Final Progress Report

1. Quantifying Electronic Health Record Usability to Improve Clinical Workflow

Principal Investigator: Zia Agha[a,d,f]

Team Members: Shazia Ashfaq[a], Alan Calvitti[a], Harry Hochheiser[b], Kristin Bell[i], Yunan Chen[e], Robert El Kareh[j], Mark T Gabuzda[i], Lin Liu[h], Sara Mortensen[a], Barbara Gray[a] Braj Pandey[i], Steven Rick[g], Richard L Street, Jr[c], Nadir Weibel[g], Charlene Weir[k]

Organization: Veterans Medical Research Foundation, San Diego

Dates of the Project: 09/01/2012 - 12/31/2016

Federal Project Officer: **Shafa Al-Showk**

Acknowledgment of Agency Support: This study was supported by Agency for Healthcare Research and Quality

Grant Award Number: 4R01HS021290-04 REVISED

2. Structured Abstract

Purpose
This study seeks to quantify electronic health record (EHR) and clinical workflow activity, patient-provider communication patterns and relate the observed patterns to physicians' subjective workload and satisfaction obtained via surveys and interviews to understand how EHR design affects workflow.

Scope
Data for this study is collected at primary and specialty care outpatient clinics of the Veterans Administration (VA) San Diego Healthcare System and University of California San Diego (UCSD) Health, observing primary care physicians and specialists who use two EHR platforms: VistA/CPRS (VA) and Epic (UCSD). We describe activity data relating to EHR use, physicians nonverbal behavior and attention as well as patient-provider communication, self-reported workload based on survey and interviews.

Methods
We collected temporally-resolved activity data including audio, video, EHR activity, and 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 multiple streams, providing a range of process measures of EHR usage, clinical workflow, and physician-patient communication. Grouping activity patterns by physician, we rank-order the factors that account for the observed correlations and provide a multivariate regression model.

Results
Results are provided for each study aim and for the study population as a whole and for the major subgroups: by site (UCSD, VASD), by specialty (primary, specialty care) and patient status (new, established). Results reflect summary statistics for mouse and keyboard activity (EHR), nonverbal and verbal behaviors during visits as well as associations to subjective measures derived from workload surveys.

Keywords: time-motion, clinical workflow, workload, electronic health record.

3. Purpose

The study has 4 aims: (1) To understand and quantify EHR workflow (2) to objectively measure clinical workflow activity and verbal and nonverbal patient-provider communication patterns (3) to solicit physicians' workload and satisfaction for each visit via surveys and interviews (4) to understand associations between the first 3 aims and the factors that may explain observed correlations.

4. Scope

4.1 Background
Current generation Electronic Health Records (EHR) integrate poorly with clinical workflow: physicians spend too much time sifting through fragmented data, they are frustrated by burdensome documentation and the user interfaces have detrimental impact on patient-physician interaction. The Institute of Medicine, among other organizations, has called for improving EHR usability. But there is currently insufficient quantitative evidence that integrate multi-modal activities in real clinical settings with realistic time-pressures and patient case variability.

4.2 Context
Time-motion studies have been used to address patterns in clinical work. Studies measuring start and stop times either manually (ethnographic shadowing) or using software and tablet devices provide detailed records of clinician actions and interruptions. Other studies use video-recording through cameras in examination rooms. In-situ coding suffers from key limitations: granularity, as recording details specific EHR actions may be too cumbersome for time-motion studies and impossible for in-room cameras. Further, the typical focus on a single data stream means no temporal alignment across simultaneous activities (e.g., speech and EHR activity). In contrast, post-hoc coding allows open coding schema and re-analysis and allows assessing inter-coder agreement aiding replicability of findings. Interview and surveys assessing perceptions of work is applicable to diverse practices, but, don't offer temporal granularity of activity and rely on subjective reports.

4.3 Settings
The study was conducted at the VA San Diego Healthcare System at 3 outpatient clinic locations: San Diego VAMC, La Jolla, VA Mission Valley Clinic (primary care) and VA Oceanside Clinic (primary care) and at UCSD Health System at 2 locations: Hillcrest (Hillcrest Cardiology Clinic, Owen Clinic Nephrology) and La Jolla (Internal Medicine Clinic, Sulpizio Cardiovascular Clinic, Arthritis/Rheumatology Clinic, Pulmonology and Sleep Medicine Clinic). We instrumented exam rooms for activity capture.

4.4 Participants
32 physicians and 234 patients (total 266) participants consented. 1 patient and 1 primary care physician dropped out. 10 patients at the VASD and 26 patients at UCSD refused participation. 48 new and 175 established patients participated after excluding major instrumental dropouts. Between 2 to 12 unique patients were recruited per physician (median 7, IQR=(6,8)), as summarized in Table 1.

	UCSD EHR = EpicCare (epic.com)	**VASD** EHR = CPRS	**Total**
Primary	8/63 (physicians/visits)	9/64	17/127
Specialty*	7/53	8/43	16/96
Total	15/116	17/107	32/223

Table 1: Recruitment, site and specialty groupings. (*) Clinical specialties include: gastroenterology, pulmonology, cardiology, rheumatology, nephrology.

4.5 Incidence
N/A

4.6 Prevalence
N/A

5. Methods

5.1 Study Design
QUICK is an observational, hybrid design that combines prospective data collection with retrospective data extraction from EHRs (e.g., Epic access logs). The primary unit of analysis is the visit. Additional prospective (surveys, interviews) and retrospective data (EHR access log on the UCSD site) was collected outside the window of the visit. The main subgroups are: site (UCSD, VASD), physician specialty (primary or specialist), patient status (new, established).

5.2 Data Source/Collection

We recorded multiple visits per physician from 2013 to 2016, typically in two half-day sessions each. Activity recording was started before beginning of visit (when patient and physician are both in the exam room at the same time) and stopped after the end of the visit (when physician leaves the room for the last time). The recording kit: tracking devices and a laptop fit in a portable case[1]. Table 2. summarizes key data collection methods and instruments.

Data Source and (Instrument)	**Description**
Room Audio/Video (Webcam)	Exam room activity was captured via a wide field of view webcam mounted above the EHR monitor. Audio and video are recorded using MORAE (techsmith.com) usability software.
In-visit EHR Activity (Morae)	EHR activity during visit was captured using MORAE to track mouse and keyboard activity (masking passwords) and display video. EHR display video enables reading text and manual coding of EHR activities. MORAE allows annotation of mouse clickstream by highlighting in a spreadsheet and in the replay of EHR display video.
Eyetracking (SMI Red-M + Epiphan Video capture card)	Physician's gaze on the monitor was captured using a remote eyetracker (smivision.com). Combined with a video capture card we record the screen content in tandem with the calculated gaze path and physiologic measurements including pupil diameter
Body tracking (Kinect for Windows)	A Kinect for Windows device, which uses color and depth cameras to detect and track human body motion, was positioned facing the physician in order to record their body motion and gaze direction in the environment.
Spatial Audio (Microcone + Kinect for Windows)	The Kinect device used for body tracking also houses a directional microphone array. This in tandem with a 360 degree direction microphone array (Microcone) allowed for the detection of angle of audio incidence, or calculation of where sound was coming from.
UCSD EHR Access logs (Epic log) Aim 1 - Epic only	Epic access logs for a window of 2 weeks before and after each visit to quantify pre-post visit EHR activity. Access logs show only which Epic function was used and user roles on care team for that encounter. Functions accessed during 1-year period prior to the study visit were extracted to profile patient's complexity.
NASA Task Load Index (TLX) Survey	Modified NASA Task Load Index (TLX) survey was used to measure physicians' workload. The six (20-point Likert) scales of the original TLX (mental demand, physical demand, temporal demand, performance, effort, and frustration) were augmented by a question assessing patient interaction.
Pre/Post-visit Workload Survey	We asked physicians to estimate EHR workload before and after each visit (e.g., none, 5min, 10min, 15min, more than 15min). Questions were related to overall time, use of their own previous notes, other physicians' notes, new notes preparation, order entry, labs or imaging review, administrative tasks and clinical reminders.
Physician Interviews / Stimulated Recall	Semi-structured interviews with physicians, addressed perceptions of the EHR systems; use of the EHR in preparation of visits, and during consultations; strategies for balancing patients and clinical demands; physicians' attitudes, and care delivery models..

