We describe methods for capturing and analyzing EHR use and clinical workflow of physicians during outpatient encounters and relating activity to physicians’ self-reported workload. We collected temporally-resolved activity data including audio, video, EHR activity, and eye-gaze along with post-visit assessments of workload. These data are then analyzed through a combination of manual content analysis and computational techniques to temporally align streams, providing a range of process measures of EHR usage, clinical workflow, and physician-patient communication. Data was collected from primary care and specialty clinics at the Veterans Administration San Diego Healthcare System and UCSD Health, who use Electronic Health Record (EHR) platforms, CPRS and Epic, respectively. Grouping visit activity by physician, site, specialty, and patient status enables rank-ordering activity factors by their correlation to physicians’ subjective workload as captured by the NASA Task Load Index survey. We developed a coding scheme that enabled us to compare timing studies between CPRS and Epic and extract patient and visit complexity profiles. We identified similar patterns of EHR use and navigation at the 2 sites despite differences in functions, user interfaces and consequent coded representations. Both sites displayed similar proportions of EHR function use and navigation, and distribution of visit length, proportion of time physicians attended to EHRs (gaze), and subjective workload as measured by the task load survey. We found that visit activity was highly variable across individual physicians, and the observed activity metrics ranged widely as correlates to subjective workload. We discuss implications of our study for methodology, clinical workflow and EHR redesign.
integrated data describing physician EHR activity from multiple perspectives and at many levels of abstraction [23].

1.1. Measurement of clinical and EHR workflow

A variety of techniques have been used to collect data assessing activity patterns. Time-motion studies measuring start and stop times for (possibly overlapping) tasks provide detailed records of physician actions and interruptions. Time-motion studies have been used to address patterns of clinical work and the impact of information systems [7,8,24–26]. Concerns over variations in time-motion data collection methodologies [27] have led to the development of both software-supported protocols such as the work observation method by activity timing [28,24,29], and proposed checklists for ensuring quality of time-motion studies [30]. Observational studies have used ethnographic shadowing [31,32] and video-recording through cameras in examination rooms [33–36].

Many in-situ observational studies published to date suffer from key shortcomings: (1) The granularity of specific EHR actions may be too cumbersome for time-motion studies and difficult for in-room cameras alone to adequately record; (2) Lack of flexibility associated with pre-defined coding schemes, as opposed to post hoc approaches capable of supporting open coding schema and lack of assessment of inter-coder agreement, therefore hindering replicability of findings; (3) Reliance on a single data stream, with no means of temporally aligning different types of activity, such as speech and EHR interactions.

Log files generated by EHRs and related applications provide another means of inferring activity, through analyses as time-stamped entries detailing specific actions. Examples include the analysis of logs captured by handoff tools [37,38]. As log files only record activity with the software in question, they may be most useful when combined with observations or video recordings.

Interview and survey studies assessing perceptions of EHR activity share the strength of being potentially easier to scale to diverse contexts and practices but are subjective and rely on user memory [39]. Surveys collected on a per-visit basis [40] provide greater granularity than more general surveys [33–36].

1.2. Specific aims

Our goal is to develop a methodology for collecting and synchronizing multiple data streams including video, audio, screen capture, and mouse and keyboard activity to provide a quantitative and coordinated picture of physician, patient, and EHR activity during an outpatient clinic visit, and to use these data to understand factors associated with perceived workload. The rich details arising from this multi-modal approach have the potential to provide insight that would not be achievable with more limited data collection approaches.

Specific study aims include: (1) measure EHR workflow, including identifying and rank-ordering EHR functions used, navigation patterns, and time-at-task of order entry, (2) measure clinical work and patient-physician communication including nonverbal (eye-gaze) and verbal communication patterns, (3) measure physicians’ self-reported workload and satisfaction via physician surveys and interviews, and (4) explore associations between the above patterns, visit factors, and physicians’ subjective workload.

Key advantages of our approach are the ability to use events synchronized to the visit timeline to merge across discrete and continuous activity data streams. This aggregation process leads to novel association measures such as the amount of time patients are speaking while physicians are gazing at the EHR. We are also able to augment coding schemas post hoc in order to compare activity between two EHRs (CPRS versus Epic). We illustrate the advantages of our approach using selected analysis components.

2. Materials and methods

2.1. Study design and recruitment

Our data collection approach was developed as part of an observational, prospective study, Quantifying EHR Usability to Improve Clinical Workflow (QUICK), was conducted at outpatient clinics at two large healthcare organizations: the VA San Diego Healthcare System (3 locations) and UCSD Healthcare (2 locations) from 2013 to 2016. Participating primary and specialty care physicians at VA sites used the VA’s Computerized Patient Record System® (CPRS) while UCSD physicians used EpicCare®. Sites differ in key characteristics of patient population, visit length, and EHR software (Table 1). Recruitment and research plans were approved by UC San Diego Human Research Protections Program (HRPP), IRB Project Number: 140215 and the VA San Diego Healthcare System IRB Number: H120065. New and established patients were recruited from each participating physicians’ panels sequentially as they arrived to scheduled visits. Between 2 to 12 unique patients were recruited per physician (median 7, IQR = (6,8)) as per Table 2.

2.2. Data collection

We recorded multiple visits for each physician, typically in two half-day sessions. Recording started before a patient entered the exam room and was stopped after the end of each consented visit. The beginning of each visit was defined as when patient and physician are both in the exam room at the same time. The end of visit was defined as the time when physician leaves the room for the last time. Data containing Protected Health Information (PHI) was secured behind institutional network firewall for subsequent de-identification. Although specific keystrokes were de-identified, room video and EHR screen capture were not. The recording kit consisting of activity tracking devices itemized below, plus a control and storage laptop, fit in a large portable case that requires typically 15 min per day to set up and calibrate. Specific devices and technologies developed for the QUICK study are detailed in [42]. Fig. 1 shows a schematic pipeline shared by visit activity data layers and Fig. 2 shows various activity layers synchronized to an example study visit.

2.2.1. Room audio/video

Exam room activity was captured via a wide field of view web-camera mounted typically above the EHR monitor. Audio and video are recorded using MORAE (techsmith.com) usability software, which enables review and coding of EHR screen and room video along a common timeline via picture-in-picture feature.

2.2.2. EHR activity

EHR activity during visit was captured using MORAE to track UI events, including mouse and keyboard activity (masking passwords) and display video. At full resolution, the EHR display video is sufficiently clear to enable reading text and enables manual coding of EHR activities. MORAE allows annotation of individual mouse clicks by highlighting them in a spreadsheet and also in the replay of EHR display video. Physician’s gaze on the monitor was captured using a slim profile eyetracker (smivision.com) that can resolve ~1 degree angle from physicians’ normal sitting position, approximately one line of text in a typical EHR display. In case of dual-monitor use, we asked physicians to confine their work to the single monitor captured by MORAE.
2.2.3. Epic access logs

Epic access logs were analyzed for UCSD visits, as corresponding logs for CPRS were not available. We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window. Unlike MORAE, access logs show only which Epic function was used, not individual mouse clicks. Logs also reveal the user roles on care team for that encounter (e.g., study physician, LVN, NP, etc.). We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window.

