efforts to capture clinical visit knowledge recognized that tacit knowledge, non-verbal interpersonal interactions, and the necessity of understanding 'embodied cognition' all emphasize the need to video-capture interactions between physicians and their patients in a natural setting.[6,9–11] This data collection also enables the documentation of activities in their entirety and complexity.[12] Video capture is not sufficient in health care settings, because of the increasing role of electronic health records, web applications, and other in-room technology.[13] Combining video data with a view of the activities taking place on the computer (EHR logs, clinical notes, and visit-related surveys) as well as a room layout showing the position of the technology with respect to the patient provides a multifaceted view, allowing for qualitative and quantitative analysis and a view supporting multimodal analysis of the encounter. This perspective is especially intriguing given current advances in multimodal AI.[14–20]

There are also concerns that the process of video recording may bias the content of the encounter.[21]
To address these problems and embrace the need for multimodal data, we created the Observer Repository (Repository), an open and FAIR (Findable, Accessible, Interoperable, Reusable) library designed to enable what we term sociotechnical ethnographic research (research involving qualitative and quantitative methods to examine how people interact with technology within a social context). This paper describes those efforts and initial observations.

# Methods

The overarching goal of Observer is to provide a library of clinical exam rooms and the interactions among the patient, any other caregivers, the primary care provider (PCP), and technology.

We began designing Observer by enlisting two steering groups of adult primary care stakeholders. The first group consisted of adult patients and family members, while the second group consisted of physicians, advanced practice nurses, registered nurses, and an ethicist. These groups provided seminal suggestions on the types of data to collect, including adding a visit satisfaction survey from both the patient and the PCP and adding a camera worn by the patient to record the visit from their point of view (egocentric). Both groups also provided suggestions about our recruiting processes.