Table 2. Data collection and instrumentation.

5.3 Coding

Activity signals are generally coded to discrete signals as either point-event streams with a single timestamp per event (e.g., mouse clickstream) and interval-event streams based on 2 timestamps representing onset- and offset- (e.g., while a person is speaking, or non-verbal eye-gaze event). Point events can be counted, while interval event time can be totaled, as each spans a definite time duration. Coding schemas are summarized in Table 3. by data layer. To assess quality, a second human coder independently coded a randomly selected 10% subsample of visits.

Aim	Activity Layer	Data Coding Description
Aim 1	EHR Mouse Clickstream	Usability software MORAE provides a coarse trace (e.g., timestamps of mouse, keyboard clicks) but is not integrated with EHR functions. Rather, these were linked via manual review. CPRS and Epic differ in their features, two schemas were developed and later harmonized to CPRS "Tab" level.
	CPRS Mouse	For CPRS we use a "flat" coding scheme based on a previous usability study. Each mouse click is tagged to several levels of abstraction. At the top level (1) the 13 CPRS Tabs.. At a second level (2) to ~50 tasks grouped into "order entry", and "views". At a 3rd level, we tagged order entry activity (CPOE).
	Epic Mouse	Epic workflow, features and naming convention from CPRS. This required an alternate coding scheme. Using a computer-assisted method, each mouse click was coded to a variable-length "tag-path" descriptor using names of windows and menu items. Tag-path sequences were pooled and rank-ordered. We mapped the most frequent 100 tags to CPRS Tab equivalent.
	Computerized Order Entry	Order entry activity was coded as a separate level of detail and can be compared between CPRS and Epic. Coders manually tracked individual orders from beginning to end, noting order type and specific items (eg medication name), enabling counting number of items, mouse clicks per item and time at task per visit.
	EHR Keyboard	MORAE's keyboard clickstream is a direct record of physicians typing. To avoid exposing identifiable data, we mapped keys to a small set of pseudo-characters, distinguishing alphabetic keys, numeric keys, backspace, password masked keys, and control keys.
Aim 2	Physician Nonverbal Behavior	Physicians' nonverbal behaviors were coded based on room video review to a set of discrete behaviors (eg, physicians' eye gaze). A start point and end point gives interval stream. This enables summing time and merging conditional on other data (eg patient vocalization, questions asked)
	Patient and Physician Vocalization	Patient-provider verbal communication was based on manually coded audio for talk-silence of physician, patient, and patient companion.
	Patient and Physician Discourse	Vocalization doesn't indicate the content of the conversation. So we additionally coded for and counted occurrences of physicians' and patient speech that reflects participation (eg question asking, responses)

Table 3. Data coding description by study Aim and activity layer. Aim 3 (Cognitive Load and Satisfaction) and Aim 4 have no human coding components.

5.4 Quantitative Analysis

The distribution of patient's and provider's characteristics as well as other covariates will be assessed and summarized using descriptive statistics (median and interquartile range) for continuous measures, and frequency and percentage for categorical measures) and graphs. Our primary units of analysis are individual visits, segments within a visit (eg, beginning versus end), and groups of visits aggregated by covariate factors (eg, primary care versus specialty). The variation in usability, workflow, communication and cognitive load by study site, and by specialty , and the association between usability, workflow, communication and cognitive load are assessed. Association between two continuous or ordinal measures are assessed using Spearman's correlation coefficients, and the association between a continuous and a categorical measure will be assessed using the Wilcoxon rank sum test or the Kruskal-Wallis test. For categorical measures, we will use Fisher's exact or the Chi-square test. Since patients are nested within providers, we assess association via two complementary methods. First, using random effects model to account for correlation between patients from same provider. Multivariable model are also used to study these associations with adjustment for patients and provider characteristics. Since the visit length is a key confounder, all regression analysis will be adjusted for visit length. Second, we use a non-parametric method where associations of individual covariates are rank ordered based on the distribution of the empirical correlation of visit activity, patient complexity and other measures, against TLX subscales aggregated at the level of the physician.

5.5 Qualitative Analysis
Iterative qualitative review of the interview transcripts were conducted using methods recommended by Patton[3]. Analysis includes 1) initial inductive identification of emergent concepts, 2) discussion and development of associated quotations content to identify constructs, and 3) aggregation constructs into thematic categories.

5.6 Measures
Table 4. summarizes key measures for the study.

Aim	Measures with units
Aim 1	Mouse Clicks (count)
	Keyboard keystrokes (count)
	Order Entry (number of orders, timing, mouse clicks per order)
Aim 2	Visit Length (timing) NonVerbal (timing, gaze dom ratio) Vocalization (timing, voc dom ratio) Behaviors (count)
Aim 3	TLX / Effort and Performance subscales
Aim 4	*Correlation between aims 1,2,3

Table 4. Table of measures derived in this study
*Non-parametric component, we used Spearman Rho. For regression, L2 (Expectation Maximization)

5.7 Limitations

5.7.1 Study Design Limitations
Study limitations include the observational design (eg patients are not randomized, there is no intervention); cognitive processes and confounders like complexity of the patient and visit are not easily observable. Manual coding limits sample size, while testing for the combination of medical conditions, population and visit characteristics requires larger samples to achieve adequate statistical power (alternative ethnographic methods either share this limitation or would be limited in the number of data layers analyzed). We capture most activity data during the visit window only, yet physicians or their staff often perform EHR tasks outside the visit. It is difficult to treat variables such as scheduled visit length or whether it began late (adding to time pressure) and medical reasons for the visit.

5.7.2 Site and EHR Differences
While most data collection, coding and analysis proceeded in the same way across sites and specialties, there were differences between sites in organizational structure, care delivery model, and incentives (see Discussion). Key differences between sites include scheduled length of visit (20/40min at UCSD versus 30/60min at VASD) and support staff for pre/post visit work and variations in EHR features. Different features and naming conventions in CPRS and Epic complicated comparison. Epic functions like association of medication orders to diagnosis seem to have no counterpart in CPRS. Epic also has automation support for restructuring patient history across multiple interfaces and importing in notes (whereas CPRS has a basic template import) and enables working on notes without blocking other activity, further complicating comparison of EHR activity.

5.7.3 Instrumental Limitations
Our data collection is limited to a narrow set of clinical environments, specifically: outpatient settings during well defined point of care encounters in exam rooms. Tracking instruments require expertise and collaboration

with clinical IT staff. Some visits were marred by instrumental failures and dropouts, which further reduced sample size or introduced noise. Eyetracking could not be registered to specific EHR since the display content changed dynamically. Eyetracker calibration accuracy required that the monitor not be moved, inhibiting physician screen sharing behaviors with patients. Tracking devices (e.g., Kinect) and spatial audio array (Microcone), introduced to automate coding of nonverbal behavior and vocalization, were not sufficiently accurate could not classify all manually coded behaviors (e.g., physical exam), thus are not reported here.

6. Results

6.1 Principal Findings

We describe results for the study as a whole and for principal populations: by site (UCSD vs VASD) and by specialty (primary vs specialty care) and patient status (new vs. established). Due to space limitations, we only report statistics at the physician level of aggregation ad-hoc and exclude other important groupings (e.g., by patient or provider demographics, by clinic location).

6.1.1 Patient and Physician Demographics

Patient demographics are summarized in Table 5a. and physician demographics and EHR and institutional experience are given in Table 5b.