2.2.4. NASA task load Index (TLX) survey

A modified NASA Task Load Index (TLX) survey [43] was used to measure physicians’ workload. The six (20-point Likert scale) items of the TLX measure mental demand, physical demand, temporal demand, performance, effort, and frustration. These were augmented with a question assessing patient interaction: “How satisfied did you feel with the level of interaction with the patient (looking at, talking with, examining) while using the EHR?”. Participants answered these seven questions after each visit, omitting the importance ranking of the subscales [44].

Table 1
Site comparison (UCSD versus VASD) in terms of care delivery model, staff support, and EHR features.

<table>
<thead>
<tr>
<th>Factor</th>
<th>UCSD</th>
<th>VASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient study population</td>
<td>Balanced male/female</td>
<td>Predominantly male</td>
</tr>
<tr>
<td>Scheduled visit lengths</td>
<td>20/40 min visits (Follow up/New patient visit)</td>
<td>30/60 min visits (Follow up/New patient visit)</td>
</tr>
<tr>
<td>EHR</td>
<td>EpicCare Ambulatory</td>
<td>CPRS (Computerized Patient Record System using VistA back-end)</td>
</tr>
<tr>
<td>EHR features, and configuration</td>
<td>- Typically single monitor, but 9 doctors use the dual window</td>
<td>- Dual monitor present in ~35% of visits</td>
</tr>
<tr>
<td></td>
<td>- More levels of menus, objects and paths</td>
<td>- CPRS functions (Notes, Orders etc.) takes up full screen, blocking other functions (even on dual monitor PCs)</td>
</tr>
<tr>
<td></td>
<td>- Associations (Dx to Rx) (no CPRS counterpart)</td>
<td>- Associations for Consults and Imaging but not Dx</td>
</tr>
<tr>
<td></td>
<td>- Real time Care coordination - Patient instructions filled in → printed out (often Nurse out of room sees change in real-time visit status “scheduling”)</td>
<td>- Real time Care coordination (patient status) not in CPRS but available elsewhere</td>
</tr>
<tr>
<td></td>
<td>- Epic access logs to profile pre/post work</td>
<td>- Computerized clinical reminder work</td>
</tr>
<tr>
<td></td>
<td>- Epic logs to profile patient complexity</td>
<td>- Order imaging has more mouse clicks</td>
</tr>
<tr>
<td></td>
<td>- Voice recognition used only 2 visits)</td>
<td>- No separate history documentation UI – only notes</td>
</tr>
<tr>
<td></td>
<td>- Dual windows allow e.g., working in Notes without blocking other functions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Scheduling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Web links available in Haiku and Canto apps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- History documentation interface is structured</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Recruitment and site and specialty groupings in Quick study. (*) Specialties included gastroenterology, pulmonology, cardiology, rheumatology, nephrology.

<table>
<thead>
<tr>
<th></th>
<th>UCSD</th>
<th>VASD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>8/63 (physicians/patients)</td>
<td>9/64</td>
<td>17/127</td>
</tr>
<tr>
<td>Specialty*</td>
<td>7/53</td>
<td>8/43</td>
<td>16/96</td>
</tr>
<tr>
<td>Total</td>
<td>15/116</td>
<td>17/107</td>
<td>32/223</td>
</tr>
</tbody>
</table>

2.2.3. Epic access logs

Epic access logs were analyzed for UCSD visits, as corresponding logs for CPRS were not available. We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window. Unlike MORAE, access logs show only which Epic function was used, not individual mouse clicks. Logs also reveal the user roles on care team for that encounter (e.g., study physician, LVN, NP, etc.). We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window.

2.2.4. NASA task load Index (TLX) survey

A modified NASA Task Load Index (TLX) survey [43] was used to measure physicians’ workload. The six (20-point Likert scale) items of the TLX measure mental demand, physical demand, temporal demand, performance, effort, and frustration. These were augmented with a question assessing patient interaction: “How satisfied did you feel with the level of interaction with the patient (looking at, talking with, examining) while using the EHR?”. Participants answered these seven questions after each visit, omitting the importance ranking of the subscales [44].

Table 1
Site comparison (UCSD versus VASD) in terms of care delivery model, staff support, and EHR features.

<table>
<thead>
<tr>
<th>Factor</th>
<th>UCSD</th>
<th>VASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient study population</td>
<td>Balanced male/female</td>
<td>Predominantly male</td>
</tr>
<tr>
<td>Scheduled visit lengths</td>
<td>20/40 min visits (Follow up/New patient visit)</td>
<td>30/60 min visits (Follow up/New patient visit)</td>
</tr>
<tr>
<td>EHR</td>
<td>EpicCare Ambulatory</td>
<td>CPRS (Computerized Patient Record System using VistA back-end)</td>
</tr>
<tr>
<td>EHR features, and configuration</td>
<td>- Typically single monitor, but 9 doctors use the dual window</td>
<td>- Dual monitor present in ~35% of visits</td>
</tr>
<tr>
<td></td>
<td>- More levels of menus, objects and paths</td>
<td>- CPRS functions (Notes, Orders etc.) takes up full screen, blocking other functions (even on dual monitor PCs)</td>
</tr>
<tr>
<td></td>
<td>- Associations (Dx to Rx) (no CPRS counterpart)</td>
<td>- Associations for Consults and Imaging but not Dx</td>
</tr>
<tr>
<td></td>
<td>- Real time Care coordination - Patient instructions filled in → printed out (often Nurse out of room sees change in real-time visit status “scheduling”)</td>
<td>- Real time Care coordination (patient status) not in CPRS but available elsewhere</td>
</tr>
<tr>
<td></td>
<td>- Epic access logs to profile pre/post work</td>
<td>- Computerized clinical reminder work</td>
</tr>
<tr>
<td></td>
<td>- Epic logs to profile patient complexity</td>
<td>- Order imaging has more mouse clicks</td>
</tr>
<tr>
<td></td>
<td>- Voice recognition used only 2 visits)</td>
<td>- No separate history documentation UI – only notes</td>
</tr>
<tr>
<td></td>
<td>- Dual windows allow e.g., working in Notes without blocking other functions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Scheduling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Web links available in Haiku and Canto apps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- History documentation interface is structured</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Recruitment and site and specialty groupings in Quick study. (*) Specialties included gastroenterology, pulmonology, cardiology, rheumatology, nephrology.

<table>
<thead>
<tr>
<th></th>
<th>UCSD</th>
<th>VASD</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>8/63 (physicians/patients)</td>
<td>9/64</td>
<td>17/127</td>
</tr>
<tr>
<td>Specialty*</td>
<td>7/53</td>
<td>8/43</td>
<td>16/96</td>
</tr>
<tr>
<td>Total</td>
<td>15/116</td>
<td>17/107</td>
<td>32/223</td>
</tr>
</tbody>
</table>

2.2.3. Epic access logs

Epic access logs were analyzed for UCSD visits, as corresponding logs for CPRS were not available. We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window. Unlike MORAE, access logs show only which Epic function was used, not individual mouse clicks. Logs also reveal the user roles on care team for that encounter (e.g., study physician, LVN, NP, etc.). We extracted logs for a 1-year period up to the study visit date to profile patient complexity in terms of the amount of documentation, size of care team involved, and number of patient encounters during that time window.

