

Master's thesis proposal [rev_4.5]

Richard Oldham, DDS

University of Pittsburgh
Department of Biomedical Informatics

September 19, 2012

Abstract

The benefits of Computerized Patient Records (CPRs) are well documented. Yet current dental CPRs inadequately support chairside use as evinced in low adoption rates for operatory computers. If dental professionals have any hope of realizing the advantages CPRs offer, information systems must adapt to their workflow and not *vice versa*. To that end, research suggests tablet devices hold promise as an alternative to traditional desktop PCs' ergonomic shortcomings in the clinical environment. Designing a goal and task specific tablet-based treatment planning interface may help bridge the adoption gap between office and chairside computer use in dentistry. Evaluation can determine if users are faster, more successful, less error prone, or more subjectively satisfied with the tablet application over traditional CPRs.

Contents

1	Background	3
1.1	Adoption Gap	3
1.2	Usability	4
1.3	Attitudes and Use	6
1.4	User Model	7
1.5	Conclusions and Solutions	10
2	Tablets and Usability	11
3	Proposed Research	13
3.1	Treatment Planning Prototype	13
3.2	Evaluation	15

3.2.1	Users	17
3.2.2	Instruments	17
3.2.3	Power Analysis	18
4	Proposed Timeline	21
	Appendices	26
A	CPR Screenshots	26
A.1	Dentrix	26
A.2	Dentrix user mouse tracking data	27
A.3	EagleSoft	28
A.4	Screen controls	29
A.5	Tablet prototype	30
B	Artifact Model	31
C	Physical Models	32
C.1	Rear-delivery	32
C.2	Side-delivery	33
C.3	Chair-delivery	34
D	Power Analysis	35

List of Tables

1	Office personnel participation (percentage) in data entry tasks by information category. Data from telephone survey of 88 dental practices [1].	9
2	Summary comparison of differences across medicinal and dental domains.	12
3	Proposed two part within-subjects evaluation design using a tablet-based prototype and conventional CPR for comparison	16

1 Background

First, I summarize applicable findings from usability and informatics research. Next, I discuss a user model for treatment planning in dentistry. Lastly, I discuss conclusions and potential solutions.

1.1 Adoption Gap

The adoption of Computerized Patient Records (CPRs) by healthcare professionals has long been recognized as a key step in improving patient care [2–6]. Dentists recognized the power of computers assisting with practice administration — such as scheduling and billing — becoming early adopters of computers in-office (Figure 1) [1].

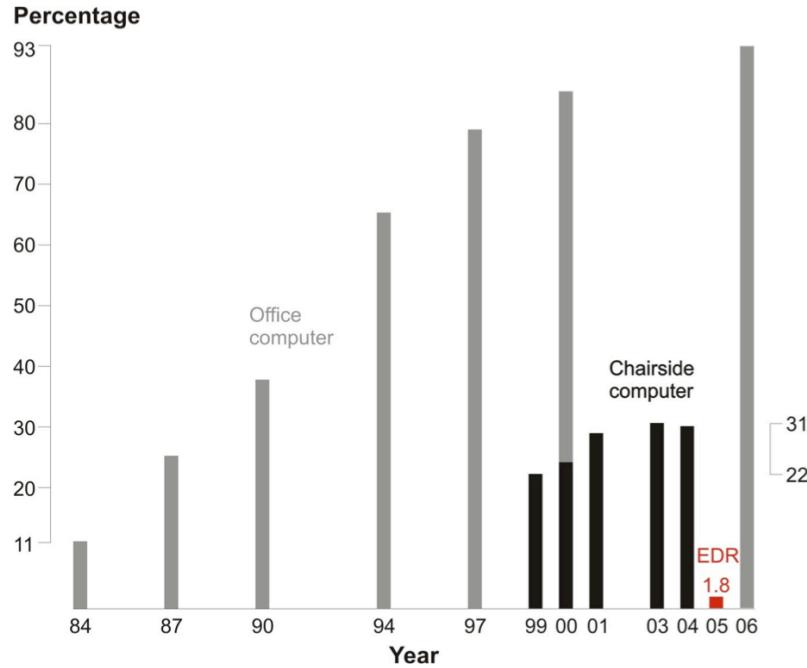


Figure 1: Year of first computer installation for office and chairside use by US dentists. The red bar represents completely paperless practices.

Survey data from 2010, the most recent available, reported 92% of respondents using Practice Management Software (PMS) [7].[†] However, the conclusions of two usability studies on four dental CPRs, which together comprise more than a 70% market share for such systems, suggest poor usability is a barrier to dental CPR adoption for chairside use [8, 9]. The gap between office adoption of computers and the adoption of computers for chairside use is evident in Figure 1.

1.2 Usability

The first of the two cited usability studies employed heuristic evaluation methods [10, 11]. Three expert reviewers examined each of the four CPRs and tabulated violations of Nielsen's heuristics [12]. Evaluators found 229 violations in total across EagleSoft, Dentrix, PracticeWorks, and SoftDent. Consistency and Standards, Match Between System and the Real World (Appendix A.4), and Error Prevention were the most commonly violated heuristics with the distribution of violations roughly equal among the four CPRs inspected. The authors admit their conclusions offered no comparative assessment between the CPRs but the magnitude of violations suggested significant room for usability improvement.

The second usability study compared [13] these same market-leading CPRs by assigning common tasks to five novice users for each of the four systems ($n = 20$). Users were purposively sampled with four full-time dental faculty members, eight practicing dentists, and eight senior dental students from the University of Pittsburgh School of Dental Medicine comprising subjects for the assessment. Two experts coded task completion success and associated usability problems. The authors used Wilcoxon rank-sum tests, Spearman correlation coefficient tests, and Kruskal-Wallis tests to determine the correlation between usability problems and task outcomes. The frequency of observed usability problems correlated positively with the frequency of task failures for all tasks except two ($p < 0.05$): assigning a tooth as missing and recording soft tissue measurements. This is a surprising result because these two tasks are often difficult to complete. On average, users correctly completed 43% of the assigned tasks leading to a similar conclusion: the most commonly used dental CPRs suffered from a steep learning curve and have ample room for usability improvement.

The results of these two studies point to inadequate support for chairside

[†]Here I use PMS to refer to software that handles office administration whereas a CPR handles patients' clinical data. Most CPR software is also PMS but not the other way around.

use. All four systems included in these studies began life as PMS with clinical functions added later. The abundance of nested menus, tree selections, and drop down lists for clinical and non-clinical functions alike are evidence of this (Appendix A.1). Dentrix, for instance, provides sometimes two or three different input methods for the same function on the same screen (button on a pallet, tree selection, and drop down menu). All four systems evaluated by Thyvalikakath et al. employ screen designs that include nearly all input options at once rather than the options needed for the specific task or goal (Appendix A.3). Interfaces with many options can make finding the function of interest more difficult, increase cognitive load, and slow the user [14]. One could argue complicated interfaces are not universally poor if such an interface is appropriate for the job (for example, an airplane cockpit). But given the evidence, it seems the gulf of execution and evaluation [15] in dental CPRs is unnecessarily wide. Dentrix and Eaglesoft screen design, for example, confused novice users with intermixed controls for entering existing conditions and planned procedures on the same palette. And performing basic tasks, such as marking a tooth as missing, require four to six mouse clicks across a wide swath of screen real estate (Appendix A.2).

Fitts's law [16, 17] predicts the average task completion time, T , as primarily a function of D , the distance from the starting point to the center of the target and W , the width of the target measured along the axis of motion such that $T = a + b \log_2 \left(1 + \frac{D}{W}\right)$ where a and b are experimentally determined constants. Though not measured by Thyvalikakath et al. (2008), one might expect poor task completion time for the surveyed CPRs in light of mousetracking data that illustrated long distances and narrow targets (Appendix A.2). Compounded over the course of an initial exam appointment, this method of data input is time prohibitive for a single user. Meaning, these input design choices are far from efficient for a user that requires efficiency.

Since the information input and retrieval processes were not integrated into the concepts of tasks and workflow, users were required to divide their attention between navigation and task completion. In both usability studies, the authors noted information that would typically appear on a single paper form was spread out over two or more screens in the CPR. The comparative evaluation cited difficulty with tasks that required screen switching as a principle source of user frustration. Evidence suggests these problems are not unique to dental health information technology (HIT) applications [18, 19].

Readers of these two usability studies offered valid criticisms [20]. One

such critique centered on the authors' choice of novice users for the evaluation. Critics pointed to the low likelihood of a user being completely unfamiliar with the CPR in a live office setting. Training office staff to use the CPR is routine and thus the steep learning curve cited in the conclusion may not reflect the learning curve faced by actual users. Another critique, from the same citation, questioned the assumption of causality between poor usability and low adoption rates of CPRs by dentists. The author sees private practice dentists' primary concern as one of return on their investment in Information Technology over potential usability issues discovered post purchase. Regarding heuristic evaluation, Nielson and Molich (1990) developed this discount inspection method at a time when it was expensive to gain access to users. Nielson and Molich warn readers of haphazard application of their method and the need for calibrating expert reviewers. Heuristic inspection can determine the extent to which an interface adheres to a set of best practices; to make broader claims about generalizability of results would be problematic.

1.3 Attitudes and Use

In a telephone survey of 102 general dentists in the United States, researchers discovered two main sources of user frustration in using dental CPRs: steep learning curves and difficulty entering clinical data [1]. Respondents considered 'better input methods,' 'smaller computers,' and 'better user interface design' as principal areas for improvement. Another important concern included lack of infection control inherent with traditional keyboard and mouse input. These responses track closely with those elicited from British users through a similar survey [21].