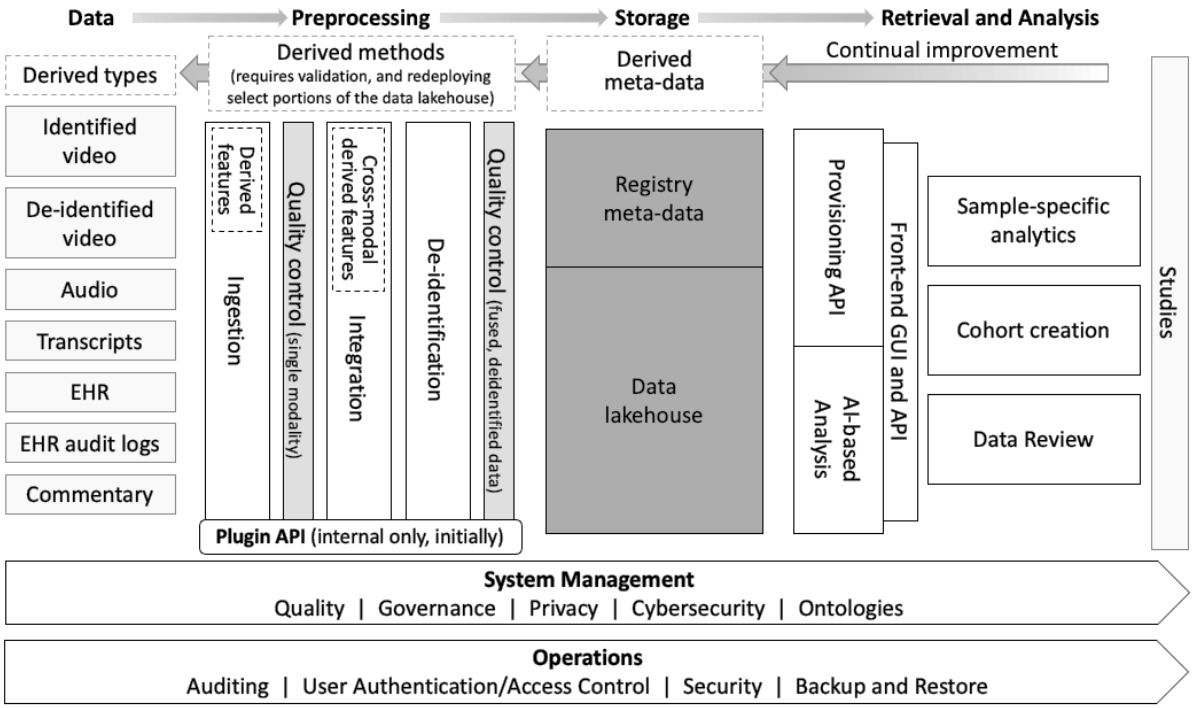



*Figure 1 Design Vision for the Observer Repository. From left to right, raw data and any derived data types go through a preprocessing pipeline, resulting in both meta data and stored data in various states of identifiability. Meta data about Observer may be accessed using a front-end interface or API endpoint to create a cohort that can be sent to the requestor for their studies or downloaded directly using the front-end tools. After retrieval and analysis, and to support continual improvement (from right to left) data derived from some projects deemed useful for subsequent queries can be deposited into the meta-data registry. The same may be true for improvements in preprocessing methods or in the data derived from analyses that are deemed useful to add as a new data type (such as audio waveform data.)*

Based on these conversations and a review of similar medicine-related data platforms such as REDCap[22], Databrary[23], the Linguistic Data Consortium[24], we constructed a design vision for Observer, as shown in Figure 1. The library consists ambulatory care visit videos linked to the EHR user action log, encounter summary, and list of current and past diagnoses, problems, medications, allergies, and all visit-related correspondence from 48 hours before until 48 hours after the recorded visit. Each of these videos is accompanied by transcripts and final progress notes.

## Equipment Selection

We tested numerous cameras, microphone setups, and screen capture software tools under simulated conditions before selecting our current equipment for collecting encounter information. After tests and discussions with health system security, we ruled out using screen-capture software due to the difficulty in aligning those data with room video, challenges with some software capturing video from more than one application (such as from both the EHR and a web browser) and other equipment challenges, (e.g., bringing in a laptop with enterprise licenses for the PCP to use during the visit).

After testing various configurations, we selected the following:

- Room-view camera with microphone array and 360-degree room view, connected to a local zoom session, monitored by a research assistant outside the room.
- Camera worn on the head at eye level, providing both a microphone and a video from the wearer's perspective (egocentric view) offered to both the PCP and the patient. These egocentric cameras were also able to be monitored from outside the room.
- For telemedicine encounters, we selected Zoom, which supported secure downloads of video, and with which clinical teams and patients were familiar.
- To standardize and facilitate remote consent for patients, we initially used an established secure messaging platform.

More technical details may be found at https://observer.med.upenn.edu/.

## Recruitment Process

We focused on patients and PCPs in adult primary care clinics throughout the University of Pennsylvania Health System, an integrated delivery system. We adopted the suggestion of our multistakeholder steering group and required that PCPs be enrolled first, followed by a convenience sample of their patients. For the introductory sessions with PCPs, we created a slide deck summarizing the study goals and procedures. After answering questions about the study, we provided informed consent forms (ICFs) to each PCP. The ICFs also included options for the degree of de-identification each PCP required for broad consent. A copy of the consent form is available at https://observer.med.upenn.edu/.

After PCPs signed ICFs and gave approval to contact their selected patients, the study team contacted their patients scheduled for visits within the PCP's date and time preferences. When patients agreed to participate, a member of our research team would send them the patient-focused ICF for signature or meet the patient with the ICF before beginning the visit. As offered to providers, patients also had the option to blur their face and/or obscure their voice with an additional option to blur any body parts they specified. Both PCPs and patients were provided with a $50 gift card after their completion of a visit and accompanying post-visit annotations. The University of Pennsylvania IRB approved this process.

Before a visit between the consented PCP and patient, a research coordinator set up the in-room camera, then helped patients place the egocentric camera on their head. If any other people accompanied the patient, they were asked to complete a photo release form before the data collection could proceed. Upon arrival of the PCP, the research coordinator helped the PCP with

the egocentric camera and left the room. After the visit, the research coordinator collected all recording equipment and asked the patient to complete a short survey about the visit. Once the surveys were completed, the research coordinator provided the patient and PCP with a gift card (or, after completing 10 visits, the PCP was offered a subscription to ambient scribing software.)