2.2.4. NASA task load Index (TLX) survey

A modified NASA Task Load Index (TLX) survey [43] was used to measure physicians’ workload. The six (20-point Likert scale) items of the TLX measure mental demand, physical demand, temporal demand, performance, effort, and frustration. These were augmented with a question assessing patient interaction: “How satisfied did you feel with the level of interaction with the patient (looking at, talking with, examining) while using the EHR?”. Participants answered these seven questions after each visit, omitting the importance ranking of the subscales [44].

![Fig. 1. Visit activity data pipeline showing instruments, data collection, coding (including human and computational) analysis components. At left, the schematic for exam room represents the visit time window. Some data components represent pre and post visit work. Activity signals are generally of two types: (*) point-event streams with a single timestamp per event (e.g., mouse click) and (**) interval-event streams based on 2 timestamps on- and off- spanning a definite time duration.](image-url)
2.2.5. Physician interviews

Semi-structured interviews with physicians addressed perceptions of the EHR systems; use of the EHR in preparation or visits, and during consultations; strategies for managing demands from patients’ and the clinical situation; and physicians’ attitudes, and perceptions of the care delivery model of their healthcare organizations. Interviews were transcribed and are amenable to computer-assisted content analysis. At the time of writing, findings from interviews have not been systematically integrated in the study, but we note some preliminary findings.

2.3. Activity coding

2.3.1. Physicians’ nonverbal behavior

Physicians’ nonverbal behaviors were manually coded on room video review to a set of discrete behaviors including: gaze to patient (or companion), gaze to computer, gaze to paper records, physical exam, temporarily leaving exam room, using phone or pager, interacting with nurse or a second physician. For each behavior, a start point and end point is recorded along the timeline yielding an interval event stream. Behaviors were defined to generate a partition of the time axis of the visit, i.e., at any given time, the physician is engaged in exactly one behavior (in a previous study, concurrent behaviors such as sharing of the EHR display with the patient, were coded as a separate activity layer.) Unlike similar studies of interaction styles [36,45], we did not code patient’s nonverbal behaviors. Since each behavior interval event occupies a definite time span, we are able to sum time conditional on specific behaviors in other data layers (e.g., patient vocalization, questions asked) as well as to quantify duration of individual events, indicating task switching frequency.

2.3.2. EHR mouse clickstream

Usability software like MORAE provides a trace of user interface actions (e.g., timestamps of mouse, keyboard clicks) but is not integrated with EHR functions or details. By manually reviewing the EHR display video, we can temporally align the clickstreams by coding to specific EHR functions and tasks. As CPRS and Epic at the two sites differ in their features and workflow, two coding schemas were developed and later harmonized.

2.3.2.1. CPRS.

For CPRS we used a “flat” coding scheme based on a previous usability study [46]. After filtering out mouse wheel events, each mouse click is tagged to several levels of abstraction. The figure illustrates this process, showing how the mouse clickstream is temporally aligned with other data layers (e.g., eye tracking, audio/video) to provide a rich, time-resolved association study for a short 20 s window near the beginning of the visit, from which we can estimate time-at-task of, e.g., physician’s attention to patient vs EHR conditional on the speaker.
At the top level (1) we coded for the 13 CPRS Tabs (e.g., Notes, Orders, Meds, Labs, Reports etc). Each of these functions (corresponding to VistA packages) completely fills the display area; physicians work in a single Tab at any given time. At a second level (2) we categorized each click into one of a set of 50 codes that were adapted from prior work and further developed in this study. Finally (3), we segmented the clickstream and tagged clicks related to order entry (CPOE) activity. Order types include medications, labs, imaging studies, consults, return to visit and reminders, as well as the specific item being ordered (e.g., specific medication name). This enables us to count items being ordered and measure proportion of clicks and time-at-task for each item ordered.

2.3.2.2. Epic. Epic workflow and features different in layout and naming convention from CPRS, with (seemingly) more functions, nested menus and parallel paths to complete any specific task. This difference required development of an alternate coding scheme using a computer-assisted method. Each mouse click was manually coded to a variable-length “tag-path” descriptor. For example, a 3-click sequence may be tagged as (1) “Order Entry”, (2) “Order Entry - Pref List”, (3) “Order Entry - Pref List - Preference List Browser - Accept”, faithfully reproducing the names and sequence of windows and menu items that users encounter. We then organized the pool of tag-path sequences (containing roughly 10,000 unique tags including specific medications, labs and other accessed menu items) by extracting and rank-ordering the most common tags. We then manually mapped the most frequent 150 tags to the closest CPRS Tab equivalent (e.g., Notes, Orders, Labs, Meds).

2.3.2.3. Order entry. Order entry activity was coded as a separate level of detail and can be compared in a straightforward manner between the CPRS and Epic. Human coders manually tracked and itemized individual orders from beginning to end. The resulting activity trace enables comparison of the number of items, number of mouse clicks per item and time at task for medications, lab tests, imaging, consults, reminders and return to clinic orders.

2.3.2.4. Navigation. Navigation patterns in EHRs are quantified by segmenting the mouse clickstream and detecting changes across major sections, i.e., screens. If the individual EHR functions are considered as the nodes of a graph, the transitions between functions are identified with the links between nodes. We consider both the directed and undirected versions of such transition graphs.

2.3.3. EHR keyboard clickstream

MORAE’s keyboard clickstream is a direct record of physicians’ typing, including notes, and can be expected to contain personally identifiable information. To avoid this complication, we coded keys to a small set of pseudo-characters, distinguishing alphanumeric keys, numeric keys, backspace, masked keys (password fields automatically masked by MORAE) and control keys (e.g., arrow keys). This step was performed programmatically based on a character lookup table.

2.3.4. Verbal communication

Patient-provider verbal communication was based on coding room audio/video recordings.

2.3.4.1. Vocalization. We manually coded audio for talk-silence sequences for physician, patient, and patient companion (when present). From this vocalization stream we developed indicators such as ‘turn taking between participants’ and ‘latencies in vocalization patterns’. This methodology has been used to study different communication patterns by doctors and patients [47–49]. This results in a continuous, interval-event stream.

2.3.4.2. Discourse. Vocalization only reflects when each person spoke, not the content of the conversation. Although coding the full content of patient-physician conversation during visits is out of scope, we highlight the possibility of re-analysis enabled by video capture. We coded each visit for occurrences of patient participation verbalizations, including question-asking, assertive responses, and expressions of concern, which may influence a doctor’s behavior and the content of the visit. Partnering-supported talk includes verbal acts that either encourage patients to express their opinions, ask questions, talk about their feelings, participate in decision-making, or affirm or accommodate the patient’s participation [50–52]. This scheme has been used in a number of studies for multiple clinical contexts [53–56], which consistently find that patients communication improves partnership-building and supportive communication. More active patient participation or less physician controlling behavior has also been associated with better health outcomes [57,58]. Discourse was coded as discrete events that can be counted in a visit.