Dental practices use their computers primarily for scheduling patients and tracking completed procedures (Figure 2). Because dental CPRs are poorly designed for chairside use, some users accommodate by working around the CPR. Some practices record temporary notes on a patient napkin or tray cover and commit these notes to the electronic record later while others record histories on paper and treatments on the computer. While transcribing notes from paper to the computer is not an inherent problem, such a practice has the potential for error. Storing some information in paper format and some in electronic format (Figure 2), similarly, is not *prima facie* poor practice but requires the consumer of the information to switch between the paper-based and computer-based records. Fragmentation of data into multiple locations has one inevitable drawback: switching requires cognitive overhead and time. This burden can slow the consumption of in-

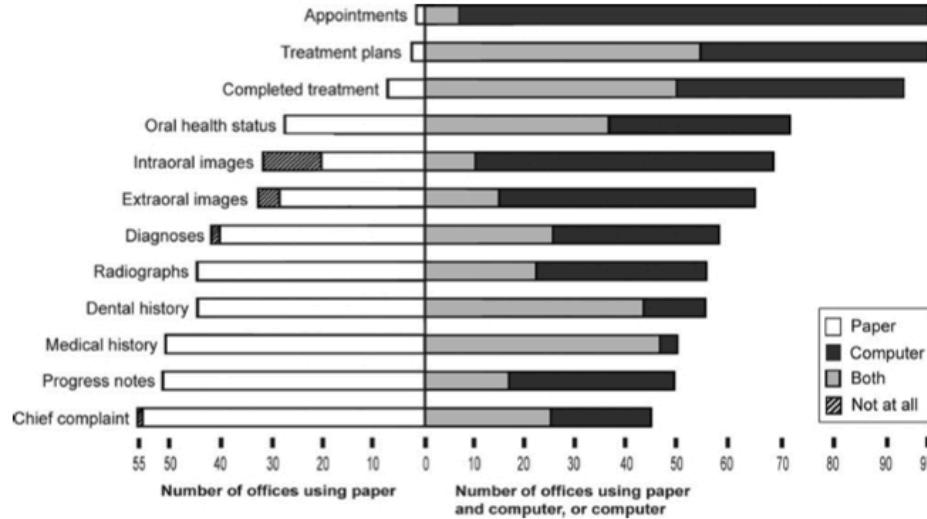


Figure 2: Storage of major clinical information categories on paper/computer, sorted by utilization of computer-based storage in descending order [22].

formation by the clinician or produce transcription errors [23]. The need to work around the CPR is due, in some part, to the difficulty of integrating the necessary hardware into an already cramped dental operatory [24, 25] and, in some part, to the time-prohibitive cost of entering data into a poorly designed system. The ergonomics of one dentist, one assistant, and one patient gathering, entering, retrieving, and presenting information are complex (Appendix C). The suggested areas of improvement noted in the Schleyer, et al. (2006) survey highlight the frustration of using software designed without these ergonomics in mind. The task of finding the required information inside the patient record is often criticized as time-intensive and distracting [26]. This is especially concerning given the increased information needs of clinicians in an increasingly time-pressed environment. If we accept the premise that HIT can improve patient care, and there exists a large adoption gap, the necessity of generating and understanding a user model becomes apparent.

1.4 User Model

Recognizing the interplay between usability, practice culture, and operatory configuration, researchers set out to model the workflow of an initial exami-

nation and treatment planning appointment through contextual inquiry [27]. During this appointment the dentist documents the patients current condition, the patient-specific goal for oral health, and generates a sequenced plan to proceed from present state to the goal state. This process, with its findings, diagnoses, and resulting plan, must then be discussed with the patient. Depending on the complexity of problems, multiple treatment plans may be required. Of any appointment, this one generates the most clinical artifacts such as radiographs, photographs, hard and soft tissue diagnostic charting, and study models. It follows that the initial exam requires the most data input of these artifacts into the CPR. And artifacts must be quickly retrieved for consumption by the clinician and for discussion with the patient. Following the initial visit and generation of a treatment plan, each subsequent visit requires comparatively little interaction with the CPR. The dentist may check the medical history or reference the planned procedures, for example. Thus the initial appointment requires the most interaction with the CPR for data entry and retrieval, making it an ideal candidate for study.

A thorough initial exam and treatment planning appointment is vital for not only diagnosing every present condition but communicating the condition, its consequences, and treatment options as well. Generating a treatment plan is analogous to peeling an onion: the dentist starts with the most obvious signs and symptoms and continues examination, testing, and imaging until all signs and symptoms have been mapped to a problem list. The dentist may ascribe diagnoses to items on the problem list and then generate a treatment to address each item. This process is non-deterministic, across patients and dentists, and time-intensive, often requiring two people. Typically one operator will diagnose and call out findings while an auxiliary transcribes the findings into the computerized record. While this task may be handled differently in sequence from one dentist or patient to another, the steps are the same and mappable to common set of terms and abstractions.

Irwin, et al. (2009) observed an auxiliary may not always be available and dental team members used a variety of workarounds in such a situation. For example, hygienists often perform recare examinations without an assistant, noting their preliminary findings on a piece of paper prior to discussion with the dentist. Some clinicians postpone any data entry tasks to a time after the patient has been dismissed. This presents the problem of having to remember what was done and why. Even under ideal circumstances, that is, when an assistant was present, over 60% of the Irwin's observed breakdowns were due to technology. One implicit conclusion from the Irwin model was dentists spend little time entering data with the majority of CPR interaction handled by hygienists and assistants. Schleyer, et al. (2006) reported

dentists participating in only 23% of data entry tasks, chief among them being progress notes. Hygienists and assistants handled most other data entry tasks such as diagnoses, oral health status, and histories. Given the fact non-dentists perform about 75% of data input, learnability of a CPR becomes an important concern. Data input categories and frequency by office role are summarized in Table 1. Whether dentists would employ more clinical functions in their CPR given improved interfaces, as Irwin, et al (2009) described, “with the flexibility in sequencing, granularity and comprehensiveness” for data entry, remains to be seen. But the conclusions from Irwin, et al (2009) mirror the conclusions from the Schleyer, et al. (2006) survey: CPR design must accommodate the demands of dental workflow.

Table 1: Office personnel participation (percentage) in data entry tasks by information category. Data from telephone survey of 88 dental practices [1].

Information Category	% of Individuals Participating in Data Entry			
	Dentists	Hygienists	Assistants	Front Desk
Progress notes	40	31	25	5
Extraoral images	36	26	35	2
Intraoral images	35	27	36	2
Diagnoses	32	29	32	7
Treatment plan	27	25	31	17
Chief complaint	23	23	32	22
Completed treatment	23	28	31	19
Oral health status	21	37	37	6
Dental history	20	30	40	11
Radiographs	16	34	48	2
Medical history	13	21	29	36
Appointments	6	27	24	43
Total	23	28	33	16

1.5 Conclusions and Solutions

Several important conclusions can be gleaned from the cited literature. First, dental operatories are small and require careful planning to integrate the CPR's supporting hardware. The suggestion of using smaller computers from the Schleyer, et al. (2006) survey reveals user frustration with this integration burden. Next, the current paradigm in dental CPR design and ergonomics requires an auxiliary for efficient data entry. Most dentists are consumers of patient information while hygienists and assistants perform most data entry. Following the generation of a treatment plan, subsequent appointments require comparatively little data entry or retrieval. Given the results of the user model study, the highest value targets for CPR improvement are the functions used during the initial visit: namely, diagnoses and treatment planning.

Schleyer et al. (2006, 2007) and Irwin et al. (2009) report the common practice of using multiple or duplicate data entry. A user may write something on paper, such as routing slips or preliminary findings, later transcribing this information into the electronic record. Alternatively, one might record the medical history on paper and use the CPR to record planned and completed treatments. This is due, in part, to the fact that dental CPRs cannot capture the variety of data with the flexibility of paper forms [22]. Depending on the survey of paper forms and the CPR used for comparison, CPRs can lack 80–90% of the data elements contained in paper records. Duplicate data entry might also be used to accommodate the non-linear and variable process by which patients are escorted through this initial visit. Therefore, an alternative system needs to allow the user a temporary and mutable data store to enter information before committing the entry to the record. And the clinical vocabulary of the information system must be at least as large as that of the clinician. Efforts are underway to enumerate and control this vocabulary [28–31] but providing the most commonly used data elements and allowing the user to customize the type and sequence of information capture may suffice.

Little research effort has been dedicated to studying mobile devices in dentistry. While many dental-specific mobile applications exist and several vendors are exploring mobile platforms, most literature are case reports. For over a decade, dental clinicians have recognized the potential for hand-held computing devices to address the limitations of traditional PCs [32,33]. Coupling the advantages of mature tablet devices with a novel interface, which incorporates lessons learned though previous research, may hold promise.

2 Tablets and Usability

Tablets offer several advantages over traditional desktop PCs in clinical use. Because CPR access is not universally available in medical exam rooms, tablets' smaller size and portability were central to their adoption by physicians. Tablets allow the clinician to enter orders, notes, and review information while remaining mobile. Similarly, tablets' smaller footprint and portability could address the ergonomic challenges in dental practice. A tablet is ideal for consuming information, which is the principle concern of the clinician. Tablets can at least match the traditional desktop PC in data entry speed and accuracy with use [34,35]. They can be easily shared between the provider and patient and easily disinfected [36]. Properly designed touch-screen mobile interfaces have demonstrated promise in a variety of medical settings [37–41]. Finally, if users are as capable on a mobile device as a traditional PC, the ergonomic and portability advantages can increase user subjective satisfaction with a mobile system over a conventional desktop [42].