	Study Pop.	**By Site**		**By Specialty**		**By Patient Status**	
	n=223 (100%)	**UCSD** n=116 (52%)	**VASD** n=107 (48%)	**Primary** n=127 (57%)	**Specialist** n=96 (43%)	**New** n=48 (22%)	**Established** n=175 (78%).
Gender (count)	M= 148 (66%) F= 75 (34%)	48 (41%) 68 (59%)	100 (93%) 7 (7%)	84 (66%) 43 (34%)	64 67%) 32 (33)	32 (67%) 16 (33%)	116 (66%) 59 (34%)
Age (years)	63 (51,70)	66 (56,76)	59 (48,67)	64 (48,72)	62 (55,68)	59 (39,67)	64 (54,71)
Race (count)	White=153 (69%) Afr Am.= 22 (10%) Asian=11 (5%) Other,N/A= 37 (17%)	82 (71%) 7 (6%) 6 (5%) 21 (18%)	71 (66%) 15 (14%) 5 (5%) 16 (15%)	94 (74%) 11 (9%) 10 (8%) 12 (9%)	59 (61%) 11 (11%) 1 (1%) 25 (26%)	36 (75%) 4 (8%) 2 (4%) 6 (13%)	117 (67%) 18 (10%) 9 (5%) 31 (18%)

Table 5a. Patient Demographics. Acronyms for Gender: M=Male, F=Female. For Race, Other includes American Indian or Alaskan Native, Native Hawaiian or Pacific Islander, Native Hawaiian, Native American/Eskimo, Filipino, Unanswered, declined to answer, unknown or missing or other race.

	Study Pop.	**By Site**		**By Specialty**	
	n=32 (100%)	**UCSD** n=17 (53%)	**VASD** n=15 (47%)	**Primary** n=17 (53%)	**Specialist** n=15 (47%)
Gender (count)	M=19 (59%) F=13 (41%)	9 (60%) 6 (40%)	10 (59%) 7 (41%)	11 (65%) 6 (35%)	13 (87%) 2 (13%)
Years with institution	8 (5,17)	8 (5,19)	7 (4,15)	11 (5,21)	7 (4,13)
Years with current EHR	8 (5,9)	8 (5,9)	7 (4,13)	9 (5,9)	7 (4,9)
Multiple EHR use (count)	Yes=22 (69%) No=8 (25%) N/A=2 (6%)	10 (67%) 4 (27%) 1 (7%)	12 (71%) 4 (24%) 1 (5%)	9 (53%) 7 (41%) 1 (6%)	13 (87%) 1 (7%) 1 (7%)

Table 5b. Physician Demographics, institutional and EHR experience.

The patients at UCSD are significantly older than those at VA (VASD: 66 years vs. UCSD: 59 years, p=0.003). As expected VA has more male patients than UCSD (p<0.001). There is no significant difference in patient characteristics between primary care visits and specialty care visit. There is no significant difference in provider characteristics between two study sites or between primary care providers and specialists.

6.1.2 Visit and Patient Complexity

We report the complexity of the visit in terms of physician's assigned CPT code. Patient complexity is assessed based on Charlson Comorbidity Index[2] based on the active problem list at the time of the encounter, extracted from the EHR. For UCSD, we extracted active problem list (consisting entirely of ICD-9-CT codes) from Epic logs. For VASD, we manually extracted active problem list (consisting of ~85% ICD-9-CT and 15% SNOMED-CT) from CPRS. These distributions are summarized in Table 6.

	Study Pop.	By Site		By Specialty		By Patient Status	
	n=219 (100%)	UCSD 112 (51%)	VASD 107 (49%)	Primary 127 (58%)	Specialist 92 (42%)	New 48 (22%)	Established 171 (78%)
CPT (count)	99214= 97 (44%) 99213= 55 (25%) 99204= 17 (8%) Other= 50 (23%)	72 (64%) 17 (15%) 10 (9%) 13 (12%)	25 (23%) 38 (36%) 7 (7%) 37 (35%)	59 (46%) 44 (35%) 1 (1%) 23 (18%)	38 (41%) 11 (12%) 16 (17%) 27 (29%)	6 (13%) 11 (23%) 15 (31%) 16 (33%)	91 (53%) 44 (26%) 2 (1%) 34 (20%)
CCI* (raw)	0 (0,1)	0 (0,2)	0 (0,0)	0 (0,0)	0 (0,3)	0 (0,1)	0 (0,1)
CCI* (age adjusted)	2 (1,4)	2 (1,4)	2 (1,4)	2 (1,3)	2 (1,5)	2 (0,4)	2 (1,4)

Table 6. Summary of patient and visit complexity extracted from active problem list and CPT codes. CCI, or Charlson Comorbidity Index.

6.1.3 Intercoder Agreement

We measure intercoder agreement for a subset of activity signals as shown in Table 7 and falls in two categories: For discrete (point-event) activity like mouse clicks, we compare the sequential activity on a click by click basis, generating a 2x2 matrix. Since both coders use the same timestamps generated by Morae, the total number of clicks is a common denominator. 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. For continuous activity signals (NonVerbal, Vocalization) we use a time-resolved method where the coded events tagged by specific behavior or code (e.g., Gaze to EHR) determine a set of intervals which are intersected to determine their common time extent, and then normalized by the independently estimated visit length. For example, if Coder 1 coded the interval [1sec,5sec] as Gaze to EHR, while Coder 2 coded the event as [1.5sec,5.5sec] (i.e., a relative shift of 1/2 sec) their intersection is the interval [1.5sec,5sec], which is an overlap of 3.5sec, even though they have the same total duration of 4sec. Since Vocalization activity evolves on a second-to-second basis (based on histogram of speaking-event durations), even small time shifts by one coder relative to the other can degrade agreement. In fact this time-resolved method results in significantly lower intercoder agreement for Vocalization than for NonVerbal. This implies that NonVerbal signal is suitable for time-resolved association studies such as timing of EHR activity conditional on physician gazing to EHR, whereas Vocalization signal would be less appropriate for such association timing studies (e.g., proportion of time that the patient speaks while physician is gazing at EHR). Therefore, we also report for Vocalization the agreement metric based on the total time across speakers. Throughout a visit, patients and physicians take turns, and the mean total time errors are averaged and normalized by visit length to yield the per-visit agreement score.

Measure	Sample Size (Visits)	Intercoder Agreement	
		Method	**Agreement:** (Median, IQR)
EHR CPRS (Aim 1)	n=15 (15 VASD)	Sequential Tab-level comparison	0.98 (0.97-1.0)
EHR Epic (Aim 1)	n=11 (11 UCSD)	Sequential CPRS-equivalent Tab-level comparison	0.92 (0.69-0.94)
NonVerbal (Aim 2)	n=21	Time-resolved comparison	0.94 (0.86-0.95)
Vocalization (Aim 2)	n=7	Time-resolved comparison Averaged sum of speaker time comparison	0.64 (0.56-0.7) 0.96 (0.88,0.99)

Table 7. Summary of data coding quality in terms of intercoder agreement across dual-coded visits.

6.1.4 Variation in Sample Size

The baseline sample size for the study is n=223 visits. In subsequent analysis, smaller sample sizes, varied by aim and specific analyses as specified in each table. This variation is due to a combination of these factors: instrumental failures, such as Morae mouse activity dropout vs video capture dropout; only a subsample of visits human-coded, need to intersect between data layers, e.g. EHR timing conditional on human-coded NonVerbal, and associations in Aim 4.

6.2 EHR Activity (Aim 1)

6.2.1 Mouse Clicks, Mouse Path Length and Keystrokes

Table 8. summarizes top-level measures of EHR activity based on raw Morae capture, before human coding to specific EHR functions.