2.3.5. Sensors to automate activity tracking

In addition to the human-coded activity data layers, we deployed a variety of sensors to help automate physician’s activity and patient-provider communication. We used a SMI eye-tracker to capture where on the EHR display the physician is gazing (or more broadly when they are looking at the computer). Kinect was used to track the physicians’ skeletal motion, including head angle and whether they were sitting in front of the EHR or out of the chair, and a microcone spatial microphone to pinpoint who is speaking (patient or physician). However, the data quality was insufficient to replace human-coded data layers. Therefore we will not present analysis methods or results from these instruments.

2.4. Coding quality

Although 2 or more coders have been used in time-motion studies based on in-situ shadowing and real-time coding, our approach based on replayable media enables multiple human coders to review the same visit post hoc, providing, in addition to open-ended activity coding, replicable means of assessing inter-coder agreement. Fig. 2 shows these activity traces for an example visit. Data coding schema were developed iteratively, after several cycles of review of exploratory data visualization. To assess coding quality, we randomly selected approximately 10% of study visits from each site. A second human coder independently coded this sample.

2.4.1. Agreement for discrete events

For discrete (or point-event) activity like mouse clicks, where the timestamps are generated by MORAE, we measure agreement on a click by click basis, generating a $2 \times 2$ confusion matrix. Although coding was done at multiple levels of detail (e.g., Tab, Task, Order Entry, Detail), we only compare the coarsest - ie Tab-level – to assess quality for CPRS and Epic.

2.4.2. Agreement for continuous events

For continuous (interval-event) activity (e.g., NonVerbal, Vocalization) we measured agreement using 2 methods. The more conservative method compares the total time for each type of event (e.g gaze to EHR) per visit (eg, total time gazing at EHR for NonVerbal, total speaker time for Vocalization). We also used a more nuanced measure of agreement based time-resolved comparison that takes into account overlap between coded events (intervals which are intersected to determine their common overlap, then normalized by the independently estimated visit length.)
3. Analysis

In this section we focus on the framework to organize the derivative measures of the activity data layers, the population nesting, and hierarchical modeling. Finally we itemize the most salient methods for specific data components and derivative measures of activity. The unit of analysis in QUICK study is the visit. In this study, each physician sees a variable number of unique patients. Further, physicians are nested by site (UCSD versus VASD), and by specialty groups (we only distinguish primary care from specialty care, and not specific specialties). In addition, we considered nesting by visit factors, e.g., “new” versus “established” patients. Patients are considered “new” if they are new to the specific specialty (i.e., Cardiology), rather than to the care providing organization.

3.1. Organization of activity data

To organize the multiple activity data layers, we use a conven-
ient framework developed by Neerinxc and colleagues based on
naval military settings to model how operators deal with technol-
ogies like communication and radar systems [59,60]. Key factors
influencing workload on human operators are (1) percentage of
time occupied with the technology, (2) level of task switching,
and (3) complexity of information processing. This model was
validated in the context of seaborne operations and found that “sub-
stantial performance decrease” of human operators is observed
near the “overload” region (all three axes high). However, empiri-
cal research is necessary to determine the boundaries between
critical regions in other specific domains of study [59,60]. In our
clinical context, axis (1) leads to consideration of the propor-
tion of time during which physicians are engaged in the EHR. Axis (2),
multitasking, can be quantified in several ways, such as proportion
of task switching between patient and EHR as well as across EHR
functions (e.g., navigating screens). Axis (3), complexity of in-
formation processing, is the most challenging to quantify and remains
evasive.

3.2. Synchronization to visit timeline

All layers of activity data are timestamped either to the MORAE
clock or to a sensor device clock (e.g., eyetracker). Two methods
were explored to allow for shifting of times into a synchronized
state. The first made use of manual review based on auditory
and visual cues. The second method used an implementation of a
dynamic time warping algorithm MATCH [61] to compute the off-
set based on audio. These timestamps were then synchronized to
the visit start time and end time, with the portions of the records
outside these markers trimmed off. Synchronizing point-events
(e.g., mouse or keyboard clickstream) is trivial, while interval
events require an additional step: If an event straddles visit-
begin or visit-end boundaries, it is trimmed so that its boundaries
coincide with the visit. While mathematically trivial (involving
only a series of subtractions) these operations require consistent
indexing, time format management. The bulk of the data workflow
was implemented in Mathematica (wolfram.com). Source code is
available by request.

3.3. Association studies

Because of the limited visit sample size per physician (median 7
visits), we use a nonparametric method to correlate visit activity
metrics to TLX responses based on physician-level aggregation
and comparing correlation distributions [62]. Specifically, we
rank-ordered the median Spearman rho correlation measure (this
only considers whether two observations increase or decrease

3.4. Derived activity measures

Based on the time-resolved activity data, we defined derivative
measures of EHR usage, clinical workflow, and patient communica-

3.4.1. Summary of clinical activity

Summary measures derived from summary statistics integrate
across the visit time of patient-physician communication include
vocalization dominance (ratio of total physician vocalization
duration and total patient vocalization duration) and gaze dominance
(ratio of physician gaze to EHR to gaze to patient) per visit.

3.4.2. Summary EHR activity

From EHR activity we consider (1) the distribution of number
of mouse clicks per specific task (e.g., mouse clicks per MRI order); (2)
the ratio of screen transitions (i.e., across CPRS Tabs) per completed
task (e.g., switching between medication order and labs and back
to medications); (3) the time to completion per task (e.g., time to
complete medication renewal).

3.4.3. EHR navigation

Transitions between EHR screens and functions reflect physi-
cians’ navigation patterns across multiple screens. Although a
number of metrics can be derived based on clickstream sequence,
we will consider only the number of transitions across screens
or major functions (from pilot analysis, typically the majority of con-
secutive clicks occur in the same function with only 10% of mouse
clicks are navigation clicks).

3.4.4. EHR order entry

Computer order entry activity (CPOE) is profiled by 3 metrics: a
tally of the number of items being ordered (CPOE) by type, includ-
ing medications, laboratory tests, imaging studies, consults, return
to clinic and completed reminders, (2) the number of mouse clicks
per order and (3) the time-at-task per order. These are based on
manual clickstream segmentation in MORAE. Order entry clicks
are relatively easy to identify and track in CPRS and Epic, whereas
other clinically relevant tasks (e.g., medication reconciliation), are
more difficult to define as task units due to physicians navigating
across multiple functions, (e.g., medication, lab history, progress
notes).

3.4.5. Time-at-task and time-resolved associations

We exploit the time-resolved activity traces by applying data
processing rules based on temporal algebra [64] to develop more
nuanced measures of associations across data layers. Fig. 2C shows
an short-time window example of such a time-resolved association
study: patient vocalization conditional on physicians’ gaze-to-EHR.
We compute time-at-task by segmenting the clickstream and filter the time windows during which physicians are actually engaged with the computer (this minimizes or eliminates bias due to time accumulated while physicians are engaged with patients). Associations between mixed signal types are handled by suitable preprocessing. For example, EHR time-at-task conditional on physician's gaze to EHR is conducted by first generating a derivative EHR clickstream converting successive pairs of clicks into interval events that can be intersected with nonverbal events.