Tablets make ideal devices for which to design interfaces for fast task completion time. Because their screen size is small and users make input choices with their fingers, buttons must be large. Most tablet devices will not discriminate capacitive touch with less than 8–9mm surface area. Further, because tablet screen real estate is limited, one must carefully consider the interface elements and sequence of interactions. This economy of choice requires a tablet-based interface to be well adapted to user tasks and goals. Thus, given a larger button size on a smaller screen, Fitts's law would suggest tablets as a logical platform choice.

Some remain skeptical of the virtues of tablet computers in clinical use [43,44] and, compared to the medical literature, few investigators have studied tablet use by the dental team [45]. Applying the insights from medical informatics research to dental problems presumes exchangeability of the domains. This may not be the case (Table 2). A key difference to note is the small number of diagnoses in dentistry compared to medicine; dentists cover a smaller portion of the body compared to physicians. But for each diagnoses there are a larger number of choices for treatment. One might explain this difference by pointing to the larger role patient-provider preferences play in dental treatment planning [46] or the smaller role of evidence-based practice in dentistry compared to medicine [47]. Another key difference is ergonomic: physicians stand or sit bedside while the dental ergonomic is unique. Because dental patients are usually less sick, the tablet interface needs fewer input controls. And because the number of diagnoses is small and therapies per diagnoses large, several rule-based functions could help

Table 2: Summary comparison of differences across medicinal and dental domains.

Feature of practice	Domain	
	Dentistry	Medicine
Patient status	Usually not sick	Usually sick
No. of encounters per day	Few	Many
Average visit length	Long	Short
No. of diagnoses	Few	Many
No. of therapies per diagnoses	Many	Few
No. of initial visit artifacts	Many	Few

generate treatments from problems saving time. Given the success enjoyed by tablets in complex medical settings, I remain optimistic about their performance in the dental operatory. Creating a usable and useful interface for the tablet is a separate consideration.

Thyvalikakath, et al. (2008) recognized the disconnect between task flow and screen design in dental CPRs ahead of a formal modeling study. Several conclusions from the 2008 comparative usability study regarding software design:

- Users should be able to identify their treatment preferences and software should support that approach.
- Data entry and retrieval controls available on screen should correspond with the tasks to be completed. Unrelated or extraneous controls should not be shown. Information that belongs to a specific task context should be shown together or be easily accessible. For instance, hard-tissue and periodontal findings often must be reviewed together to make a clinical decision and should not be separated unnecessarily.
- Data entry and functional controls should be organized and labeled clearly according to input class (findings, planned/completed procedures, notes, pictures, annotations, etc.).

Lastly, the User-Centered Design (UCD) approach to software engineering has demonstrated promise in general [48–51] and HIT specifically [52]. Thyvalikakath, et al. (2007) recognized UCD as a potential solution to the multiple heuristic violations noted during expert review. UCD is a broad term to describe design processes in which end-users influence the software design. Linking the iterative design process to feedback from users has a proven track record of producing highly usable software. I elaborate on the proposed incorporation of the UCD approach in Section 3.2.

3 Proposed Research

Based on research evidence, tablet devices make a logical hardware choice. However, hypotheses about tablet use in dentistry remain untested. The extent to which tablet devices and touch interfaces can improve data entry and user satisfaction is a question left to evaluation. But first an interactive tablet-based treatment planning prototype is needed.

Research is currently underway to qualify and quantify the data private practice dentists capture in their CPRs. Given the time constraints of clinical practice and poor usability demonstrated by market-leading CPRs, we hypothesize dentists predominantly document completed procedures and few diagnostic findings (Figure 2). This is accepted practice by third-party payers in contrast to medicine. Given the lack of open-domain knowledge of type and quantity of data users store, design choices must rely on literature and best practices. But once completed, the data from this study can help inform system design — ranked list UI elements, for example.

I propose testing two hypotheses: (1) tablet computers, using platform-native UIs, can at least match conventional desktop computers in data entry speed and task completion; (2) users will prefer using a mobile touch screen device over the conventional desktop PC.

3.1 Treatment Planning Prototype

To effectively evaluate the comparative usability of tablets and desktop PCs a measure of functional equivalence across the two systems is required. That is to say the tablet prototype must be able to input and retrieve the same clinical artifacts necessary for treatment planning as a mature software system like Dentrix or EagleSoft. Such artifacts include:

- Patient histories: both medical and dental histories are critical to patient safety and disease management

- Hard tissue: both existing restorations and conditions that would indicate treatment
- Soft tissue: periodontal measures such as clinical attachment level, soft tissue margins, and common periodontal findings (e.g. furcation location and classification)
- Picture Archival and Communication System: radiographs as well as intraoral and extraoral photographs play an increasingly important role in dental practice. Such artifacts provide not only valuable information to the clinician for decision-making but also aid in communicating that information in a patient-friendly manner.

These are the same class of items required to work through case examples in the seminal textbook on dental treatment planning by Sefanic and Nesbit [53]. Further these artifacts are the same ones used by non-digital dentists (Appendix B) and correspond with the task and user models generated by Irwin et al., thus offering a consensus for high value targets for data input and retrieval. Assuming the four mentioned artifacts are necessary and sufficient to complete a dental treatment plan, and given that such functionality is already present in current CPR vendor offerings, what is the value proposition?

From survey data we know that dental practices value digital functionality related to financial reimbursement. Poor usability demonstrated in evaluation offers a plausible explanation for this and may help explain the practice of keeping some artifacts digital and others on paper. Why would a clinician bother with the many clicks and screen shuffling required to document all of their findings in the digital record if they are paid only for the completed treatments? Indeed many clinicians either use planned treatments as their primary method of documentation or use paper records as a workaround.

If a system were designed with goals and tasks in mind, however, it is feasible that a small number of rules could unambiguously and programmatically render diagnoses from problems—generating knowledge from data and saving the operator valuable chair-time in the process. These rule-based functions are a necessary first step to improving patient care by changing the work flow process from “what treatment should address the problem?” to “why is there pathology present?” Doing so allows a potentially valuable hook into decision making process with which to provide clinical decision support.

In order for the prototype to be interactive, The functional specification includes:

- The ability to document the most common diagnostic findings in a clinically linear fashion (find all sites affected with diagnostic condition, move on to the next condition)
- The ability to automatically generate diagnoses from the documented findings
- The ability to automatically map a list of treatments, based on provider preferences, from the list of problems and diagnoses
- The ability to bin and sequence the list of treatments to form one or several treatment plans

Following the generation of a treatment plan (or plans), the provider next discusses the findings with the patient. The ability to summarize findings for the providers use in treatment planning is as important as communicating these findings to the patient. Using the treatment plan “patient view” to facilitate this discussion could help drive treatment plan acceptance. This feature would undoubtedly be beneficial but may fall outside the scope of this project. The “patient view” function is but one example of the many functions available to users once patient data is present in the CPR. However, addressing the problem of cumbersome data entry is a necessary prerequisite to reaping the benefits of computerized records.

3.2 Evaluation

Dsesigning a novel tablet interface lends itself well to the evaluation process. Were I to evaluate only the interfaces, there would no advantage to use of a smaller device. Subjects could just as easily evaluate the novel interface on a desktop PC alongside a conventional CPR. Therefore, I propose a two stage evaluation using fractional factorial design. For the vector of participants of length n , each participant p is tasked with entering one of two comparably difficult patient cases $c_{1,2}$ into the conventional CPR and tablet system. If Bob were tasked with entering c_1 into Dentrix, he would then enter c_2 into the tablet system. For the next participant, Alice, the case assignment switches. The design is fractional factorial because not every combination of cases and methods will be administered for each subject in order to avoid learning effects from entering the same patient case twice. This experiment

design is well suited to comparing the relative efficiency of data entry across the tablet and conventional CPR because of within-subjects comparison. Each user must use both methods of data input and thus acts as their own control when comparing to an alternative. After gathering the data, I can examine the relative impact of case selection, person, and data input method on task completion time and error rate. Significant effort will be made to select two patient cases of equal difficulty and comprehensiveness in order to minimize the case effect on variance of task completion time and error rate.

A second evaluation, using the same design but different setting, can determine the relative ergonomic advantages to a tablet device. Subjects will sit in a dental operatory and receive paper artifacts to enter each system. Using video recording, I can measure subject motion over time and determine the relative efficiency of each device in the physical environment.

Table 3: Proposed two part within-subjects evaluation design using a tablet-based prototype and conventional CPR for comparison

Evaluation	Users	Setting	Design
Usability test	Naïve with 30 minutes of training	Lab-based	Users sit at a desk in ‘clean room’ enter one task set per system; tests interface with time
Function study	Naïve with 30 minutes of training	Clinical	Users sit in a clinical environment enter one task set per system; tests interface & ergonomic with time & motion

Evaluation of the tablet application requires comparing it to an alternative. Two logical points of comparison are the most popular CPRs in private practice: Dentrix and EagleSoft. Vendors have granted the CDI access to these CPRs and using one of them for the purposes of evaluation should present no challenges.

As previously described, the UCD approach to development is ideally suited to the proposed research. The University of Pittsburgh’s Center for Dental Informatics and School of Dental Medicine can provide ready access to potential participants for design feedback ahead of the formal evaluation.

This feedback could take the form of expert heuristic inspection or small-scale usability tests. The purpose of this feedback is to confirm the design specification and implementation. Depending on the results of this feedback, one or more rounds of iteration may be necessary.