	Study Pop.	By Site		By Specialty		By Patient Status	
		UCSD	VASD	Primary	Specialist	New	Established
Mouse Clicks*	**87 (46,142)** n=222 (100%)	**66 (37,110)** n=116 (52%)	**112 (63,207)** n=106 (48%)	**98 (52,171)** n=127 (57%)	**74 (37,128)** n=95 (43%)	**102 (39,142)** n=48 (22%)	**85 (47,137)** n=174 (78%)
Keystrokes	**432 (98, 1137)** n=194 (100%)	**463 (87,1121)** n=107 (55%)	**412 (98,1305)** n=87 (45%)	**347 (87,830)** n=107 (55%)	**671 (104,1530)** n=87 (45%)	**1157 (120,2015)** n=45 (23%)	**356 (87,895)** n=149 (77%)
Mouse Path**	**90 (54,143)** n=201 (100%)	**66 (42,103)** n=116 (58%)	**138 (87,190)** n=85 (42%)	**103 (64,153)** n=109 (54%)	**84 (46,125)** n=92 (46%)	**90 (44,151)** n=44 (22%)	**90 (58,142)** n=157 (78%)

Table 8. Summary data based on non-human-coded Morae activity tracking of EHR (*) mouse clicks exclude wheel events. (**) Mouse Path length is Euclidean (L2) distance in kilo-pixels. Due to various types of Morae dropouts, sample sizes vary by data layer.

We did not find any significant difference in keystrokes and mouse clicks between sites or between primary care providers and specialists from random effects regression analysis with adjustment for visit length. There is a significant difference in keystrokes by patient status.

6.2.2 Specific EHR Functions Accessed

We report a comparison of the two EHRs: CPRS and Epic, 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 have in common (Notes, Orders, Labs, meds and Reports) and group all remaining Tabs as "Other" as shown in Table 9a. Timing in Table 9a. is estimated based on time-resolved association between EHR and NonVerbal activity and 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. Table 9b. shows this EHR activity by study subgroups.

	CPRS (VASD) n=89 (16668 mouse clicks)		Epic (UCSD) n=106 (8280 mouse clicks)		
Common and Frequent Tabs	**Timing** (min)	**Count**		**Timing** (min)	**Count**
Notes	578 (58%)	8300 (50%)	Notes	311 (41%)	1842 (21%)
Orders	198 (20%)	4547 (27%)	Orders	117 (16%)	1960 (23%)
Labs	55 (5%)	1084 (7%)	Labs	41 (5%)	639 (7%)
Meds	43 (4%)	666 (4%)	Meds	20 (3%)	274 (3%)
Reports	24 (2%)	403 (2%)	Reports	64 (9%)	693 (8%)
Other category (Consults, Cover, Discharge, Patient Selection, Problems, Review/Sign, Surgery, Unidentified)	107 (1%)	1668 (10%)	Other category (Association, Cover, Patient Selection, Problems, Review/Sign, Surgery, or Unidentified)	195 (26%)	3272 (38%)

Table 9a. Comparison of EHR function activity between the two sites based on mouse clicks. VASD uses CPRS, while UCSD uses Epic.

	Study Pop.		**By Site**		**By Specialty**		**By Patient Status**	
	Visits with tab use (n=195)	n=195 (100%)	UCSD n=106 (54%)	VASD n=89 (46%)	Primary n=113 (58%)	Specialist n=82 (42%)	New n=41 (21%)	Established n=154 (79%)
Notes	182 (95%)	23 (8,59)	13 (4,28)	65 (17,120)	22 (8,57)	26 (7,59)	30 (3,104)	22 (8,47)
Orders	180 (94%)	23 (8,45)	16 (5,29)	44 (16,80)	31 (11,55)	16 (5,31)	20 (7,44)	23 (9,45)
Labs	124 (65%)	4 (0,13)	2 (0,8)	9 (0,19)	4 (0,12)	3 (0,16)	4 (0,12)	4 (0,13)
Meds	108 (56%)	1 (0,6)	1 (0,3)	1 (0,11)	2 (0,7)	0 (0,3)	0 (0,1)	2 (0,7)
Reports	116 (60%)	2 (0,7)	2 (0,7)	1 (0,5)	1 (0,5)	3 (0,8)	1 (0,8)	2 (0,7)
Other	188 (98%)	20 (9,34)	27 (17,39)	12 (3,26)	22 (12,40)	17 (7,28)	17 (7,33)	20 (10,35)

Table 9b. Tabulation of mouse clicks in screens common to CPRS and Epic. The small proportion of clicks to Meds is due to order entry of medications is carried out in the Orders Tab.

6.2.3 EHR Navigation

We report EHR Navigation patterns in terms of the number of "Tab" (screen-level) transitions during the visit, based on mouse click activity (Table 10.), as well as the most frequently occurring Tabs in these transitions. Most consecutive mouse clicks occur in the same screen: from a total sample of ~25k clicks, we observed ~4k transitions between Tabs. 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 to and from {Notes,Other} (26%), {Orders,Other} (24%), {Notes,Orders} (12%), {Labs,Other} (7%), {Labs,Notes} (6%), {Meds,Notes} (5%). Notes, and Orders were hubs of activity to which physicians navigated multiple times during the visit.

	Study Pop.	**By Site**		**By Specialty**		**By Status**	
Count	n=195 (100%)	UCSD n=106 (54%)	VASD n=89 (46%)	Primary n=113 (58%)	Specialist n= 82 (42%)	New n=41 (21%)	Established n=154 (79%)
Screens	16 (9,27)	22 (13,34)	12 (8,22)	21 (11,32)	14 (8,22)	15 (10,26)	17 (9,29)

Table 10. Number of "Tab" or screen changes based on mouse clicks tagged to the top-level screen or "Tab" coding as described in Section 6.2.2.

6.2.4 EHR Order Entry Comparison by Site Due to varying naming conventions between CPRS and Epic, we grouped orders as follows: on the VASD side, Allergies and Procedures were grouped as Reminders, and Nursing and miscellaneous other orders as Other. VA and Non VA medications were grouped as Meds. On Epic, Health Maintenance were considered as Reminders. Aggregating by site, Table 11. 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, Consults and Imaging orders take the longest to complete on VASD using CPRS, followed by medications, and that labs take less time and fewer clicks on average. The statistics for UCSD side using Epic are similar. One major difference is the preponderance of Reminders activity by VA primary providers.

	UCSD			**VASD**		
Order Type	**Visits** n= 106 (100%)	**Timing per item ordered** (sec)	**Mouse clicks per item ordered**	**Visits** n=89 (100%)	**Timing per item**	**Mouse clicks per item**
Consult	27 (25%)	49 (29,75)	11 (8,15)	44 (49%)	52 (36,82)	16 (11,24)
Imaging	22 (21%)	47 (28,86)	8 (6,11)	20 (22%)	43 (28,70)	12 (6,17)
Lab	32 (30%)	12 (6,28)	4 (3,6)	44 (49%)	12 (7,19)	5 (3,7)
Med	54 (51%)	38 (24,78)	10 (6,13)	54 (61%)	26 (18,47)	9 (7,12)
Other	5 (5%)	3 (25,93)	6 (5,8)	6 (7%)	6 (8,32)	4 (4,14)
Reminder*	5 (5%)	9 (6,11)	5 (4,6)	33 (37%)	19 (14,32)	5 (4,7)
Return to Clinic	58 (55%)	12 (7,30)	3 (2,4)	37 (42%)	25 (16,47)	9 (8, 14)

Table 11. Comparison of computerized order entry frequency, time-at-task per order and EHR user interface burden as measured by numbers of clicks to complete each order. in Epic Reminders are called Health Maintenance. The Visits columns reflect the number and proportion of visits in which the specific order type is encountered.