## Protocol, Data Collection Modifications, and Repository Creation

After our first 10 recorded clinical encounters, we reviewed our recruitment, enrollment, and data collection processes. We made numerous small changes in our protocols, including:

- Less involvement from the PCP in deciding which patients should not be contacted (because of concern for bias.)
- Replacement of the camera/zoom session configuration with a standalone room video camera, tripod, and data stored on an encrypted SD card. This step improved set-up time and overall video quality.
- Mitigating overheating in the room video camera by disabling non-essential features and using an external battery pack for longer sessions.
- Ensuring reliable egocentric recording by verifying that both the PCP and patient cameras are actively recording before leaving the room when activated simultaneously.
- Preventing inadvertent changes in video orientation by standardizing the record button activation method.
- Options for PCPs to receive either a gift card or a subscription to an ambient-scribing technology.
- Changing from asking the patient and PCP to annotate the visit's processes and discussion to asking the patient and PCP to complete a survey summarizing their satisfaction with the visit.
- Abandoning the secure messaging strategy to talk with patients about the study, after finding no improvement in patient contact or in study-related communication speed.

Our final protocol is summarized in the Supplemental Figure 1. Of note, the use of wearable cameras was favorably received by all participants; therefore, we continued with that practice.

Observer and all recruitment materials are available at https://observer.med.upenn.edu/.

## Privacy Preservation

To preserve privacy in the Observer Repository, we developed and implemented MedVidDeID[25], a modular pipeline designed to de-identify clinical audio-video (AV) data in compliance with HIPAA Safe Harbor standards.[26] This system integrates state-of-the-art, open-source tools into a six-stage workflow that treats audio and video modalities independently. First, we extracted transcripts using WhisperX and de-identified them with a customized version of PHIlter. Audio de-identification was performed by redacting timestamped PHI segments and transforming voice characteristics to obscure speaker identity. Concurrently, video frames were processed with YOLOv11 to detect and blur facial features using pose estimation and Kalman filtering. The de-

identified audio and video were then recombined and subjected to manual quality control to ensure completeness of PHI removal. An analysis of our first 11 hours of clinical encounters demonstrated a 91% video de-identification rate and reduced need to conduct manual processing time by over 60%. This framework enables scalable, secure sharing and analysis of AV clinical data while preserving patient privacy.