3.2.1 Users

Prior to a usability evaluation, I propose vetting the prototype design with a small group of experts (committee members, CDI faculty, etc.). This expert review could take the form of heuristic inspection or a think aloud. The purpose of this feedback is to find low-hanging fruit for usability improvement ahead of the comparative evaluations.

Because of the small tablet's small size and portability it could be used by anyone involved with patient care. In private practice, the majority of data input is performed by assistants, hygienists, and front office staff (Table 1). Therefore, the sample of users should include potential users most likely to be performing data entry tasks in the operatory: dental assistants, dental hygienists, and dentists in decreasing order of frequency of data input by category.

3.2.2 Instruments

(a) *Task completion success*

Logical summative measure for overall usability measured by determination of user completing the task within a specified time range (say 10 minutes), regardless of accuracy of the input.

(b) *Task completion time, if successful*

Useful metric because of quantitative nature allowing statistical modeling of the relative effects of the method, case, and subject on the task completion time. Useful for comparing data input speed as measured with a stop watch by observer.

(c) *Subjective user satisfaction*

One of the only means of measuring user preferences across data input methods. Measured through survey of users with a validated instrument, of which there are many [54–59].

(d) *Error rate*

Also quantitative measure, if secondary to task completion time because users are expected to finish the data input tasks but not expected to

commit errors. Useful for gauging effectiveness or quality of user input. Measured through match of paper artifacts to electronic records as determined by observer with domain expertise.

3.2.3 Power Analysis

Little consensus exists on the appropriate sample size for usability evaluations [60–63]. While some researchers in the HCI domain acknowledge parameter settings for a power analysis are little more than a best guess, we concluded an inspection of the literature and effort towards sample size determination would be worthwhile. The experiment design makes use of quantitative measurements and because of this we can use generalized linear models to simulate the data and determine power accordingly. Thyvalikakath et al. (2008) argue:

“We chose five participants per group because this number is considered adequate for usability testing as evidenced in the usability engineering literature [13].”

Nielsen offered no guidance on sample size in the cited article, however.

Virzi (1992), Nielsen and Landauer (1993), and Lewis (1994) have published influential articles on the topic of sample size in usability testing. In these articles, the authors presented a mathematical model of problem discovery rates in usability testing. These models used the Poisson distribution of rare events to simulate problem discovery rates. The authors determined it was feasible to determine the sample size needed to uncover a given proportion of problems in an interface during one test and presented empirical evidence for the models. Some important claims from these papers:

- Most usability problems are detected with the first three to five subjects.
- Running additional subjects during the same test is unlikely to reveal new information.
- Return on investment (ROI) in usability testing is maximized when testing with small groups using an iterative test-and-design methodology.

Nielsen and Landauer (1993) extended Virzi’s (1992) original findings and reported case studies that supported their claims for needing only small samples for usability tests. They and Lewis (1994) identified important assumptions about the use of the formula for estimating problem discovery

rates; among them, prior knowledge of error rates and distribution being the most useful.

After discussion with Dr. Roger Day, we arrived at two methods for determining the needed sample size: simulation and heuristics. To ensure adequate statistical power, I include both (code included in Appendix D). While it is plain to see the numbers used in the analysis were made up, the important determination is not the values used but the sensitivity of the result to changes in assumption. Other assumptions, by way of parameter settings, are detailed below:

- Test statistic: for task completion time and error rates the sample means from each method are used to determine significance
- Sources of variance in task completion time: built into the model are assumptions about case, method, person, and measurement (or unsystematic) variance
- Model distributions: I use the normal distribution, centered at the task completion mean for each group and the Poisson distribution with lambda equal to mean error rates for each group
- Effect size: the measured difference in task completion time and error rate across methods

After varying the parameters and running the experiment 2,000 times for each setting, I determined the required effect size for statistically significance in values of $n < 40$ and power ≥ 0.80 is between 15–17%. This translates into a two to three minute time savings for a mean task completion time of 17 minutes.

Performing a two-way within-subjects ANOVA on the simulated data yields:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Method	1	107.45	107.45	15.507	0.00118 **
Case	1	6.22	6.22	0.898	0.35737
as.factor(Person)	17	194.22	11.42	1.649	0.16177
Residuals	16	110.87	6.93		
<hr/>					
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

The summary of ANOVA shows the greatest source of variance is the method followed by the person, which is consistent with the model parameters. Using the power method from the **NCStats** package, which allows the online

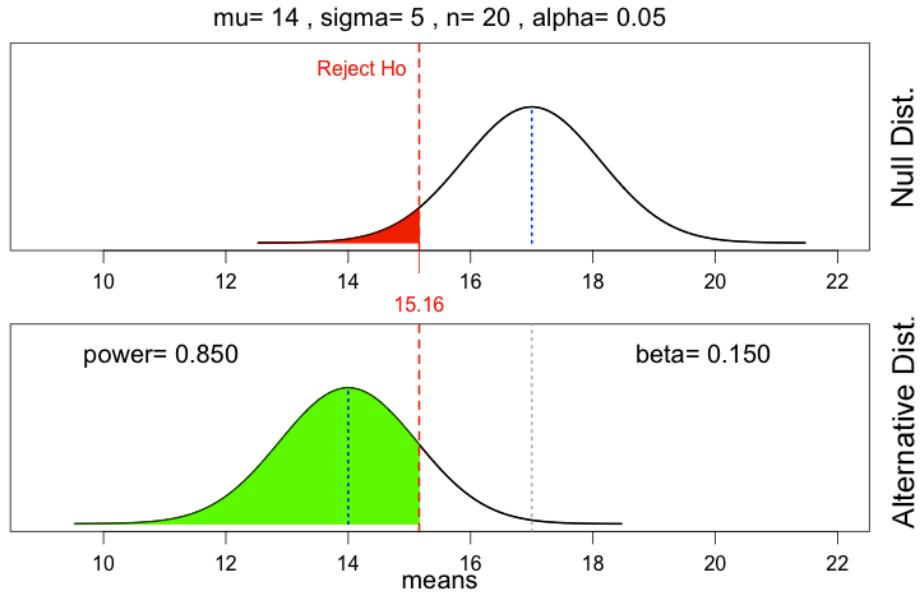


Figure 3: Plot of null and alternative distributions; the rejection region and the power region are highlighted [64].

adjustment of model parameters, we can see a value of $18 \leq n \leq 20$ will provide power above 0.80 for liberal assumptions of variance (Figure 3). An analysis similar to Turner (2006) could be performed on error rates. Given the even smaller amount of *a priori* knowledge of expected error rates and distribution, the use of Poisson sampling for such an analysis would be less powerful than the results of task completion simulation.

4 Proposed Timeline

<p>April Initial proposal May Begin development</p> <p>June Begin contextual inquiry</p> <p>August Complete contextual inquiry August Model building</p> <p>September Screen design and wireframe revision</p> <p>September Complete preliminary expert review</p> <p>September Complete proposal October Begin principal development</p> <p>October Second round of expert review</p> <p>December Third round of expert review</p> <p>January Complete interactive prototype</p>	<ul style="list-style-type: none">• April Begin thesis••••• August Begin sketches and paper prototypes• September Preliminary expert review•• September Complete functional specification•• November Begin design iteration•• December Design polish•• January Begin evaluation
---	--

Because I will be living in Virginia after January 2013 it is hard to predict the evaluation schedule. I am committed to regular trips back to Pittsburgh to conduct expert reviews and the comparative evaluations. In Virginia, I have access to many private practices — some of whom may be interested in participating in the evaluation. My aim is to complete the evaluation by the spring and complete the thesis by the summer.

Acknowledgements I would like to thank Roger Day, Gerry Douglas, Harry Hochheiser, and Titus Schleyer for their invaluable questions and insights. This research made possible by the National Library of Medicine through grant 2T15LM007059-26.

References

- [1] TK Schleyer, TP Thyvalikakath, H Spallek, MH Torres-Urquidy, P Hernandez, and J Yuhaniak. Clinical computing in general dentistry. *J Am Med Inform Assoc*, 13:344–52, 2006.
- [2] J Chasteen. A computer database approach for dental practice. *J Am Dent Assoc*, 123:26–33, 1992.
- [3] J Eisner, J Chasteen, T Schleyer, and C Feldman. *The computer based oral health record: A new foundation for oral health information systems*. American Fund for Oral Health, Chicago, IL, 1993.
- [4] TG Thompson and DJ Brailer. The decade of health information technology: Delivering consumer-centric and information-rich health care. U.S. Department of Health and Human Services, Office of the National Coordinator for Health Information Technology, 2004.
- [5] R Spicer. Bytes and bites—using computerized clinical records to improve patient safety in general dental practice. *Dent Update*, 35(9):618–9, Nov 2008.
- [6] T Schleyer, U Mattsson, R Ní Ríordáin, V Brailo, M Glick, RB Zain, and M Jontell. Advancing oral medicine through informatics and information technology: a proposed framework and strategy. *Oral Dis*, 17 Suppl 1:85–94, Apr 2011.
- [7] N Levine. Tech experts chart a digital path. Technical report, Dental Products Report, September 2010.
- [8] TP Thyvalikakath, TK Schleyer, and V Monaco. Heuristic evaluation of clinical functions in four practice management systems. *J Am Dent Assoc*, 138:208–18, 2007.
- [9] TP Thyvalikakath, V Monaco, HB Thambuganipalle, and TK Schleyer. A usability evaluation of four commercial dental computer-based patient record systems. *J Am Dent Assoc*, 139:1632–42, 2008.
- [10] J Nielsen and RL Mack. *Executive summary*, pages 1–24. Usability inspection methods. John Wiley & Sons, New York, 1994.
- [11] TK Schleyer. Assessing software usability through heuristic evaluation. *J Am Dent Assoc*, 138(2):211–12, Feb 2007.
- [12] TP Thyvalikakath, V Monaco, HB Thambuganipalle, and TK Schleyer. Comparative study of heuristic evaluation and usability testing methods. *Stud Health Technol Inform*, 143:322–27, 2009.
- [13] J Nielsen. Proceedings of the SIGCHI conference on Human factors in computing systems: “Celebrating Interdependence”. CHI ’94—Boston, Massachusetts, USA, pages 152–158, New York, NY, April 1994. ACM.
- [14] H Qian, R Kuber, and A Sears. Towards developing perceivable tactile feedback for mobile devices. *Int J Hum-Comput St*, 69:705–19, Jul 2011.
- [15] DA Norman. *The design of everyday things*. MIT, 1989.
- [16] PM Fitts and JR Peterson. Information capacity of discrete motor responses. *J Exp Psychol*, 67(2):103–12, Feb 1964.
- [17] PM Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *J Exp Psychol*, 47(6):262–69, June 1954.