6.2.5 Thematic Analysis of Physician Interviews (Aim 1)

The interviews revealed that physicians had ambivalent views of EHRs. On the one hand, they provide easy access to information (compared to paper); on the other hand, EHR there is often too much information to sift through in a typical visit time frame. There is no “overview” or index for efficient navigation. The main barriers of EHR usage identified in the study include: a) **Inefficient interfaces:** EHR requires many mouse clicks to navigate and long search strings for information retrieval, documentation and order entry. b) **Lack of robustness** for complex clinical situation, the effort of using EHR increases forcing them to use various effort-reducing strategies and workarounds. Almost everyone mentioned the differences between new patients, returning patients, and the frequent visitors. c) **Lack of training**: The difficulty of using the EHR also showed up in the many complaints about the lack of training: most said that received no training and it is difficult to share and learn efficiency tips from other clinicians. f) **EHR is not the only factor.** EHR is just one factor among many that influences workflow and perceived task load, other factors included institutional rules, support staff levels, and scheduling flexibility.

6.3 Clinical Workflow and Communication (Aim 2)

6.3.1 Visit Length and Association to Pre/post visit work

Visit length was measured based on study protocol where visit begins when patient and physician are first in the room together, and ends when the physician finally exits the room. Pre and post-visit work was estimated based on physicians' workflow survey as follows: We mapped survey responses (time ranges) to time estimates as follows: ("None" = 0 min, "0-5 min" = 2.5 min, "5-10 min" = 7.5 min, "10-15 min" = 12.5 min, ">15 min" = 22.5 min. We assume a ceiling of 30 min per visit, for each of pre and post work. Total outside-visit time is not significantly associated with visit length. The average visit length is 6.5 mins longer at VA than at UCSD, this is statistically significant (p=0.02) from random effects regression analysis.

	Study Pop.	**By Site**		**By Specialty**		**By Patient Status**	
(minutes)		UCSD	VASD	Primary	Specialty	New	Established
Visit Length	**18.5 (13.3-24.9)** n=223 (100%)	**16.5 (12.1-20.6)** n=116 (52%)	**20.4 (15.5,29.7)** n=107 (48%)	**17.8 (12.6,24.1)** n=127 (57%)	**19.4 (15.1,25.2)** n=96 (43%)	**24.7 (18.2,33.4)** n=48 (22%)	**17.4 (12.5,21.9)** n=175 (78%)
Pre-visit work	**2.5 (0,7.5)** n=214 (%)	**2.5 (2.5,7.5)** n=116 (%)	**2.5 (0,7.5)** n=98 (%)	**2.5 (0,7.5)** n=119 (%)	**2.5 (2.5,7.5)** n=95 (%)	**2.5 (2.5,2.5)** n=48 (%)	**2.5 (0,7.5)** n=166 (%)
Post-visit work	**7.5 (2.5-7.5)** n=210 (100%)	**2.5 (2.5,7.5)** n=113 (54%)	**7.5 (2.5,7.5)** n=97 (46%)	**7.5 (2.5,7.5)** n=115 (55%)	**7.5 (2.5,7.5)** n=95 (45%)	**5.0 (2.5,7.5)** n=48 (23%)	**7.5 (2.5,7.5)** n=162 (77%)

Table 12. Visit Length and pre- and post-visit work time.

6.3.2 Physician Nonverbal (Eye-gaze) Timing

Nonverbal (physician eye-gaze) behaviors time-at-tasks are summarized in Table 13.

	Study Pop.	**By Site**		**By Specialty**		**By Patient Status**	
(minutes)	218 (100%)	UCSD n=114 (%)	VASD n=104 (%)	Primary n=122 (%)	Specialist n=96 (%)	New n=47 (%)	Established n=171 (%)
Gaze to EHR	7.4 (5.0,11.3)	6.4 (3.6,9.4)	9.2 (6.0,14.6)	7.3 (4.7,11.0)	8.2 (5.3,11.3)	9.9 (4.4,18.3)	7.2 (5.0,10.6)
Gaze to Patient	6.6 (3.8,10.2)	5.7 (3.1,8.7)	7.4 (4.3,11.0)	6.2 (3.6,9.8)	7.0 (3.9,10.9)	9.3 (6.9,13.8)	5.5 (3.3,9.4)
Physical Exam	1.6 (0.9,2.6)	1.7 (1.0,2.7)	1.5 (0.7,2.4)	1.6 (0.9,2.8)	1.5 (0.8,2.3)	1.7 (0.9,2.8)	1.6 (0.9,2.6)
Gaze Dom (ratio)	1.3 (0.6,2.4)	1.1 (0.5,2.2)	1.4 (0.7,2.4)	1.3 (0.7,2.4)	1.2 (0.5,2.1)	1.0 (0.4,2.0)	1.3 (0.6,2.5)

Table 13. Nonverbal Timing, median and IQR in minutes.

There is no significant difference in nonverbal gaze time between sites or between primary care providers and specialties from the random effects regression analysis with adjustment for visit length.

6.3.3 Physician and Patient Vocalization Timing

Table 14. summarizes the total speaking time for physicians, patients (and companion if present), silence and overlapping speech. To simplify computations, we further assume that if there is companion present, not all parties speak simultaneously.

	Study Pop.	**By Site**		**By Specialty**		**By Patient Status**	
(minutes)	n=163 (100%)	UCSD n=68 (42%)	VASD n=95 (58%)	Primary n=114 (70%)	Specialist n=49 (30%)	New n=30 (18%)	Established n=133 (82%)
Patient/ Comp*	5.8 (3.2,8.5)	5.8 (3.2,8.3)	6.0 (3.2,8.8)	6.0 (3.1,8.5)	5.7 (3.8,8.2)	7.3 (4.0,10.3)	5.8 (3.1,8.3)
Physician	7.5 (5.0,11.5)	5.9 (3.8, 8.7)	8.4 (6.3,13.2)	7.0 (4.5,10.9)	9.1 (6.7,12.0)	8.9 (6.5,13.7)	7.2 (4.5,11.2)
Overlap	0.5 (0.1,1.0)	0.7 (0.2,1.3)	0.3 (0.1,0.8)	0.4 (0.1,0.9)	0.7 (0.1,1.5)	0.4 (0.1,0.6)	0.5 (0.1,1.0)
Silence	4.6 (2.6,7.6)	3.8 (2.3,5.3)	5.8 (2.9,9.2)	4.7 (2.9,7.6)	4.0 (2.2,7.3)	5.5 (3.8,9.7)	4.4 (2.5,7.0)
Voc Dom (ratio)	1.4 (0.9,2.1)	1.0 (0.7,1.7)	1.6 (1.1,2.4)	1.3 (0.9,2.0)	1.6 (1.0,2.2)	1.4 (1.0,2.0)	1.4 (0.9,2.1)

Table 14. Vocalization timing median and IQR in minutes, and vocalization dominance ratio. [*] Pat/Comp includes patient and companion time. Vocalization Dominance is defined as the ratio of physician time over patient or companion.

Although the average time that patient or companion talk is high at VA visits, but after controlling for visit length and cluster effect of provider, we found that patient or companion talk time is significantly higher at UCSD ($p=0.03$). We also assessed the percentage of time that patient or companion talk during the visit, we found the same results, that is, UCSD patients talk more than VASD patients. We did not see any significant difference in other vocalization time between sites or between primary care providers and specialties from the random effects regression analysis with adjustment for visit length.

6.3.4. Verbal Behavior Counts

Table 15. summarizes tallies of physician and patient behaviors during visit.