# Results

## Participant Recruitment

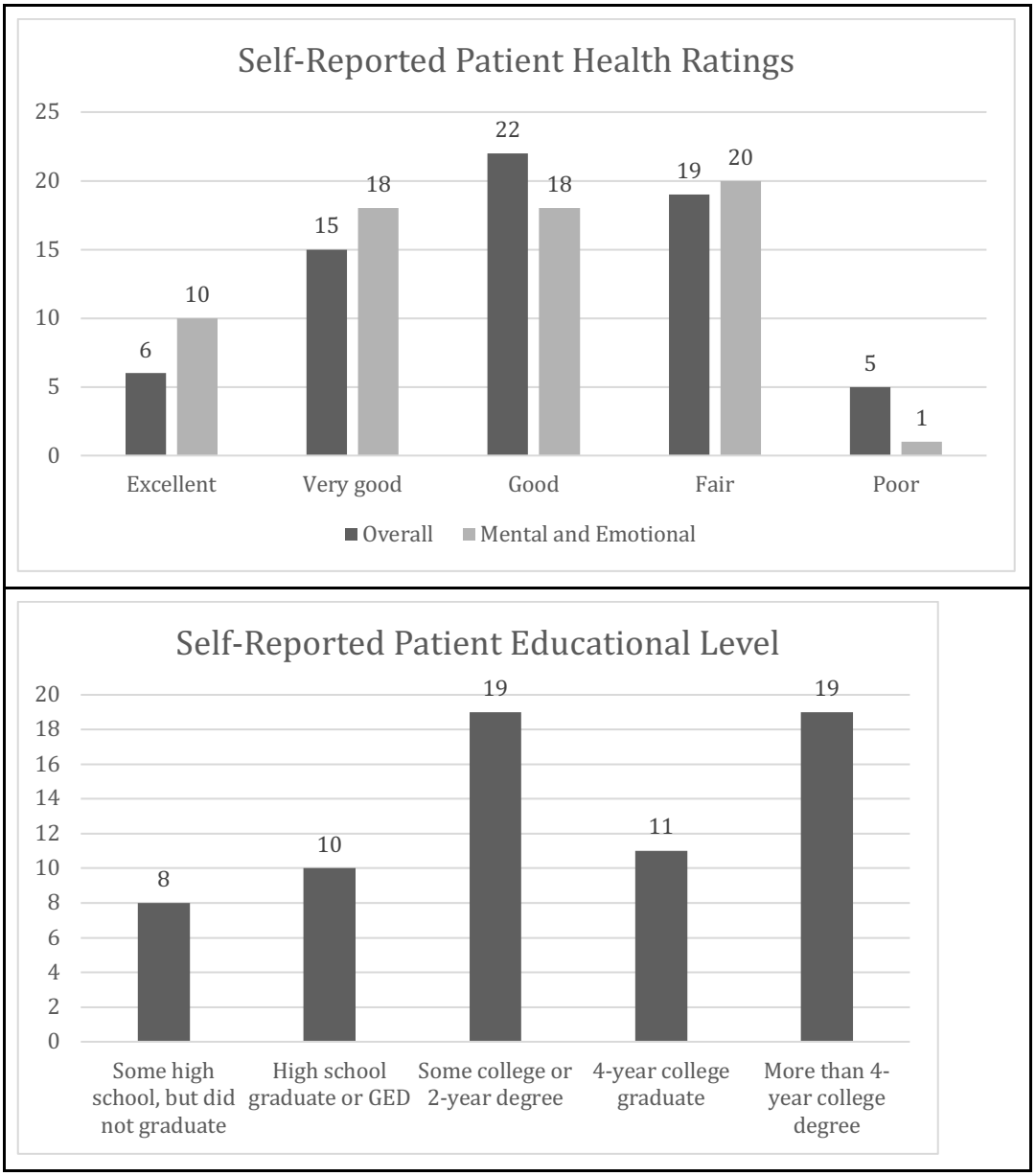
We recruited 13 PCPs from 4 clinic settings. We approached 210 patients scheduled to see these PCPs to participate in this study. A total of 129 patients agreed to participate (60%), from which we were able to capture a clinic visit for 100 (77.5%.) There were 29 patients who verbally agreed to participate but did not complete their appointment as scheduled due to illness, rescheduling, or logistical issues. Three patients participated more than once for separate visits. We contacted three patients whose primary language was non-English, of which two participated in the study (data not shown.) Figure 2 summarizes enrolled patient self-reported health and educational levels.

Table 1 summarizes the results of our first 100 visits. Visit satisfaction (via the post visit surveys) averaged 3.19 out of a score of 5 for patients and 2.63 out of 5 for PCPs, with a generally similar level of satisfaction for each visit. Visit length ranged from 5 - 100 minutes. Because of a change in our after-visit survey, the first 30 patients had no data other than overall satisfaction with the visit.

*Table 1 Visit characteristics.*

| Visit type | Count | Visit type | Count |
|---|---|---|---|
| Follow up | 34 | Check up | 6 |
| Annual Visit | 20 | Chronic disease management | 4 |
| Wellness visit | 16 | Physical | 4 |
| Return visit | 7 | Pre-op exam | 2 |
| New patient | 6 | Procedure | 1 |

*Figure 2 Patient characteristics. Data only available for the last 70 visits, due to the addition of these questions on the post-visit surveys.*




Reasons for visits included the range of visit types for primary care, as shown in the Table 2 below.