3.5. Patient clinical complexity

Patient complexity was quantified at both sites by the Charlson Comorbidity Index (CCI) [65] based on the active problem list at the time of the encounter. For UCSD visits, we extracted the problem list (consisting entirely of ICD-9-CT codes) from Epic logs. For VASD, we manually extracted the problem list, which consisted of ~85% ICD-9-CT and ~15% SNOMED-CT codes, from CPRS. ICD-9 Codes were mapped to the 17 Charlson disease groups using a table provided in [65]. SNOMED codes were manually mapped to the Charlson groupings by clinical co-investigators. In addition, on the UCSD side, we also profile patient's complexity based on the size of the patient's 1-year longitudinal health record extracted from Epic logs.

4. Results

Here we limit results to a representative sample of analysis use cases that highlight our ability to merge and compare across activity modes.

4.1. Intercoder agreement

Intercoder agreement for a subset of activity layers is shown in Table 3. We note generally good agreement (median >90%) except for time-resolved vocalization activity. This is due to the short time scales on which conversation turn taking evolves: even small relative time shifts in coding can degrade agreement. This is in contrast to gaze behavior events which are generally of longer duration. As expected, the time-resolved agreement method results in significantly lower intercoder agreement for vocalization than for nonverbal activity. This implies that non-verbal activity is suitable for time-resolved association studies such as timing of EHR activity conditional on physician Gaze-to-EHR, whereas the vocalization signal would be less appropriate for such timing studies (e.g., proportion of time that the patient speaks while physician is gazing at EHR). However, bulk visit measures such as total patient and physician speaking times and vocalization dominance ratio do not require this granular approach. Therefore, we also report for vocalization a second agreement metric based on total time per speaker (total time errors are averaged across speakers and normalized by visit length).

4.2. Patient clinical complexity

The clinical complexity of patients is summarized in terms of Charlson Comorbidity Index (CCI) in Table 4. Most patients have an unadjusted CCI of 0. Seven physicians have study patient panels where all patients have unadjusted CCI of zero. Consequently, rank-order correlations of this complexity profile to visit activity measures are undefined for these physicians.

4.3. TLX subscales

TLX Item response analysis shows that many of the items are highly correlated. We grouped items into 2 subscales based on the magnitudes of the correlations. A parallel approach using principal component analysis reveals the same subscale groupings: the 5-item “Effort” subscale comprises the arithmetic mean of TLX items: Mental, Physical, Temporal, demands, Effort and Frustration. A somewhat less correlated “NegPerformance” subscale is comprised by the mean of 2 items: Performance and Patient Interaction. We flipped the polarity of Performance and Patient Interaction items due to their wording so that all items are consistently interpreted as “worse” as they increase, hence the “Neg” prefix.

4.4. EHR function usage and navigation

We report a comparison of activity patterns in EHRs based on mouse clicks coded to the top-level Tab (or screens). To facilitate comparison, we began with an existing coding schema for CPRS and mapped Epic functions to these existing groupings. We then focus on the most frequently encountered Tabs that the two systems share (Notes, Orders, Labs, meds and Reports) and group all remaining Tabs as “Other” as shown in Table 5. Timing for each Tab is conditional on physician’s Gaze-to-EHR only: no time accrues to the Tab that is currently displayed while the physician is attending to the patient or physical exam. EHR Navigation patterns can be summarized in terms of the number of “Tab” (screen-level) transitions (or 2-grams) during the visit, based on mouse click activity (Table 6). We also rank order the most frequently occurring Tabs in these transitions. Most consecutive mouse clicks occur in the same screen: from a total sample of ~25 k clicks, we observed ~4 k transitions between Tabs: navigation patterns correspond to a higher level of EHR task abstraction than individual mouse clicks. The most frequent directed transitions (observed at least 5%) were: Other → Notes (14%), Orders → Other (13%), Notes → Other (13%), Other → Orders (11%), Notes → Orders (6%), Orders → Notes (6%). The most frequent undirected transitions were between [Notes,Other] (26%), [Orders,Other] (24%), [Notes,Orders] (12%), [Labs,Other] (7%), [Labs,Notes] (6%), [Meds,Notes] (5%). Notes, and to a lesser degree, Orders, were key hubs of activity: physicians navigated frequently between these functions during the visit (Notes = 39% and Orders = 34%).

Table 3
Summary of data coding quality in terms of intercoder agreement across dual-coded visits.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Sample size (visits)</th>
<th>Intercoder agreement</th>
<th>Agreement: (Median, IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR CPRS (Aim 1)</td>
<td>n = 15 (15 VASD)</td>
<td>Sequential Tab-level comparison</td>
<td>0.98 (0.97–1.0)</td>
</tr>
<tr>
<td>EHR Epic (Aim 1)</td>
<td>n = 11 (11 UCSD)</td>
<td>Sequential CPRS-equivalent Tab-level comparison</td>
<td>0.92 (0.69–0.94)</td>
</tr>
<tr>
<td>NonVerbal (Aim 2)</td>
<td>n = 21</td>
<td>Time-resolved comparison</td>
<td>0.94 (0.86–0.95)</td>
</tr>
<tr>
<td>Vocalization (Aim 2)</td>
<td>n = 7</td>
<td>Averaged sum of speaker time comparison</td>
<td>0.64 (0.56–0.7)</td>
</tr>
</tbody>
</table>

A. Calvitti et al. / Journal of Biomedical Informatics 69 (2017) 135–149
Fig. 3 shows navigation sequences for 1 randomly selected visit for each of the study physicians based on this same top-level mouse click activity. Extensive switching across screens and interfaces is typical of most physicians and visits regardless of site or specialty. Tasks that are typically considered as a unit, such as documentation, medication reconciliation, order entry, and (particularly) information retrieval tasks are often distributed throughout the visit, making it difficult in general to segment tasks (e.g., determining the beginning and ending of each task). These patterns make it clear that physicians switch clinical tasks throughout the visit. Some tasks such as order entry can be tracked because of the explicit user interface components accessed, e.g., a medication prescription consists of selecting (typically from pull down menus) the medication, dosage, and various directions (“sig”) components, and signing the order. All of these steps are easily identifiable.

4.5. Time-at-task distribution

The study-wide median visit length was 18.5 min IQR = (13.1, 25.0). Fig. 4 shows the distribution of non-verbal behavior timing by site and specialty grouping, as well as the median distributions of both nonverbal and EHR activity by physician. As for most of the activity profiled in this study, the variance of time-at-task (eg length of visit, gaze-to-EHR) across visits is more highly associated with individual physicians than with other subgroups studied (eg, by site, by specialty, by patient status). For example, the differences in median observed visit length can be explained by an institutional factor: the different scheduled visit length is 20 min and 40 min at UCSD versus 30 min and 60 min at VASD. However, the distribution of visit length across all visits by some physicians tend to be all (or mostly) above or below the median study visit length.