		Study Pop.	By Site		By Specialty		By Patient Status	
(counts)	**Behavior**	n=164 (100%)	UCSD n=69 (42%)	VASD n=95 (58%)	Primary n=115 (70%)	Specialist n=49 (30%)	New n=30 (18%)	Established n=134 (82%)
Patient Behavior	Questions	3 (1,6)	3 (1,5)	3 (1,7)	2 (1,5)	5 (2,6)	3 (1,8)	3 (1,6)
	Assertive Resp.	3 (2,6)	3 (1,4)	4 (2,6)	4 (2,6)	3 (2,4)	3 (1,4)	3 (2,6)
	Expr. of Concern	0 (0,0)	0 (0,0)	0 (0,1)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Physician Behavior	Partnership Build.	3 (2,4)	2 (1,3)	4 (2,6)	3 (2,5)	3 (2,4)	3 (2,4)	3 (2,5)
	Supportive Talk	0 (0,1)	0 (0,0)	1 (0,3)	0 (0,1)	0 (0,1)	0 (0,1)	0 (0,1)

Table 15. Verbal Behaviors counts

6.3.5 Thematic Analysis of Physician Interviews (Aim 2)

The following themes extracted from the interviews are relevant to the problem of communication and workflow. a) **Computerized Communication is ubiquitous**. The EMR is the center of care for all team members and computer-mediated communication is at the heart of information exchange. Clinicians use the EHR to communicate with their patients' care teams. They do so by assuming that others are reading their documentation, enacting their orders, or following agreed upon protocols. The result is assumed to be a shared awareness of the patient's situation. *" I try not to because that's one more thing which has been added on. Because I have GUI mail we have view alerts, we have secure messaging and the Lync is another thing. So, when I'm with patients I'm looking through my GUI mail, I'm looking through view alerts, making phone calls and then I have somebody calling in me on the Lync oh there's such and such patient. You know, it could be a call center, it could be anybody and that's another thing to my stuff. So, unless it's really urgent I avoid using it."*

b) Poor integration of clinical workflow with EHR results in degraded communication. EHRs (or Health IT in general) support multiple communication forms (e.g. notifications, view alerts) and some external sources, such as phone calls, texts and emails. However, they are not well integrated with the EHR, so the result is a commonly expressed feeling that most communication events are interruptive.

c) **Patient-provider interaction is impacted by EHR in the room**. Some rooms, particularly at UCSD, were described as being patient unfriendly requiring physicians to turn their back to the patient while using the EHR (note: computers are mounted on swing arms and are adjustable, physicians often did not adjust to optimize) Two workarounds were reported: using paper instead of EHRs during visits (this necessitates after-visit work by the clinician or staff) and multi-tasking between attending to the EHR and the patient. *" Well, I think we're all trying to multitask when we're in a clinic visit. So, we're trying to. . . some people try and take histories and type at the same time and look at lab tests all at the same time. And I think they get used to that but it may not be the best practice… it's very disruptive to the patient conversation and they get offended by it. But the way the computers are set up it's forcing us to do it."*

d) **Duplicative Workflow in terms roles:** Because of lack of automation and administrative rules, physicians do redundant work such as documenting order entry activity in notes.

6.4 Subjective Workload and Satisfaction (Aim 3)

6.4.1 NASA Task Load Survey

Subjective workload was measured using the NASA TLX. Data reduction was conducted on clinician's ratings of individual items. Specifically, we conducted an item response analysis (correlations and principal component analysis). Two factors explained 65% of the variance. Inspection of the factors indicated 2 scales: a 5-item *Effort* subscale consisting of Mental, Physical, Temporal, demands, Effort and Frustration ratings with a Coefficient Alpha of 0.93. A second subscale, *NegPerformance,* was comprised of 2 items, Performance and Patient Interaction with a Coefficient alpha of 0.51. Both scales had values ranging from 1-20, inclusive, with lower numbers indicating preferred outcomes (i.e., lower effort or greater performance).The subscale responses are summarized in Table 16.

	Study Pop.	**By Site**		**By Specialty**		**By Patient Status**	
	n=215 (100%)	UCSD n=116 (54%)	VASD n=99 (46%)	Primary n=119 (55%)	Specialist n=96 (45%)	New n=48 (22%)	Established n=167 (78%)
Median TLX	6.3 (4.3,9.1)	6.1 (4.5,8.4)	7.3 (4,9.5)	6 (4,8.4)	7.1 (5.1,9.3)	6.4 (5,9.2)	6.3 (4.1,9.1)
NegPerformance	5.5 (2.5,8.5)	6.5 (3.5,9)	4 (2.5,8)	4 (1.5,7.3)	6.5 (4.5,9.1)	6.5 (4.1,10)	5.5 (2.5,8)
Effort	6.4 (4.2,9.6)	5.7 (4.2,8.7)	7.6 (4.5,10.2)	6 (4,8.6)	6.8 (4.4,10.2)	6.4 (4.3,9.1)	6.4 (4.2,10)

Table 16. TLX Median and interquartile ranges for overall TLX scores and for 2 subscales derived from item-response analysis.

There is no significant difference in two subscales between sites or between specialties by random effects regression analysis.

6.4.2 Thematic Analysis of Physician Interviews (Aim 3)

Perceived effort is related to the experience of cognitive load and the match between the demands of the task and cognitive effort required to perform well. An overarching theme is that the workload and communication demands create cognitive load, that the EHR contributes or does not help the clinician manage this load and that providers were forced in our study have to develop their own distinctive and personalized working styles. Some providers prefer to finish all the documentation in the room, some choose to do extensive pre-preparation and some choose to work on notes in their private office after seeing patients. A few doctors even developed their own documentation tools where they use paper to take notes during consultation, then transcribe it into progress notes in the EHR. Though the working styles and workflow differ, many providers mentioned that the way they use the EHR works for them to manage their information environment, so they can accomplish their information management goals[4]. These personalized working styles are ways for providers to mitigate the high cognitive load, improve efficiencies and thereby, improve their performance. There were many strategies offered. Providers had both positive and negative views of using the EMR. On the one hand, easy access to information is important and highly useful; on the other hand, using the EMR is something akin to drinking from a water hose. The resultant ecology, set of strategies and practice culture that has resulted from this situation constitutes the bulk of this thematic category.

6.5 Association of Visit Activity to Subjective Workload (Aim 4)

We use TLX responses as measures of subjective (self-reported) workload. These responses clustered by physician, with some individuals' responses clustering in low values, others clustering toward higher values, and some with larger variance. Because of this variable location and dispersion characteristics, responses cannot be compared uniquely between clinicians.

To address the heterogeneity of subjective responses, we use two complementary methods: a parametric regression model and a non-parametric method where we rank order the visit activity factors by their physician-aggregated correlations to TLX responses. We use Spearman rho as the correlation measure which only takes ranks into account rather than activity metric values, thus can be compared across activity layers.

6.5.1 Regression Model

The associations of TLX subscales with EHR activity, nonverbal gaze time and vocalization time were assessed using random effects model to account for the correlations between patient visits within the provider. The results are given in Table 17. We found that the longer visit length is significantly associated with more TLX effort ($b=0.13$, $p<0.001$). We also found that less gaze time at patient ($b=-0.18$, $p=0.006$) and more physical exam ($b=0.34$, $p=0.02$) are associated with more TLX effort . Also, less patient talk ($b=-0.18$, $p=0.046$) and more silence ($b=0.26$, $p=0.03$) are significantly associated with more effort. For TLX performance, we found it is marginally significant associated with provider talk time ($p=0.06$) and silence time ($p=0.06$), with more silence and less provider talk are associated with more negative TLX performance. All analyses were adjusted for visit length.

	Correlates	**EffortTLX subscale**			**NegPerformanceTLX subscale**		
		b	**se**	**p**	**b**	**se**	**p**
EHR (Aim1)	Mouse clicks	0.002	0.002	0.19	0.001	0.002	0.55
	Keystrokes	<0.001	<0.001	0.81	0	0.001	0.73
Nonverbal (Aim 2)	Visit Length	0.13	0.02	<0.001	0.04	0.03	0.20
	Gaze to Patient	-0.18	0.06	0.006	-0.03	0.07	0.67
	Gaze to EHR	0.08	0.07	0.21	0.04	0.07	0.63
	Physical exam	0.34	0.15	0.02	-0.04	0.17	0.83
Vocalization (Aim2)	Provider talk	-0.07	0.10	0.50	-0.19	0.10	0.06
	Patient talk	-0.18	0.09	0.046	-0.06	0.09	0.53
	silence	0.26	0.08	0.003	0.17	0.09	0.06

Table 17. Association of visit activity to TLX - regression Model.