- [18] J Ash, M Berg, and E Coiera. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc*, 11(2):104–12, 2004.
- [19] AF Rose, JL Schnipper, ER Park, EG Poon, Q Li, and B Middleton. Using qualitative studies to improve the usability of an emr. *J Biomed Inform*, 38(1):51–60, 2005.
- [20] T Cockerell. Patient Records, Letter to the Editor. *J Am Dent Assoc*, 140(4):402–4, April 2009.
- [21] J John, D Thomas, and D Richards. Questionnaire survey on the use of computers in dental practices across the thames valley region. *Br Dent J*, 195:585–590, 2003.
- [22] TK Schleyer, H Spallek, and P Hernandez. A qualitative investigation of the content of dental paper-based and computer-based patient record formats. *J Am Med Inform Assoc*, 14(4):515–26, Jul-Aug 2007.
- [23] DD Salvucci, NA Taatgen, and PB Jelmer. Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09–Boston, MA, pages 1819–28, New York, NY, 2009. ACM.
- [24] TK Schleyer. Why integration is key for dental office technology. *J Am Dent Assoc*, 135:4S–10S, October 2004.
- [25] M Unthank. Designing your office for technology. *J Am Dent Assoc*, 135:24S–29S, Oct 2004.
- [26] E Nygren, JC Wyatt, and P Wright. Helping clinicians to find data and avoid delays. *The Lancet*, 352(9138):1462–6, October 1998.
- [27] JY Irwin, MH Torres-Urquidy, TK Schleyer, and V Monaco. A preliminary model of work during initial examination and treatment planning appointments. *Br Dent J*, 206(1):E1, Jan 10 2009.
- [28] E Kalenderian, RL Ramoni, JM White, ME Schoonheim-Klein, PC Stark, NS Kimmes, GG Zeller, and MF Walji. The development of a dental diagnostic terminology. *J Dent Educ*, 75(1):68–76, Jan 2011.
- [29] JM White, E Kalenderian, PC Stark, RL Ramoni, R Vaderhobil, and MF Walji. Evaluating a dental diagnostic terminology in an electronic health record. *J Dent Educ*, 75(5):605–15, May 2011.
- [30] A Acharya, DP Mital, and TK Schleyer. Electronic dental record information model. *Int J Med Eng Inform*, 1(4):418–34, 2009.
- [31] L Emmott. Electronic dental records in dentistry. *J Am Coll Dent*, 77(1):10–12, 2010 Winter.
- [32] MH Taylor. Handheld computing in dentistry. *Dent Clin North Am*, 46(3):549–51, Jul 2002.
- [33] M Jablow. Your practice in the palm of your hand. *J N J Dent Assoc*, 74(1):8–9, 2003 Winter.
- [34] J Kirby and AL Rector. The pen & pad data entry system: from prototype to practical system. *Proc Annu Symp Am Med Inform Assoc*, page 709, 1996.
- [35] S Mackenzi and W Soukoreff. Text entry for mobile computing: models and methods, theory and practice. *Human-Computer Interaction*, 17(2 and 3):147–98, 2002.

- [36] D Mayrhofer. Pen-based electronic medical record on a tablet pc in a medical practice. Masters thesis, Vienna University of Technology, 2007.
- [37] G Haller, DM Haller, DS Courvoisier, and C Lovis. Handheld vs. laptop computers for electronic data collection in clinical research: a crossover randomized trial. *J Am Med Inform Assoc*, 16(5):651–59, Sep-Oct 2009.
- [38] EM van Mulligen, H Stam, and AM Van Ginneken. Clinical data entry. *Proc Annu Symp Am Med Inform Assoc*, page 81, 1998.
- [39] DC Baumgart. Personal digital assistants in health care: experienced clinicians in the palm of your hand? *Lancet*, 366:1210–22, Oct 2005.
- [40] YC Lu, Y Xiao, A Sears, and JA Jacko. A review and a framework of handheld computer adoption in healthcare. *Int J Med Inform*, 74(5):409–22, Apr 2005.
- [41] N Seneviratne and B Plimmer. Improving stylus interaction for emedical forms. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, OZCHI ’10, pages 280–87, New York, NY, 2010. ACM.
- [42] E Cole, ED Pisano, GJ Clary, D Zeng, M Koomen, CM Kuzmiak, BK Seo, Y Lee, and D Pavic. A comparative study of mobile electronic data entry systems for clinical trials data collectio. *Int J Med Inform*, 75(10-11):722–29, 2006.
- [43] T Kaneshige. iPad in Healthcare: Not So Fast, Nov 2011.
- [44] W Mashman. The iPad in cardiology: Tool or toy? *JACC Cardiovasc Interv*, 4(2):258–9, Feb 2011.
- [45] E Frank. iPad—tool or toy? *Int J Comput Dent*, 13(4):341–9, 2010.
- [46] EJ Kay, NM Nuttall, and R Knill-Jones. Restorative treatment thresholds and agreement in treatment decision-making. *Community Dent Oral Epidemiol*, 20(5):265–8, 1992.
- [47] M Tellez, SL Gray, and S Gray. Sealants and dental caries: Dentists’ persepectives on evidence-based recommendations. *J Am Dent Assoc*, 142:1033–40, Sept 2011.
- [48] LJ Bannon. Issues in design: Some notes. In DA Norman and SW Draper, editors, *User centered dsystem design: New perspectives on human-computer interaction*, pages 25–30. Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.
- [49] D Wixon and C Wilson. The usability engineering framework for product design and evaluation. In M Helander and T Landauer, editors, *Handbook of Human-Computer Interaction*, pages 653–88. Elsevier, Amsterdam, 1997.
- [50] JD Gould. How to design usable systems. In M Helander and T Landauer, editors, *Handbook of Human-Computer Interaction*, pages 231–54. Elsevier, Amsterdam, 2nd edition, 1997.
- [51] C Abras, D Maloney-Krichmar, and J Preece. *Encyclopeda of Human-Computer Interaction*. User-Centered Design. Sage Publications, Thousand Oaks, 2004.
- [52] CM Johnson, TR Johnson, and J Zhang. A user-centered framework for redesigning health care interfaces. *J Biomed Inform*, 38:75–87, 2005.
- [53] SJ Stefanac and SP Nesbit. *Treatment Planning in Dentistry*. Mosby, St. Louis, 2nd edition, 2006.

- [54] JP Chin, VA Diehl, and KL Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '88, pages 213–18, New York, NY, 1988. ACM.
- [55] DF Sittig, GJ Kuperman, and J Fiskio. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. *Proc Annu Symp Am Med Inform Assoc*, 1999.
- [56] BH Wixom and PA Todd. A theoretical integration of user satisfaction and technology acceptance. *Inform Syst Res*, 16(1):85–102, March 2005.
- [57] SH Han, KJ Kim, MH Yun, SW Hong, and J Kim. Identifying mobile phone design features critical to user satisfaction. *Hum Factor Ergon Man*, 14(1):15–29, 2004.
- [58] WJ Kettinger and CC Lee. Perceived service quality and user satisfaction with the information services function. *Decision Sci*, 25(5-6):737–66, Sept 1994.
- [59] AM Aladwani and PC Palvia. Developing and validating an instrument for measuring user-perceived web quality. *Inform Mgmt*, 39, 2002.
- [60] RA Virzi. Refining the test phase of usability evaluation: How many subjects is enough? *Hum Factors*, 34(4):457–468, Aug 1992.
- [61] JR Lewis. Sample sizes for usability studies: additional considerations. *Hum Factors*, 36(2):368–78, Jun 1994.
- [62] CW Turner, JR Lewis, and J Nielsen. Determining usability test sample size. In W Karwowski, editor, *International Encyclopeida of Ergonomics and Human Factors*, volume 3, pages 3084–88, Boca Raton, FL, 2006. CRC Press.
- [63] J Nielsen and T Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, CHI '93, pages 206–213, New York, NY, 1993. ACM.
- [64] DH Ogle. Simulation of power calculation in 1-sample mean problems.
RForge: <http://www.rforge.net/doc/packages/NCStats/powerSim.html>, September 2012.