Overall, physicians spend slightly more time looking at the EHR than at patients (median 7.4 min versus 6.6 min, yielding a gaze dominance of 1.1). Physicians at UCSD spend more time looking at the patient than VASD physicians relative to visit length, but less absolute time due to their shorter visits (median 5.7 min versus 7.4 min). Physical exam time was similar between sites and specialties. EHR activity was similarly more variable by physician than by site.

4.6. Correlation of visit activity to TLX

As with most visit activity metrics, TLX response patterns varied greatly by physician, with some individuals clustering toward low values across visits, others clustering toward higher values, and some with larger variance across visits. Because of this variable location and dispersion characteristics responses cannot be compared directly across physicians. Instead, we aggregate visit activity at the level of the physician. Table 7 shows a rank-ordering of the top visit activity measures most correlated (in magnitude) to the TLX subscales based on the (median of) physician-aggregated correlation distribution. We use Spearman rho as the correlation measure as this is only sensitive to rank rather than numerical value of a measure. We arbitrarily truncate this list to absolute rho values greater than 0.2.

### Table 4
Summary of patient and visit complexity extracted from active problem list and CPT codes. CCI, or Charlson Comorbidity Index, is based on active problem list at the time of the visit.

<table>
<thead>
<tr>
<th>Study pop. n = 219 (100%)</th>
<th>By site</th>
<th>By specialty</th>
<th>By patient status</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSD 112 (51%)</td>
<td>VASD 107 (49%)</td>
<td>Primary 127 (58%)</td>
<td>Specialist 92 (42%)</td>
</tr>
<tr>
<td>CCI (raw) 0 (0.1)</td>
<td>0 (0.2)</td>
<td>0 (0.0)</td>
<td>0 (0.3)</td>
</tr>
<tr>
<td>CCI (age adjusted) 2 (1.4)</td>
<td>2 (1.4)</td>
<td>2 (1.4)</td>
<td>2 (1.5)</td>
</tr>
</tbody>
</table>

### Table 5
Comparison of EHR function activity between the two sites based on mouse clicks and timing based on physicians’ gaze-to-EHR.

<table>
<thead>
<tr>
<th>CPRS (VASD) n = 89 (16688 mouse clicks)</th>
<th>Common and frequent tabs in CPRS and epic</th>
<th>Epic (UCSD) n = 106 (8280 mouse clicks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tab-level transitions (count)</td>
<td>Timing (min)</td>
<td>Count</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>578 (58%)</td>
<td>8300 (50%)</td>
<td>Notes</td>
</tr>
<tr>
<td>198 (20%)</td>
<td>4547 (27%)</td>
<td>Orders</td>
</tr>
<tr>
<td>55 (5%)</td>
<td>1094 (7%)</td>
<td>Labs</td>
</tr>
<tr>
<td>43 (4%)</td>
<td>666 (4%)</td>
<td>Meds</td>
</tr>
<tr>
<td>24 (2%)</td>
<td>403 (2%)</td>
<td>Reports</td>
</tr>
<tr>
<td>107 (1%)</td>
<td>1698 (10%)</td>
<td>Other*</td>
</tr>
</tbody>
</table>

### Table 6
Distribution of navigation across EHR functions based on human-coded mouse activity. Tab (or screen)-level transitions based on mouse clicks tagged to the top-level screen or “Tab” coding.

<table>
<thead>
<tr>
<th>Study pop. n = 195 (100%)</th>
<th>By site</th>
<th>By specialty</th>
<th>By status</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSD 106 (54%)</td>
<td>VASD 89 (46%)</td>
<td>Primary 113 (58%)</td>
<td>Specialist 82 (42%)</td>
</tr>
<tr>
<td>New 41 (21%) Established 154</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tab-level transitions (count)</td>
<td>median IQR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 (9.27)</td>
<td>22 (13.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 (8.22)</td>
<td>21 (11.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 (8.22)</td>
<td>15 (10.26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 (9.29)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The accompanying scatterplots in Fig. 5 show the variability of correlations of physician TLX responses to the 2 highest correlated factors to each subscale. Visit length and EHR navigation (Tab-level transitions) are the top correlates to Effort subscale (median rho values 0.49 and 0.42, respectively). The clustering of (19 of the 29) physicians in the upper-right (positive correlation) quadrant of the Effort subscale plot indicates that these visit activity measures can predict the subjective effort of most physicians. On the other hand, subjective performance is not as easy to relate to visit activity, as shown by the wider distribution of physicians in the NegPerformance plot (only 8 of 21 physicians in the positive quadrant), despite moderate correlations of this subscale to the top 2 factors: EHR navigation and patient complexity as measured by Charlson comorbidity (median rho values of 0.42 and 0.26, respectively). Moreover, for both subscales effort and performance, a significant fraction of physicians essentially show the opposite correlation. An extreme example, physician D27 (in the lower-left or negative quadrant of both scatterplots) reported longer visits with high EHR navigation and high comorbidity patients as less effortful as well as higher performance.

4.7. EHR order entry comparison by site

The combination of human-coded, itemized order entry and time-at-time task (based on the time-stamped mouse clickstream) enables comparison of the quantity of orders and time necessary to complete orders at each site, enabling comparison between CPRS and Epic. Due to variable naming conventions, we grouped certain types of orders: for VASD visits, Allergies and Procedures were grouped as Reminders, and Nursing and miscellaneous other orders as Other. VA and Non-VA medications were grouped as Meds. For UCSD visits, Health Maintenance activities were considered as Reminders. Fig. 6 shows the order entry activity (total number of items ordered in visit) and the self-reported TLX Effort score per visit for 2 physicians selected from opposite extremes of the observed correlation spectrum. This shows that no single activity measure is likely to fully capture the variety of physicians’ personalized workloads.

Aggregating by site, Table 8 shows that the same underlying order entry measures can be used to compare the frequency of orders by type as well as to provide baseline for comparing the EHRs user interface burden in terms of mouse clicks and time-at-task per order. Consistent with previous studies of CPRS [46], we see that Consults and Imaging orders take the longest to complete for VASD physicians, followed by medications. We can also see that labs take less time and fewer clicks on average. The statistics for UCSD side using Epic are strikingly similar. One major difference is the preponderance of Reminders activity by VA providers (largely confined to primary care physicians). To summarize, at both sites consults and imaging orders require more mouse clicks and time to complete (per item ordered) than medications or laboratory tests. However, the latter two are ordered more frequently. The difference in the user requirements across order type may be due to the complexity of the information set to complete consults and imaging, rather than user interface inefficiency. Such quantitative profiling can be used to guide EHR redesign by focusing on the most burdensome and frequent tasks.
5. Discussion

Patel and Kannampallil describe cognitive informatics as focusing on “understanding work processes and activities within the context of human cognition and the design of interventional solutions” [66]. Typical dimensions considered by relevant studies include the specific cognitive framework, study type, setting, data collection methods, and participants. Within this framework, the QUICK study can be seen as a naturalistic (study type), clinical (setting) study of comprehension, decision-making, distributed cognition, and usability (cognitive framework), involving patients and physicians (participants). Our multi-modal recordings of more than 200 visits in two large care organizations provides a detailed data set for a larger sample size than previous cognitive informatics studies [66]. Our preliminary analyses demonstrate the power of this approach for enabling detailed analyses of EHR and clinical activity by physicians in outpatient visits. This work also provides both valuable lessons regarding the challenges of collecting and analyzing a multitude of synchronized activity streams at scale, and provides insights on the relation and relative importance of activity patterns to subjective workflow.