6.5.2 Physician-Aggregated Correlations

In addition to the regression model, we also use a non-parametric approach to estimate the association of activity measures with self-reported (TLX) workload. Table 18. shows the top 10 visit activity measures most correlated in magnitude to the TLX subscales based on physician-aggregated correlation distribution (we use the robust correlation Spearman rho, which is only sensitive to rank orders rather than the numerical value of a measure). We arbitrarily truncate absolute rho values less than 0.2.

Activity Measure	**Subjective Workload** (TLX subscale)	**Sample Size** (physicians)	**Physician-aggregated Correlation** Spearman rho: median, (IQR)	**Rank** (median)
Visit Length (minutes)	Effort (5-item)	n=32 (100%)	0.49 (0.12,0.67)	1
EHR Tab (screen) changes (count)	NegPerformance (2-item)	n=29 (91%)	0.42 (-0.36,0.63)	2
EHR Tab (screen) changes (count)	Effort	n=29 (91%)	0.38 (0.14,0.61)	3
EpicLogSize (count)	Effort	n=16 (50%) UCSD only	0.35 (0.24,0.52)	4
Mouse Clicks (count)	Effort	n=32 (100%)	0.32 (0.03,0.6)	5
Gaze Dominance (ratio)	Effort	n=32 (100%)	0.28 (-0.07,0.58)	6
Charlson Comorbidity Index*	Effort	n=24 (75%)	0.27 (-0.29, 0.54)	7
Charlson Comorbidity Index*	NegPerformance	n=25 (78%)	0.26 (-0.23,0.47)	8
Mouse Path Length (pixels)	Effort	n=32 (100%)	0.23 (-0.06,0.56)	9
Verb. Patient Concerns (count)	NegPerformance	n=25 (78%)	0.21 (-0.22,0.32)	10
Keystrokes (count)	Effort	n=31 (97%)	0.2 (-0.07,0.49)	11

Table 18. 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. The NegPerformance subscale only appears as the 6th largest correlation magnitude. Charlson Comorbidity Index is based on the patients' active problem list at the time of the visit. Both a raw CCI and an age-adjusted CCI were computed as discussed elsewhere in this report. For CCI, the smaller sample size is due primarily to an entire patient panel's CCI equal to zero, in which case, Spearman rho is undefined, and these physicians were removed from the pool as dropouts.

6.5.3 Thematic Analysis of Physician Interviews (Aim 4)

Table 19. summarizes the key themes across Aims 1-3.

Thematic Title	Description	Example Quotation
Aim 1: The EHR is a "A double-edged sword" - Too many clicks - Rigid system - No training - High volume of data	Easy access to information is important and highly useful; on the other hand, using the EHR is something akin to drinking from a water hose.	*" I mean it's- if it was just a paper note our life would be much easier. But, then again, we wouldn't be as comprehensive and the bulk of information that CPRS provides us is my best dream come true, I would say, because it helps me be on top of everything."*
Aim 2: Patient –provider interaction changes as providers adjust to using the computer in the room.	Divided attention demands and poor room design has resulted in diverse strategies.	*"So, we're trying to. . . some people try and take histories and type at the same time and look at lab tests all at the same time. And I think they get used to that but it may not be the best practice… it's very disruptive to the patient conversation and they get offended by it."*
Aim 2: The EHR "dumbs-down" the work.	Physicians complain of increased clerical work, multiple interruptions and being the one shouldering all of the burden.	*"Everything is my responsibility, I'm doing hundred percent of the work so it definitely brings, gives you more control, if you're control freak little bit. But, yes, it just puts everything on your table."*
Aim 2: Virtual communication may be an illusion	Everyone assumes that everyone is reading everyone's notes – so there is an assumption of shared situation awareness that may not be real.	*" Because I have GUI mail we have view alerts, we have secure messaging and the Lync is another thing. So, when I'm with patients I'm looking through my GUI mail, I'm looking through view alerts, making phone calls and then I have somebody calling in me on the Lync oh there's such and such patient."*
Aim 3: Getting the "gist" is hard	Reviewing and synthesizing the patient information requires a lot of cognitive effort	*"I don't want to spend a lot of time reading a long comprehensive note when I can just hit the concise points and know how to help the patient."*
Aim 3: Workarounds are the only way to survive	Multiple, idiosyncratic adaptations to the cognitive load are ubiquitous and diverse.	*"And that strategy is to give the patient something that they have to focus on and write so that they're not expecting any interaction."*
Aim 3: Lack of expertise on utilizing EHR efficiently	Providers have a multitude of ways they have adapted to using the EHR and would like more training to be efficient.	*"…might be helpful… someone who is kind of sitting there with you and helping you find short cuts to the things that you need to have done."*
Aim 3: Using the system is frustrating.	Multiple expressions of negative emotion were expressed, including anger, frustration anxiety, etc.	*"It is extremely frustrating to have to hunt for information"*

Table 19. Findings and categories from thematic analysis of physician interview transcripts.

In summary, we found distinctive difference between simple and complex cases in the study, where complexity brings in increases in workload, as the EHR does not meet the more complex demands, which, in turn, leads to more complicated and often distracted workflow, higher cognitive load and frustration. Learning new strategies was frustrating and there was no time and little support.

6.6 Discussion

Overall, the results of this multi-focused study demonstrates vividly the complexity of healthcare delivery. This complexity is illustrated in our results and is congruent with other literature.

6.6.1 Sources of Variation

For most activity measures, the differences between individual physicians are more pronounced than by site, despite the fact that typical UCSD visits tend to be shorter, and somewhat more time is spent looking at the patient. Differences in EHR system design further limit comparability of results, as identifying high-level tasks and mapping those to workflows and command sequences as necessary for cross-site comparison requires manual identification and extraction of individual actions into higher-level semantic actions such as order entry and medication reconciliation. Despite these differences, high-levels of mastery of the EHRs was visible in many, if not all, of the video recordings, as providers confidently completed relevant tasks, even when doing so required complex and visibly inefficient sequences of operations. Clinicians were highly skilled at adapting the constraints of their local setting and EHR, using multitude of workarounds, time-saving strategies and various documentation approaches. Physicians were frequently seen moving quickly through multiple screens to complete complex orders, enter comments into clinical notes, reconcile notes with medication orders, and complete other tasks, often moving between multiple fields and tabs with little difficulty, and often doing so while engaged in conversation with patients.

6.6.2 Site Differences

There were many commonalities between sites, but some key differences help explain our findings. For example, visits were about 5 minutes shorter on average at UCSD and also for established patients, but these may be explained by the scheduled visit durations which are shorter on UCSD and shorter for established patients (at both sites) than for new patients. Another significant difference involves differences in staff support levels. Anecdotal evidence from interviews suggest that nurses and other care team members may perform some EHR tasks that VASD physicians perform.

Factor	UCSD	VASD
Scheduled visit lengths	20/40min visits (Follow up/New patient visit)	30/60 min visits (Follow up/New patient visit)
Patient study population	Balanced male/female patient demographics	Predominantly male patients
EHR Physician-facing interface	EpicCare Ambulatory	CPRS (Computerized Patient Record System)
EHR features, and configuration	• 9 doctors use the dual window • More levels of menus, objects and paths • Associations (Dx to Rx) • Non-blocking split screen (Notes) • Real time Care coordination - Patient instructions filled in --> printed out • Epic access logs to profile pre/post work • Epic logs to profile patient complexity • Voice recognition used only 2 visits) • Dual windows allows e.g., working in Notes without blocking other functions Scheduling • Web links available outside EHR • History documentation interface is structured	• Dual monitor present in ~35% of visits • CPRS functions (Notes, Orders etc) takes up full screen, blocking other functions (even on dual monitor PCs). • Associations for Consults and Imaging but not DX • Real time Care coordination (patient status) not in CPRS but available elsewhere • Computerized clinical reminder work • Order imaging has more mouse clicks • No separate history documentation UI - only notes

Table 20. Site comparison (UCSD versus VASD) in terms of care delivery model, staff support, and EHR features.