A CPR Screenshots

A.1 Dentrix

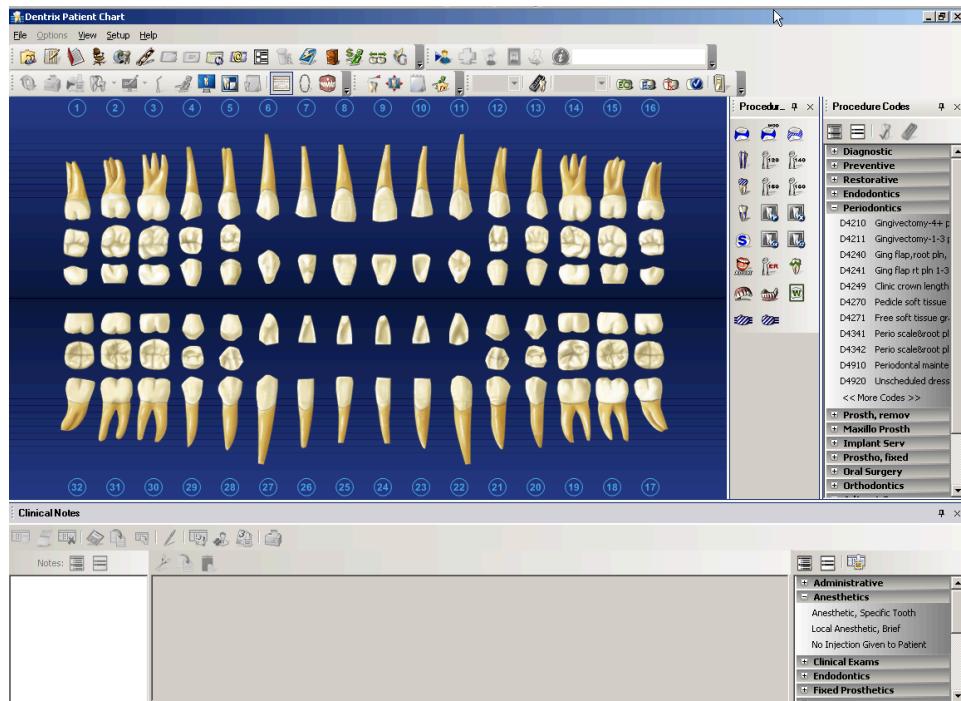


Figure 4: Screenshot of Dentrix's G4 user interface.

A.2 Dentrix user mouse tracking data

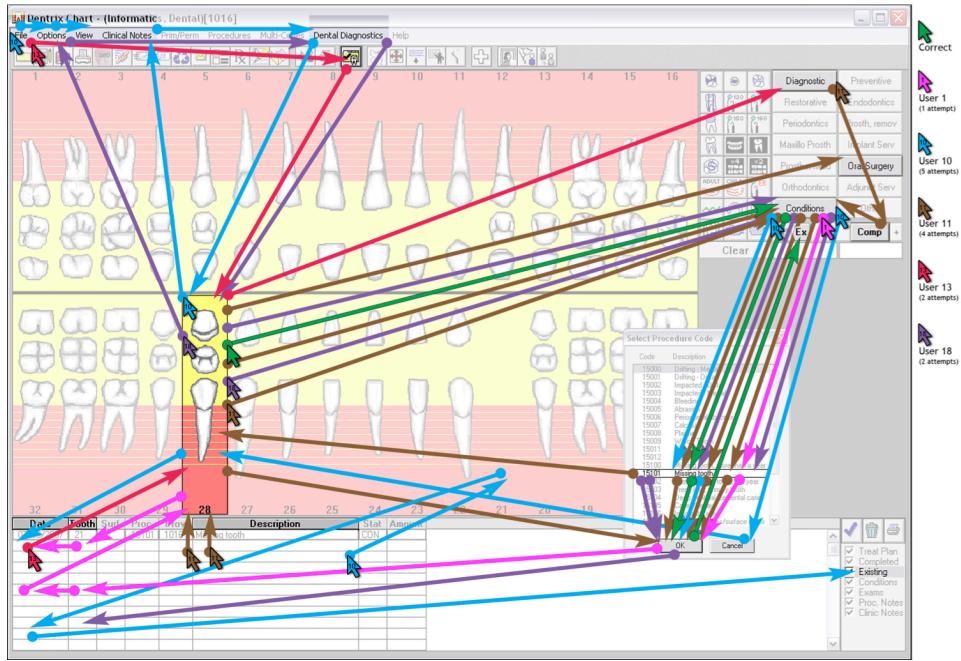


Figure 5: Screenshot showing a user's actions while trying to record tooth no. 28 as missing while using the Dentrix v10.0.36.0. The correct sequence consists of the following left mouse clicks: “tooth no. 28,” “Conditions,” “Missing tooth,” “OK” and “EX” [9].

A.3 EagleSoft

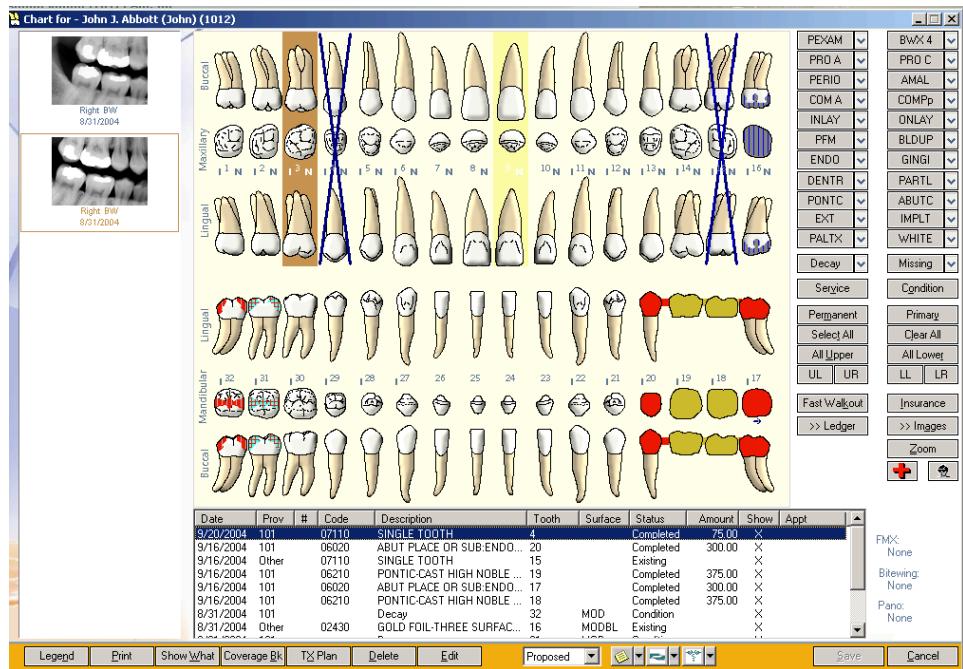


Figure 6: Screenshot of EagleSoft's v16.0 user interface.

A.4 Screen controls

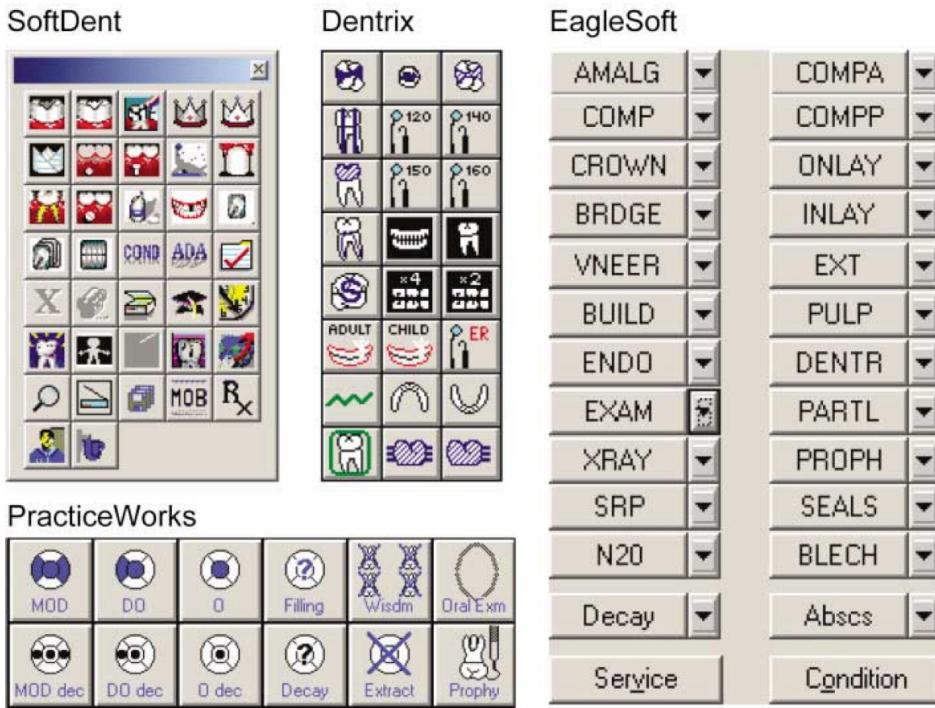


Figure 7: Charting icons for hard-tissue findings and procedures in Dentrix v 10.0.36.0, EagleSoft v 10.0, SoftDent v 10.0.2, and PracticeWorks v 5.0.2 [8]. Expert review found heuristics such as Consistency and Standards and Match Between System and the Real World were often violated by hard tissue controls. One expert found PracticeWorks icons to be ambiguous. One common theme was the lack of reliance on operating system convention to help the user with recognition rather than recall.