*Table 2 Primary care provider (PCP) characteristics*

| PCP # | Total Visits | Mean PCP satisfaction with visit (1-5, SD) | Mean patient satisfaction (1-5, SD) | Years in Practice | Technical competence | # visits impacted by recording processes |
|---|---|---|---|---|---|---|
| 1 | 12 | 3.9 (0.51) | 5(0) | 11-15 | Initial help | 0 |
| 2 | 12 | 3.8 (0.39) | 5(0) | >20 | Self-taught | 0 |
| 3 | 8 | 4.0 (0) | 4(0) | 11-15 | Self-taught | 0 |
| 4 | 21 | 4.3 (0.56) | 4.47 (0.72) | 16-20 | Initial help | 1 (positively) |
| 5 | 16 | 3.3 (1.65) | 4.56 (0.51) | 1-5 | Teacher | 0 |
| 6 | 14 | 4.0 (0) | 4.29 (0.73) | 6-10 | Self-taught | 0 |
| 7 | 6 | 4.2 (0.75) | 4 (0.89) | 6-10 | Self-taught | 4 (negatively) |
| 8 | 3 | 3.7 (0.58) | 3.67 (0.58) | >20 | Initial help | 0 |
| 9 | 4 | 4.0 (0) | 4.2 (0.84) | 1-5 | Self-taught | 0 |
| 10 | 1 | 4.0 (0) | 5 (0) | 16-20 | Self-taught | 0 |
| 11 | 1 | 5.0 (0) | 4.5 (0.71) | 6-10 | Self-taught | 0 |
| 12 | 1 | 5.0 (0) | 4 (0) | 11-15 | Initial help | 0 |
| 13 | 1 | 5.0 (0) | 4 (0) | >20 | Initial help | 0 |

Table 2 summarizes PCP characteristics. PCPs were uniformly distributed across years in practice, ranging from 1-5 years to more than 20 years in practice. Provider comfort with learning to use new technology varied, with most providers being self-taught or teaching others about technology (8/13) and the rest needing instructions or assistance.

Most PCPs reported no adverse effects from having the visit recorded. Of those who perceived some impact, one was positive, where the PCP performing a physical exam reminded a medical student that it was always important to examine the patient without a shirt on. One other PCP reported negative impact without specifying details.