6.6.3 Correlates of TLX Effort

(a) The TLX effort and performance subscales were not correlated. That is consistent with the characterization of physicians as expert users. Although some measures of activity, including mouse clicks, visit length, size of the Epic log, and gaze dominance ratio is moderately correlated with the effort subscale. These results are consistent with habitual expert behavior, as physicians quickly and easily completed complex actions requiring long sequences of mouse clicks and keystrokes. **(b)** Physician co-investigators on the project team and some of the interviews suggested that patient difficulty and complexity might influence perceived workloads more than physical activity or constraints of the EHR. Patients with complex medical needs, mental status issues, or with requests for disability evaluations, might be more challenging, even if less work is required with the EHR. Unfortunately, our data collection methods limit our ability to easily extract a complete clinical picture of the patient and high-level conceptual tasks (such as processing a disability request). **(c)** Physician interviews suggested that perceived effort and frustration may be a function of information not available in the EHR rather than limitations of interface itself. High-level summaries, support for clerical input, easier ways to communicate and document information were mentioned. The need to avoid interruption while working on a clinical problem was noted and congruent with other work. **(d)** Although further work - perhaps via more in-depth analysis of QUICK data - will be needed to better understand the nature of the relationship between activity and perceived workload, these preliminary results suggest a novel approach for understanding the role of EHR design in impact on clinical work. As experienced users who were able to effectively complete complex tasks on EHRs despite inefficient interfaces, QUICK physicians seemed more concerned about the complexity of patients than about the design of the tools. Including clinical complexity in future studies examining challenges associated with EHR might provide greater insight into the factors contributing to frustration and perceived difficulties.

6.7 Conclusions

We demonstrated the ability to capture and analyze clinical activities during outpatient visits in exam rooms at two sites with different EHRs in primary and specialty care. We further compared highly time-resolved behavior patterns with physicians' self-reported measures of task and cognitive load. Previous efforts in this area have explored the use of observations, video, workload assessments, and time-motion measurements. Unlike time-motion studies based on fixed, pre-existing codes, we captured activity on replayable media, aiding replicability of findings and further allow subsequent reanalysis and open coding.

Expert users have skill sets and workarounds to adapt to task load factors, such as time pressure (e.g. some work may be streamlined) or changes in team structure (e.g. some individuals pick up additional tasks outside the visit window). EHR log analysis may fill in the gaps in visit and patient context that are not observable during the window of the visit.

Some differences in activity patterns between Epic and CPRS have a straightforward interpretation. For example, smaller fraction of mouse clicks on Epic Notes versus CPRS notes may be explained by Epic enabling notes to be active and visible to users while they are working on other functions.

Analysis of interviews provide context to understand these patterns. For example, staff support levels are greater at UCSD, where nursing staff carry out more EHR tasks such as order entry. That may potentially explain, beyond the difference in scheduled visit lengths, why UCSD visits tend to be several minutes shorter on average than at VASD, even when accounting for covariates like new/established patient and physician specialty.

Preliminary workload analyses suggest a potentially complex relationship between levels of measurable EHR activity and perceptions of effort and task performance. Further work will be needed to better understand the complex interplay between extrinsic characteristics of patients' medical and psychological needs and resulting levels of EHR activity and perceptions of workload.

6.8 Significance

It' s clear from our data, that multiple factors such as visit complexity, patient factors, and technology (EHR) can influence clinical work, and variations in physicians' personalized work practices and perceived workload. Variation and clustering of activity across providers also suggests that the role of the EHR might be relatively less significant than hypothesized. A socio-technical approach that address the multiple factors identified is needed to improve our ability to integrate technology in a complex environment of clinical work. Usability and principles of user centered design need to go beyond simple user studies often conducted on a handful of subjects in a non-clinical environment. Perhaps ongoing evaluation of the technology and how humans integrate these technologies into daily workflow are more useful.

We recognize that such efforts can be expensive and perhaps even disruptive to care if done without concern. However, the emerging emphasis on efficiency, improved workflow and change management / quality improvement activities in many healthcare systems provides the opportunity to study and improve human-computer interaction. In terms of automating human coding of such interactions, we expect that emerging technologies like usability tracking software integrated with web-based EHR systems, ubiquitous sensors, and increasing use of enterprise data logging would complement the methods described here in future studies with larger sample sizes and diverse clinical settings. Complementing these techniques, future studies should include both contextual issues particularly patient diagnoses and social needs and higher-level abstractions of key tasks to more directly link observations of EHR use to meaningful clinical tasks.

Implications for EHR redesign physicians recognize the importance of having an EHR, the need to have information stored electronically where it can be retrieved for review and clinical decision making. However physicians are frustrated with our current EHR systems, as they lack finesse, or "easability of use." In addition, the use of multimodal data entry such as voice and natural language commands provide opportunities to reduce the burdensome use of mouse and keyboard driven UIs. Additionally, there are opportunities to reduce redundant workflows and legacy workflows and documentation practices that were carried over as we transitioned from a paper based medical record to an electronic format. These aren't purely technical challenges, but perhaps clinical and administrative practices need to evolve to keep pace with technologies.

7. List of publications and products

1. Calvitti, A., Hochheiser, H. S., Ashfaq, S.,et al. Physician Activity During Outpatient Visits and Subjective Workload. Journal of Biomedical Informatics, accepted for publication; 2017 Feb.
2. Zhang, J., Avery, K., Chen, Y., et al. A Preliminary Study on EHR-Associated Extra Workload. In Proceedings of American Medical Informatics Association, Annual Symposium; 2015 Nov 14-18; San Francisco, USA.
3. Ashfaq, S., Avery, K., Agha, Z., et al. Analysis of Computerized Clinical Reminder Activity and Usability Issues. In Proceedings of American Medical Informatics Association, Annual Symposium; 2015 Nov 14-18; San Francisco, USA.
4. Calvitti, A., Weibel, N., Hochheiser, H. S., et al. Can eye tracking and EHR mouse activity tell us when clinicians are overloaded? Human Factors Quarterly, Veteran Health Administration. 2014 Sep.
5. Weibel, N., Rick, S., Emmenegger, C., et al. LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office. Personal and Ubiquitous Computing. 2015 Feb 1;19(2):317-34.
6. Weibel, N., Ashfaq, S., Calvitti, A., et al. Multimodal data analysis and visualization to study the usage of Electronic Health Records. In Proceedings of 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 May 5. (pp. 282-283). IEEE.
7. Rick, S., Street, R. L., Calvitti, A., et al. Understanding Patient Physician Communication and Turn-taking Patterns with Directional Microphone Arrays. In Proceedings of International Conference on Communication in Healthcare; 2015 Oct; New Orleans, USA.
8. Rick, S., Calvitti, A., Agha, Z., et al. Eyes on the clinic: Accelerating meaningful interface analysis through unobtrusive eye tracking. In Proceedings of 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2015 May 20 (pp. 213-216). IEEE.

Bibliography

1. Weibel, N., Rick, S., Emmenegger, C., et al. LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office. Personal and Ubiquitous Computing. 2015 Feb 1;19(2):317-34.
2. Quan, H., Sundararajan, V., Halfon, P., et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data.Med Care. 2005 Nov;43(11):1130-9.
3. Patton, M. Qualitative Research & Evaluation Methods: Integrating Theory and Practice 4th Edition ed. Thousand Oaks, CA: Sage; 2015.
4. Weir, CR., Nebeker, JJ., Hicken, BL., et al. A cognitive task analysis of information management strategies in a computerized provider order entry environment. Journal of the American Medical Informatics Association: JAMIA. 2007;14(1):65-75.