A.5 Tablet prototype

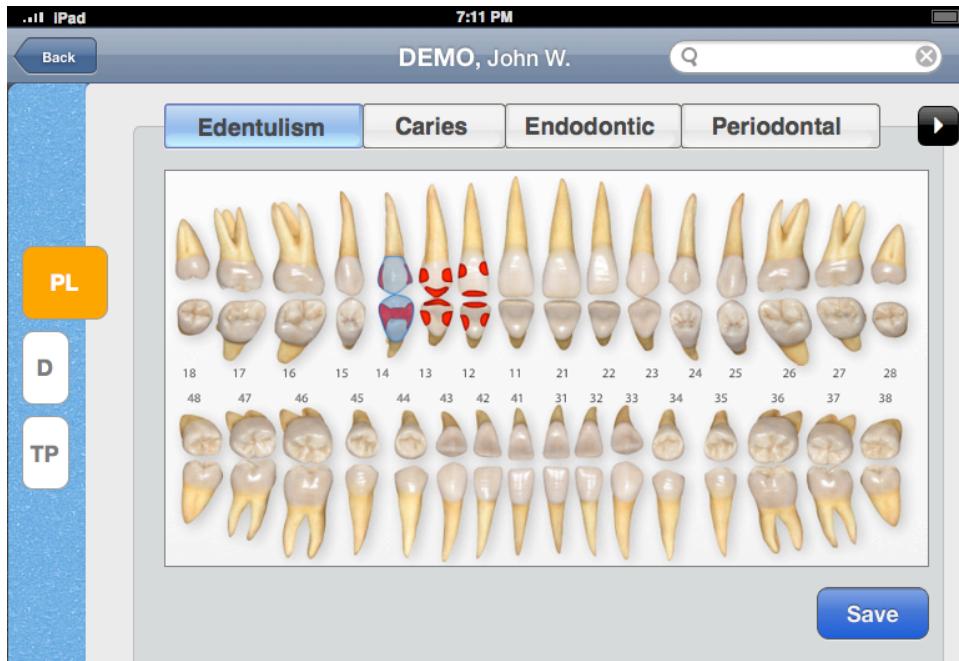


Figure 8: Mockup for tablet-based prototype. Much of the design inspiration comes from the two usability studies of dental CPRs. Button controls large enough for fingers and taken from the `StdUILibrary` within Xcode. Many other conventions like tap to select and swipe-based navigation also leverage user familiarity with iOS devices. Navigation along the top is couched in the concept of the initial exam and allows users to customize the content and sequence of data capture. And the ability to programmatically render diagnoses and treatments based on problems identified allow more efficient data entry.

B Artifact Model

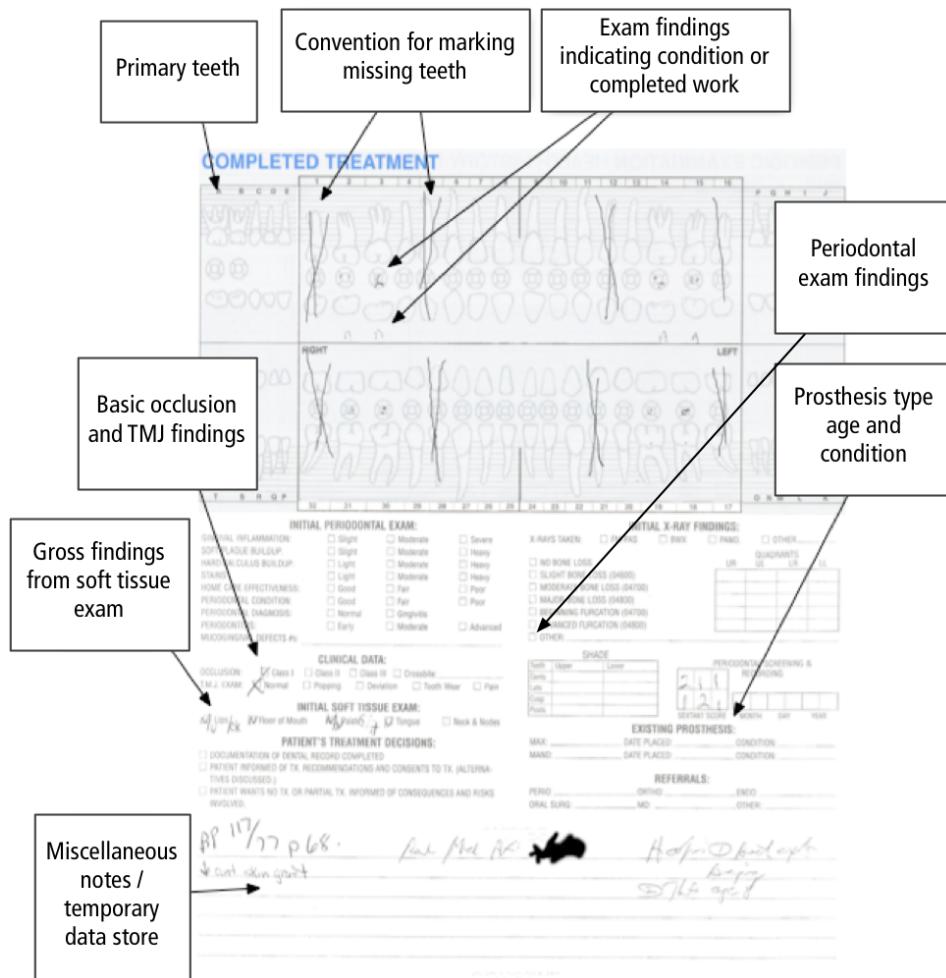


Figure 9: Artifact model generated from common paper-based treatment planning forms

C Physical Models

C.1 Rear-delivery

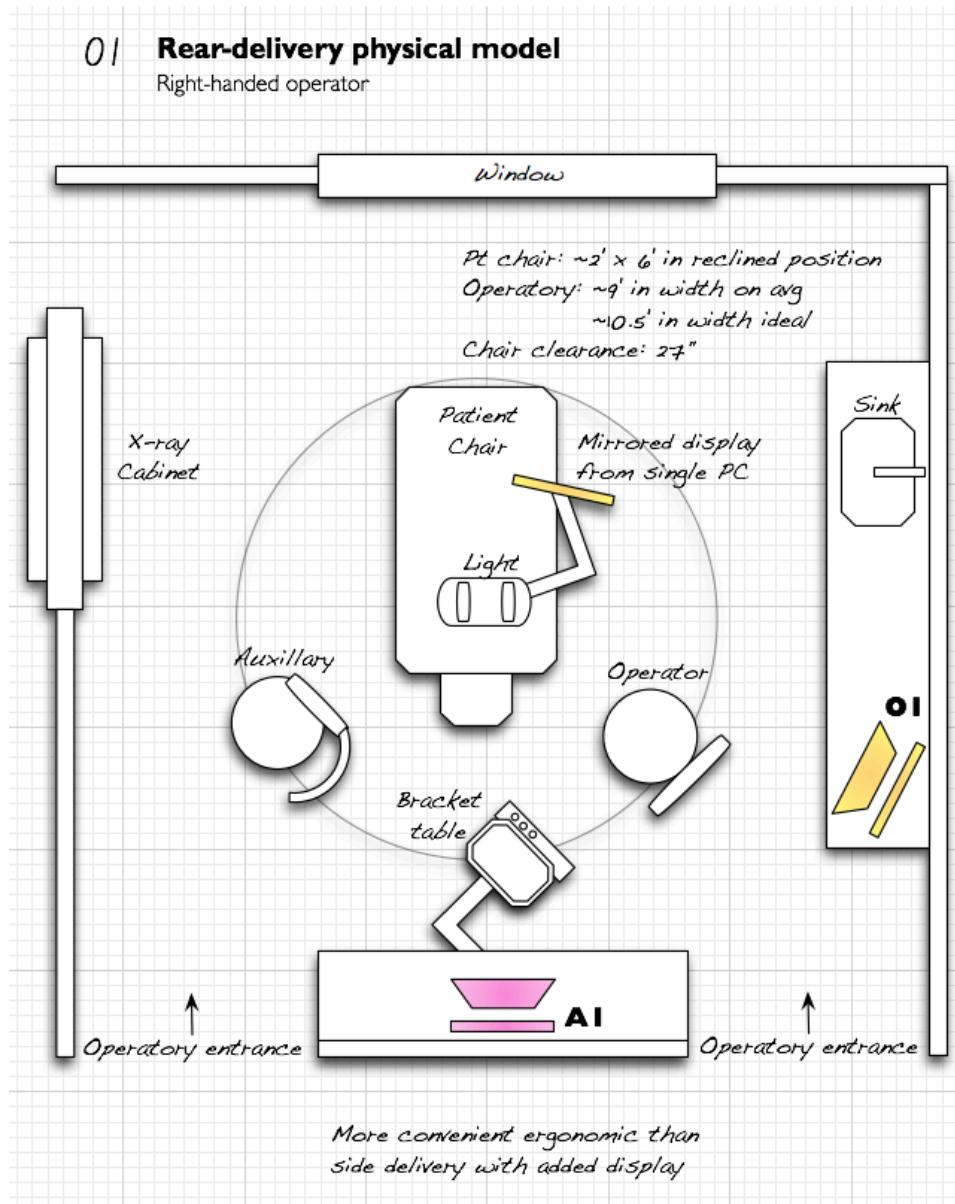


Figure 10: Rear-delivery physical model for a right-handed operator.