Table 7
Physician-aggregated empirical correlation distributions rank ordered by median absolute Spearman rho. Note the top 5 correlation magnitudes are associated with the TLX Effort subscale. Charlson Comorbidity Index is based on the patients’ active problem list at the time of the visit. For CCI, the smaller sample size is due primarily to an entire patient panel’s CCI equal to zero. Thus Spearman correlation to visit activity measures is undefined (these physicians were removed from the pool). Raw CCI is not adjusted for patient’s age.

<table>
<thead>
<tr>
<th>Activity measure</th>
<th>Subjective workload (TLX subscale)</th>
<th>Sample size (physicians)</th>
<th>Physician-aggregated correlation Spearman rho: median, (IQR)</th>
<th>Rank (based on median)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit Length (minutes)</td>
<td>Effort (5-item)</td>
<td>n = 32 (100%)</td>
<td>0.49 (0.12, 0.67)</td>
<td>1</td>
</tr>
<tr>
<td>EHR Tab (screen) transitions (count)</td>
<td>Effort</td>
<td>n = 29 (91%)</td>
<td>0.42 (0.36, 0.63)</td>
<td>2</td>
</tr>
<tr>
<td>EHR Tab (screen) transitions (count)</td>
<td>NegPerformance (2-item)</td>
<td>n = 29 (91%)</td>
<td>0.38 (0.14, 0.61)</td>
<td>3</td>
</tr>
<tr>
<td>Epic Log Size (count)</td>
<td>Effort</td>
<td>n = 16 (50%) UCSD</td>
<td>0.35 (0.24, 0.52)</td>
<td>4</td>
</tr>
<tr>
<td>EHR Mouse Clicks (count)</td>
<td>Effort</td>
<td>n = 32 (100%)</td>
<td>0.32 (0.03, 0.6)</td>
<td>5</td>
</tr>
<tr>
<td>Gaze Dominance (ratio)</td>
<td>Effort</td>
<td>n = 32 (100%)</td>
<td>0.28 (0.07, 0.58)</td>
<td>6</td>
</tr>
<tr>
<td>Charlson Comorbidity Index (raw)</td>
<td>Effort</td>
<td>n = 24 (72%)</td>
<td>0.27 (0.02, 0.54)</td>
<td>7</td>
</tr>
<tr>
<td>Charlson Comorbidity Index (raw)</td>
<td>NegPerformance</td>
<td>n = 25 (78%)</td>
<td>0.26 (0.23, 0.47)</td>
<td>8</td>
</tr>
<tr>
<td>EHR Mouse Path Length (pixels)</td>
<td>Effort</td>
<td>n = 32 (100%)</td>
<td>0.23 (0.06, 0.56)</td>
<td>9</td>
</tr>
<tr>
<td>Verbal Patient Concerns (count)</td>
<td>NegPerformance</td>
<td>n = 25 (78%)</td>
<td>0.21 (0.02, 0.32)</td>
<td>10</td>
</tr>
<tr>
<td>EHR Keystrokes (count)</td>
<td>Effort</td>
<td>n = 31 (97%)</td>
<td>0.20 (0.07, 0.49)</td>
<td>11</td>
</tr>
</tbody>
</table>

Fig. 4. Summary of gaze and EHR activity in visits. A: Non-verbal median time-at-task for the most frequent clinical behaviors (gaze to EHR, gaze to patient, physical exam and gaze to paper, medications or other artifacts), aggregated by site and specialty. Time axis is in minutes. B: Nonverbal and EHR time-at-task by physician. The thicker bottom bar above it shows median EHR time-at-task activity coded to CPRS Tabs (VASD) or Epic CPRS-Tab equivalent functions (UCSD) conditional on physician gaze to EHR.
In this discussion, we focus on the challenges and lessons learned from QUICK study and the advantages and limitations of our approach and methods. Among major challenges are comparison across sites (UCSD and VASD), and accounting for the observed variation and clustering across physicians.

5.1. Tracking activity in clinical settings

Collecting detailed information on EHR usage in clinical settings for large sample sizes has been identified as a critical challenge in cognitive informatics for understanding health IT and its impact on care [66]. Building on prior explorations of techniques including observations [31,32], video recordings [34–36], and time-motion measurements [24,25,28,29], our protocol uses a novel combination of measurement techniques [42], including an array of sensors for eye-tracking, spatial audio, and participant motion sensing (Kinect) to collect multiple, synchronous data streams, providing previously unavailable perspectives on EHR user and physician-patient interactions.

Our approach provides several advantages over previous methods. The combination of complementary data streams provides multiple means to qualitatively and quantitatively interpret...
human and system level behaviors from multiple perspectives. Comparing synchronized records of keyboard interactions, mouse movements, and speech acts can inform understanding of visit dynamics, allowing greater understanding of how EHR content might drive clinical encounters, when changes in speech patterns might indicate greater cognitive load associated with EHR use, and possibly errors or miscues associated with distraction. Unlike time-motion studies based on fixed, pre-existing codes [28], our approach uses replayable media enabling re-analysis at multiple levels of detail, thus aiding replicability of findings and reusability of collected data via additional video and EHR activity review.

5.2. Instrumental limitations

Although advances in both software and sensors enabled the development of new and more powerful approaches to data collection [42], effective application of these tools presents multiple challenges. Our data collection sessions experienced data loss due to difficulties with software malfunctions, ambient noise, and poor calibration. We also experienced limitations in identifying the most meaningful way to represent clinically meaningful units of activity. Work in this area continues as we integrate context based on physician interviews. To minimize these problems, researchers should be prepared to carefully test and debug novel combinations of data collection technologies. Recent and prospective developments may ease some of these hurdles, as usability tracking software integrated with web-based EHR systems, ubiquitous sensors, and increasing use of enterprise data logging will likely help scale up data collection and automate coding to larger sample sizes and diverse clinical settings.

Our data collection is currently limited to a narrow set of clinical environments and workflows (e.g., outpatient settings during well defined point of care encounters). Measuring activity at this foundational level requires technological expertise and collaboration with clinical IT staff. Extensive human coding is required to contextualize MORAE's EHR clickstreams to the actual clinical EHR activities. Some visits were marred by instrumental failures or intermittent dropouts, which further reduced sample size or introduced noise. Eyetracking and mouse path was limited by inability to register to specific EHR windows or moving content. Combining tracking devices such as eyetracking, Kinect head and body tracking, and spatial audio reconstruction may partially automate some of the activity coding, particularly eye-gaze, and vocalization. This was the motivation for integrating and synchronizing these devices in a LAB-IN-A-BOX [42]. Similarly, web-based EHRs with built-in usability analytics may facilitate or automate mouse coding to EHR functions. Further, eyetracker calibration requires that the monitor not be moved, potentially inhibiting screen sharing with patient.