## Clinic Encounter Visit Details

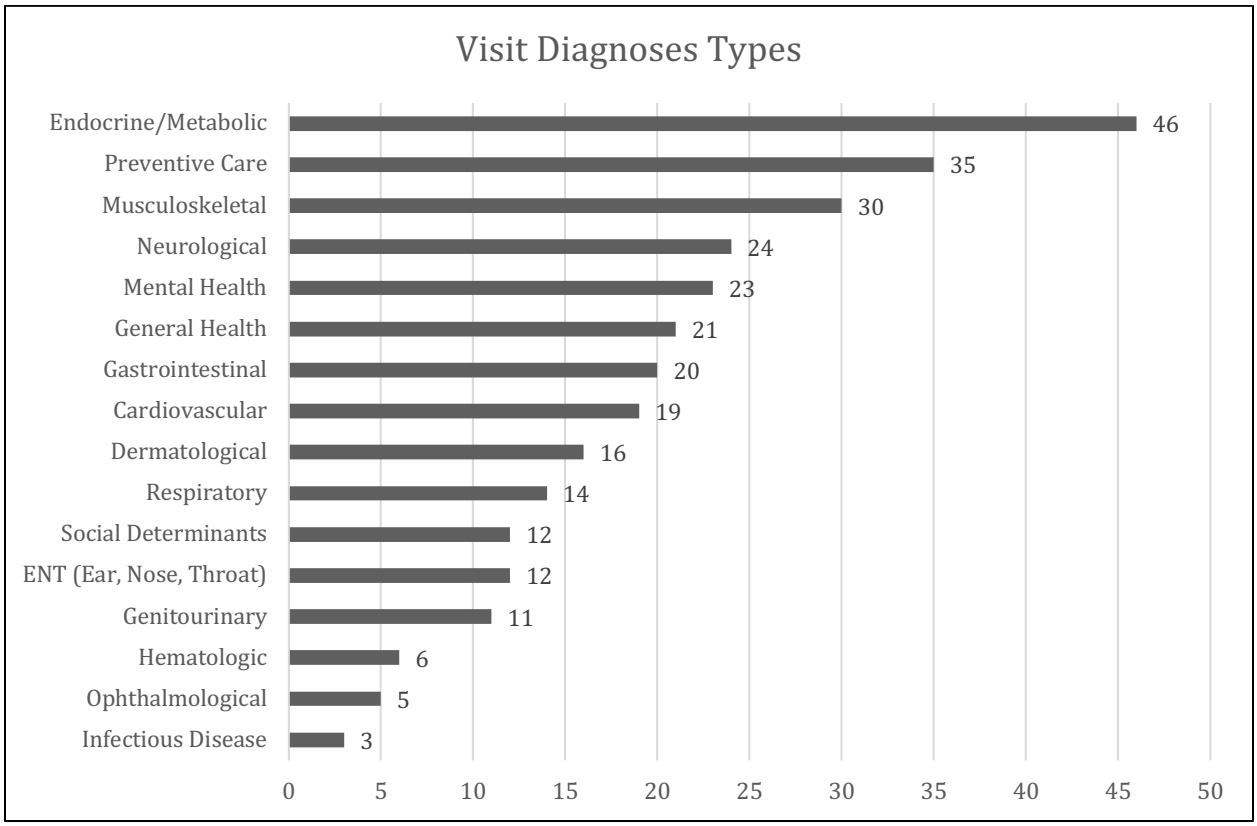



*Figure 3 Summary of visit diagnoses*

Figure 3 summarizes the visit diagnoses covered in our first 100 recordings. Patients had multiple active diagnoses with a frequency representative of a typical adult primary care setting. Figure 4 provides a sample of the different room configurations. We observed that the physical configuration of exam rooms and the EHR screens affects the ability of PCPs and patients to see each other face-to-face and to view the EHR. This reality affects patient engagement, ease of sharing of information, and even the ability of the physician to observe patients' posture, facial responses and other non-verbal actions.

*Figure 4 Sample room configurations of participating clinical sites*

## Discussion and Implications

We report our experience creating the Observer Repository–a first-of-its-kind library of visit recordings, EHR data, and many other forms of accompanying information. We anticipate that use of Observer will enable hypothesis generation and testing, as well as an evolving dataset useful for model development and testing. Ethnographic researchers will be able to use Observer to understand compliance with healthcare delivery recommendations, observe primary care provider-patient interaction, and identify opportunities to improve cognitive support for both patients and PCPs.

In addition to hypothesis generation and testing,[3] these recordings allow a complete record of consultations to be viewed repeatedly for inter-observer reliability testing and open science.[27] In addition, because of the multimodal nature of the data, we anticipate that Observer will support data analysis or model development for a variety of tasks—many of which we hope will be novel and impactful ways to transform information collection during the visit. Perhaps an even more immediate outcome of Observer is to improve medical education by providing a window into care that informs personalized curriculum development and training. The current model contains self-reported training examples that could be viewed by learners (such as when an PCP delivers

bad news to a patient.) Future researchers could add to Observer with items such as expert critiques or questions for learners.[6,8,9]

Another benefit of Observer is its ability to help catalyze multimodal AI applications in a variety of contexts, such as helping clinicians stay up-to-date on latest disease protocols, clinical decision support rules, providing real-time visit information, an almost infinite array of clinical care improvements, physician training, use and design of EHRs, the role of AI, patient facing chatbots, exam room structures, integration with other clinicians and services, including support staff and patients' families, social workers, etc. Multimodal AI, powered by collections such as these, addresses both EHR inefficiencies and benefits, and can improve cognitive support and care quality. As enumerated above, feedback from our participants proposes several actionable innovations worth exploring using informatics and data science methods.

Although we envision many uses for the Observer Repository, we recognize that any data resource of this type suffers from biases that are either measurable or implied. For example, we have observed no systematic skew in recruiting physicians or patients; however, notice of the use of video recordings could potentially introduce a selection bias in the participant population, as those who are camera-shy or have privacy concerns may choose not to participate. This necessitates vigilant monitoring and possibly, the implementation of measures to ensure a diverse and representative sample. The ethical implications of video recording and the potential risks to participant confidentiality and privacy must be meticulously addressed to maintain the integrity of the research and the trust of the participants. Balancing the benefits of rich, detailed data collection with the ethical considerations and participants' comfort is crucial in leveraging representative video recordings for research purposes.[28,29]

The request for video recordings may have raised concerns among potential participants. Some individuals may have been reluctant to participate. because of the possibility that a) the video recordings may be viewed by multiple individuals over time, b) unauthorized individuals may access video data due to improper storage, and c) an individual's identity could be more easily discerned from a video recording compared to written data. Even with our focus on privacy preservation, and even after including language guaranteeing their right to withdraw from any segment of the study whenever they choose, some patients raised concerns that ultimately affected their willingness to participate. We mitigated these concerns by providing gift cards and using a phone call-based contact method, which we found more successful for this population than a method involving secure text messages, perhaps by instilling trust. These observations may help others interested in collecting encounter video data.