C.2 Side-delivery

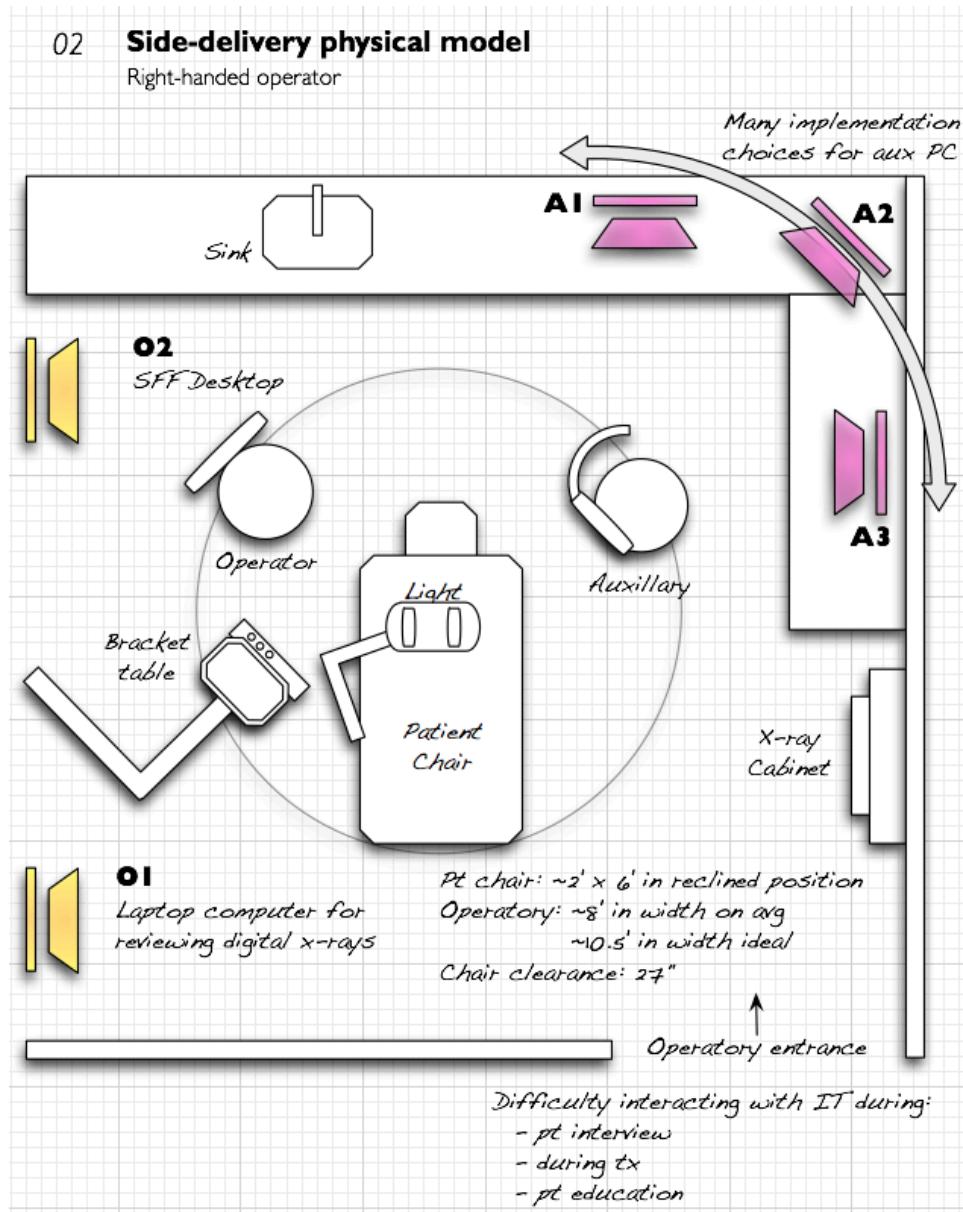


Figure 11: Side-delivery physical model for a right-handed operator.

C.3 Chair-delivery

03 Chair-delivery physical model

Right-handed operator

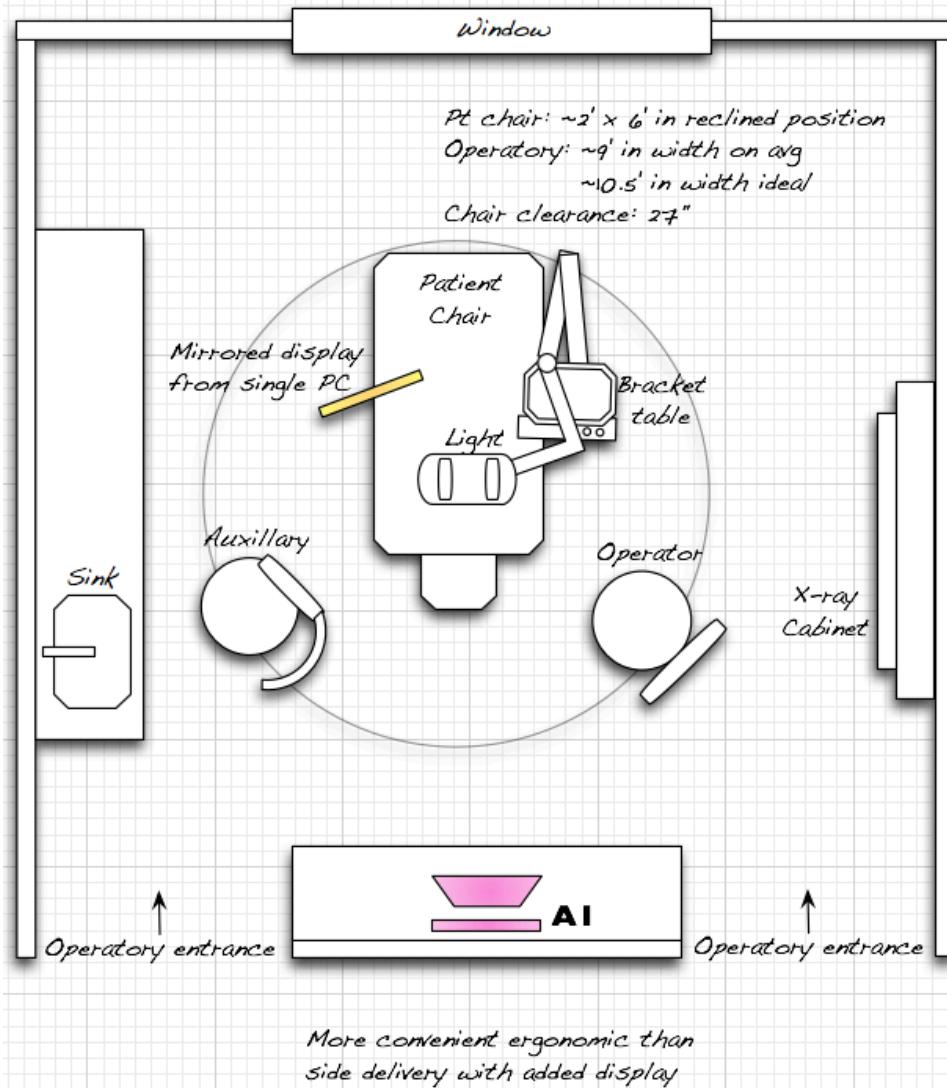


Figure 12: Chair-delivery physical model for a right-handed operator.

D Power Analysis

```

# Begin model parameters
Nsubjects = 18
# Mean task completion time for H0 and HA
method_effect= c(A=18, B=15)
# Lambda mean error rates for H0 and HA;
lambda = c(A=1, B=0.5)

# Begin sources of variance
method_sd = c(A=5, B=4)
person_effect = rnorm(Nsubjects, 0, 4)
# KEY ASSUMPTION: variance due to case?
case_effect = c(1,0)
# All other sources of variance
unsystematic_sd = 3
# End model parameters

## Construct data frame(person, case, method); no explicit use of as.factor
for subjects and case due to needing to select only a fraction of factorial
(tempOK).
temp = expand.grid(1:Nsubjects, 1:2, c("A","B"))
names(temp) = c("Person","Case","Method")

# Select fraction of total comb
tempOK = temp$Person + temp$Case + match(temp$Method, unique(temp$Method))
temp = temp[ (tempOK %% 2)==0 , ]

# Generate model task completion times
temp$Time = case_effect[temp$Case] + person_effect[temp$Person] +
  method_effect[temp$Method] + rnorm(1, 0, method_sd[temp$Method]) +
  rnorm(nrow(temp), 0, unsystematic_sd)

# Generate error rates
temp$Error = rpois(n=Nsubjects, lambda=lambda)

# Clean up and assign temp to dframe
dframe = temp[with(temp, order(Person, Method)), ]
row.names(dframe) = 1:nrow(temp)

# Simulate experiment, looping over dframe H0 and HA xNREPS
NREPS = 2000
Nsubjects = 18

pValues = sapply(1:NREPS, function(ignoreMe) {
  dframe = expand.grid(as.factor(1:2), c("A","B"), as.factor(1:Nsubjects))
  names(dframe) = c("Case", "Method", "Person")
  dframeOK = as.numeric(dframe$Person) + as.numeric(dframe$Case) +
    as.numeric(match(dframe$Method, unique(dframe$Method)))
}

```

```

dframe = dframe[ (dframe$OK %% 2)==0 , ]
dframe$Time = case_effect[dframe$Case] + person_effect[dframe$Person] +
  method_effect[dframe$Method] + rnorm(1, 0, method_sd[dframe$Method]) +
  rnorm(nrow(dframe), 0, unsystematic_sd)
model <- glm(Time ~ Method + Case + as.factor(Person), data=dframe)
summary(model)$coef["MethodB", "Pr(>|t|)"]
}
)
summary(pValues)
mean(pValues < 0.05)

#powerSim viz
require(NCStats)
powerSim(mu0=method_effect['A'], s.mua=method_effect['B'], s.sigma=5,
  s.n=Nsubjects, s.alpha=0.05, lower.tail=TRUE)

```