One of the key lessons learned is that sensors that function well in controlled lab settings may not work as well in rooms with different lighting, layouts, sound levels, or other factors beyond the control of the investigator. Basic field research survival skills apply – use a well-defined protocol, test frequently, be sure to have backups, and plan for data loss due to collection failures. Although improvements in sensor technology may address some of these concerns, working in varied and unfamiliar settings is likely to continue to be challenging.

The granularity of activity data provides unprecedented resolution but also presents a challenge. One aspect, temporal synchronization, can be managed at the analysis stage as we've demonstrated in terms of coding quality. A more problematic aspect is understanding the dynamics of interactions between the physician, patient, and EHR requires analysis that ties activity to clinically meaningful goals and tasks. Data captured by general-purpose usability software tools like MORAE provides a finely-grained record of individual keystrokes and button presses, but it does not automatically identify actions in units that reflect physician's intentions. Identification of specific elements of the EHR (e.g., clicking on the notes tab), completion of key tasks (e.g., reminders, order entry), or sequences of actions associated with relevant higher-level goals (e.g., medication reconciliation) currently requires manual review and coding. Future studies might look toward EHRs instrumented to track and log key function usage (whether in clinic or remotely) and composite tasks (e.g., medication reconciliation).

5.3. Site and EHR differences

A leading motivation for QUICK study was to understand how EHR design affects workflow, with a particular focus on differences between the two care contexts (Table 1). Although data was collected in the same way at both sites, different interfaces and features of CPRS and Epic complicated comparison. Coding challenges included lengthy clickstreams generated that are not keyed to specific EHR functions. This required development and comparison of two coding schema for CPRS and Epic. We also encountered differences in practice in terms of the room layouts, and support staff that perform EHR tasks before or after visits. For example we observed activity in Epic on functions like associations of medication orders to diagnosis that have no counterpart in CPRS. Also, Epic has more automation support for restructuring data like patient history across multiple interfaces and importing in notes, whereas CPRS has a more basic notes and template interface. Epic also enables working on one function, like notes, without blocking other activity, making it challenging to compare navigation patterns.

5.4. Implications

5.4.1. Methodology

We highlight a novel aspect of our methodology for quantitative profiling clinical activity: merging across activity layers enables quantifying aspects of activity which can differ markedly from more basic proxy measures. For example when we merged physician gaze and mouse activity to measure time spent on specific EHR functions, we found that at one site (UCSD) 41% of time was spent on Notes (reflecting documentation or reading), whereas only 21% of that time was found to be spent on Notes when based on mouse activity alone (Table 5). A similar difference is exhibited at VASD. We advance the notion that timing measurements are applicable to a wider variety of user interaction with computers than mouse click counts.

Study design could be improved by closing gaps in data collection, particularly visit factors such as visit timing (i.e., whether a late start to a visit added to temporal demand), and patient factors not quantified by Charlson comorbidity, such as mental health and behavioral issues. The comorbidity index, derived from clinical codes (ICD, SNOMED), further lacks context such as how long the condition has been extant, and degree of severity. We suggest that these facets can be extracted from longitudinal EHR logs and propose that the informatics community focus on the longitudinal aspect of patient complexity profiling that logs enable. Similarly, longitudinal data captured before or after the visit might be useful for understanding work that physicians or staff do before or after the visit.

5.4.2. Clinical workflow

We have quantified variation in EHR and clinical work between major study subgroups, including by site, specialty, and new versus established patients, as well as grouped by physician. Variations between physicians, even within the same institution, seemed to
be greater than trends within institutions. This suggest that individuals’ personalized work practices and preferences may be more important than the institutional setting, EHR interface or even patient characteristics in shaping observed patterns, such as visit length, proportion of time attending to EHR versus patient, patient-provider communication and pre-post visit work. Care team task delegation (which also depends on the institution, setting, or even individual provider team) may also account for some variation in activity and self-reported workload, effort and performance. Although physician interview analysis is beyond the scope of this paper, we expect physician comments will provide context to explain some quantitative findings. For example, a preliminary reading suggests that staff support levels are greater at UCSD, where nursing staff carry out more EHR tasks such as order entry. This difference in support availability may potentially explain, beyond the difference in scheduled visit lengths, why UCSD visits tend to be several minutes shorter on average than at VASD.

5.4.3. EHR redesign

Our approach and findings based on studies at VA San Diego clinics have provided baseline evidence of CPRS use and actionable guidance to ongoing VHA national initiatives to modernize its user interface [67–72]. Quantitative profile of the proportion of time physicians spend on EHR as well as identification of key functions used and navigation and order entry burden by order type led to proposed design improvements. For example, documentation and order entry are the most used functions, and most physicians frequently navigate between these multiple times during a visit. This finding motivated development of a VA prototype technology, “Active Notes”, using a command language syntax combining semi-structured language understanding integrated with a “computational” notes interface. In formative usability studies of the prototype versus CPRS demonstrated that combining order entry and subsequent documentation into a single step reduced overall time-at-task for medication order entry [72]. Optimizing other types of EHR tasks is complicated by a greater available combinations of user activity sequences or pathways across difference EHR functions. For example, determining clinical reminders is straightforward, whereas linking all tasks related to medication reconciliation requires resolving activity in functions as disparate as order-entry, medication look-up, and notes (e.g., to retrieve fragmented patient data).

6. Conclusion

We demonstrated a novel approach to collecting and analyzing multiple sources of data during clinical activities and integrated these streams into meaningful measures, enabling comparison across two clinical settings with different EHRs and a spectrum of primary and specialty (outpatient) care. This effort revealed a high degree of variation in observed activity and clinical practice despite accounting for similar types of visits and patient complexity. We identified similar patterns of EHR use and navigation at the 2 sites despite differences in functions, user interface and consequent coded representation. For example, both sites displayed remarkably similar proportions of EHR functions use and navigation, as well as distribution of visit length, proportion of time physicians attended to EHRs (gaze), and subjective work-load as measured by NASA Task Load survey. Commonly noted high-level clinical tasks, such as medication reconciliation or preventative care were highly distributed across the visits and very difficult to measure, suggesting the need for further levels of integration. Preliminary workload analyses suggested a complex relationship between levels of measurable physicians’ activity during visits and perceptions of effort and task performance. As no single visit activity factor was highly correlated with subjective talk load, a fuller understanding of the work-flow and cognitive flow will require integration of the physician interviews, possibly augmented with additional activity coding based on existing EHR display and room video to extract additional details about patient and visit characteristics.

This study is an early attempt at applying an interdisciplinary approach and multi-modal data collection, that will need refinement in methods, instrumentation and analysis to scale up to larger sample sizes and different clinical settings (e.g., inpatient care). Future applications of this methodology may be valuable to better understand cognitive processes in real-world situations and to help redesign information technology to better blend with clinical workflows.

Conflict of Interest

Authors declare no conflicts of interest related to the material in this manuscript.

Acknowledgments

This work was supported by the Agency for Healthcare Research and Quality R01HS021290. This material is the result of work supported with resources of the VA San Diego Healthcare System. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

References