Finally, it is important to emphasize that to our knowledge, the Observer Repository is a first of its kind data resource containing a completely de-identified view of ambulatory care visits—with multiple camera views, a visit transcript, an EHR audit log covering from 24 hours before the visit until 72 hours after the visit, and all pertinent EHR data used or accessible during the visit. Developing this Repository required significant attention to patient and PCP privacy using state-of-the art techniques. Improving its utility will require more sites contributing to the resource, as well as continual improvement in our de-identification pipeline, and possibly the development of a federated, rather than centralized data curation model.

## Data Availability

The Observer Repository is available at https://observer.med.upenn.edu

## Code Availability

Code related to the MedVidDeID pipeline is available on GitHub, at https://github.com/kbjohnson-penn/MedVidDeID


## Acknowledgements

We appreciate the thoughtful review of this manuscript by Dr. S.Trent Rosenbloom and Dr. John Holmes, as well as the proofreading of this manuscript by Lauren Malloy.


## Author Contributions

All authors contributed to the conception and design of this project. Mr. Mopidevi provided the technical work leading to this manuscript and contributed to the writing of the Methods section. Mr. Alasaly and Dr. Johnson analyzed the data reported in this manuscript. Drs. Koppel, Jang, and Eaton contributed to the interpretation of the data reported in this manuscript. All authors approve the submitted version of this manuscript.

## Competing Interests

None

# References


1. Blumenthal D, Shah, A, Gunja, MZ, Williams, II RD. *Mirror, Mirror 2024: A Portrait of the Failing U.S. Health System*. Accessed April 29, 2025. https://www.commonwealthfund.org/publications/fund-reports/2024/sep/mirror-mirror-2024

2. Stead WW. Rethinking electronic health records to better achieve quality and safety goals. *Annu Rev Med*. 2007;58:35-47. doi:10.1146/annurev.med.58.061705.144942

3. Johnson KB, Stead WW. Making Electronic Health Records Both SAFER and SMARTER. *JAMA*. 2022;328(6):523-524. doi:10.1001/jama.2022.12243

4. National Research Council (US) Committee on Engaging the Computer Science Research Community in Health Care Informatics. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*. (Stead WW, Lin HS, eds.). National Academies Press (US); 2009. Accessed June 13, 2022. http://www.ncbi.nlm.nih.gov/books/NBK20640/

5. Johnson KB, Neuss MJ, Detmer DE. Electronic health records and clinician burnout: A story of three eras. *J Am Med Inform Assoc*. 2021;28(5):967-973. doi:10.1093/jamia/ocaa274

6. Asan O, Montague E. Using video-based observation research methods in primary carehealth encounters to evaluate complex interactions. *Journal of Innovation in Health Informatics*. 2014;21(4):161-170. doi:10.14236/jhi.v21i4.72

7. Golembiewski EH, Espinoza Suarez NR, Maraboto Escarria AP, et al. Video-based observation research: A systematic review of studies in outpatient health care settings. *Patient Educ Couns*. 2023;106:42-67. doi:10.1016/j.pec.2022.09.017

8. Gilmore RO, Adolph KE, Millman DS, Gordon A. Transforming Education Research Through Open Video Data Sharing. *Adv Eng Educ*. 2016;5(2):http://advances.asee.org/wp-content/uploads/vol05/issue02/Papers/AEE-18-Gilmore.pdf.

9. Chan S. Using videos and multimodal discourse analysis to study how students learn a trade. *International Journal of Training Research*. 2013;11(1):69-78. doi:10.5172/ijtr.2013.11.1.69

10. Coleman T. Using video-recorded consultations for research in primary care: advantages and limitations. *Fam Pract*. 2000;17(5):422-427. doi:10.1093/fampra/17.5.422

11. Nassauer A, Legewie NM. Video Data Analysis: A Methodological Frame for a Novel Research Trend. *Sociological Methods & Research*. 2021;50(1):135-174. doi:10.1177/0049124118769093

12. Schaeffer JH. Videotape: New Techniques of Observation and Analysis in Anthropology. In: Hockings P, ed. *Principles of Visual Anthropology*. DE GRUYTER; 1995:255-284. doi:10.1515/9783110290691.255

13. Committee on Implementing High-Quality Primary Care, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. *Implementing High-Quality Primary Care: Rebuilding the Foundation of Health Care*. (McCauley L, Phillips RL, Meisnere M, Robinson SK, eds.). National Academies Press; 2021:25983. doi:10.17226/25983

14. Bokade R, Navato A, Ouyang R, et al. A cross-disciplinary comparison of multimodal data fusion approaches and applications: Accelerating learning through trans-disciplinary information sharing. *Expert Systems with Applications*. 2021;165:113885. doi:10.1016/j.eswa.2020.113885

15. Bundy H, Gerhart J, Baek S, et al. Can the Administrative Loads of Physicians be Alleviated by AI-Facilitated Clinical Documentation? *J Gen Intern Med*. Published online June 27, 2024. doi:10.1007/s11606-024-08870-z

16. Reed M, Rampono B, Turner W, et al. A multicentre validation study of a smartphone application to screen hand arthritis. *BMC Musculoskelet Disord*. 2022;23(1):433. doi:10.1186/s12891-022-05376-9

17. Silverman D. *Interpreting Qualitative Data*. Seventh. Sage Publishing; 2024.

18. Classen DC, Longhurst C, Thomas EJ. Bending the patient safety curve: how much can AI help? *NPJ Digit Med*. 2023;6(1):2. doi:10.1038/s41746-022-00731-5

19. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. 2022;28(9):1773-1784. doi:10.1038/s41591-022-01981-2

20. Xue C, Kowshik SS, Lteif D, et al. AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat Med*. 2024;30(10):2977-2989. doi:10.1038/s41591-024-03118-z

21. Pringle M, Stewart-Evans C. Does awareness of being video recorded affect doctors' consultation behaviour? *Br J Gen Pract*. 1990;40(340):455-458.

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. doi:10.1016/j.jbi.2008.08.010

23. Gilmore RO, Adolph KE, Millman DS. Curating identifiable data for sharing: The databrary project. In: *2016 New York Scientific Data Summit (NYSDS)*. IEEE; 2016:1-6. doi:10.1109/NYSDS.2016.7747817

24. Liberman, M, Cieri C. *The Evolution of the Linguistic Data Consortium*. Accessed January 22, 2024. https://scholar.google.com/scholar_url?url=https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/lrec1998-creation-distribution-use.pdf&hl=en&sa=X&ei=8ISuZf2RKcrJy9YPkL-OiAY&scisig=AFWwaeYm4hUhOlQ1bH2e8Q1mjVoG&oi=scholarr

25. Mopidevi VSSH, Jang KJ, Alasaly B, et al. MedVidDeID: Protecting Privacy in Clinical Encounter Video Recordings. Published online 2025. doi:10.2139/ssrn.5194782

26. *Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*. U.S. Department of Health and Human Services Accessed May 12, 2025. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html

27. Adolph KE, Gilmore RO, Freeman C, Sanderson P, Millman D. Toward Open Behavioral Science. *Psychological Inquiry*. 2012;23(3):244-247. doi:10.1080/1047840X.2012.705133

28. Maldonado-Castellanos I, Barrios LM. Ethical Issues when Using Digital Platforms to Perform Interviews in Qualitative Health Research. *International Journal of Qualitative Methods*. 2023;22:16094069231165949. doi:10.1177/16094069231165949

29. Sartori L, Theodorou A. A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. *Ethics Inf Technol*. 2022;24(1):4. doi:10.1007/s10676-022-09624-3