

ABSTRACT

Title of dissertation: INTERACTIVE VISUAL DISPLAYS FOR RESULTS
MANAGEMENT IN COMPLEX MEDICAL WORKFLOWS

Sureyya Tarkan, Doctor of Philosophy, 2013

Directed by: Professor Ben Shneiderman
Department of Computer Science

Clinicians manage medical orders to ensure that the results are returned promptly to the correct physician and followed up on time. Delays in results management occur frequently, physically harm patients, and often cause malpractice litigation.

Better tracking of medical orders that showed progress and indicated delays, could result in improved care, better safety, and reduced clinician effort. This dissertation presents novel displays of *rich tables* with an interaction technique called *ARCs (Actions for Rapid Completion)*. Rich tables are generated by *MStart (Multi-Step Task Analyzing, Reporting, and Tracking)* from a workflow model that defines order processes. Rich tables help clinicians perceive each order's status, prioritize the critical ones, and act on results in a timely fashion. A second contribution is the design of an interactive visualization called *MSProVis (Multi-Step Process Visualization)*, which is composed of several *PCDs (Process Completion Diagrams)* that show the number and duration of in-time, late, and not-completed orders. With MSProVis, managers perform retrospective analyses to make decisions by studying an overview of the order process, durations of order steps, and performances of individuals.

I visited seven hospitals and clinics to define sample results management workflows. Iterative design reviews with clinicians, designers, and researchers led to refinements of the rich tables, ARCs, and design guidelines. A controlled experiment with 18 participants under time pressure and distractions tested two features (showing pending orders and prioritizing by lateness) of rich tables. These changes statistically significantly reduce the time from nine to one minute to correctly identify late orders compared to the traditional chronologically-ordered lists. Another study demonstrated that ARCs speed performance up by 25% compared to state-of-the-art systems. A usability study with two clinicians and five novices showed that participants were able to understand MProVis and efficiently perform representative tasks. Two subjective preference surveys suggested new design choices for the PCDs.

This dissertation provides designers of results management systems with clear guidance about showing pending results and prioritizing by lateness, and tested strategies for performing retrospective analyses. It also offers detailed design guidelines for results management, tables, and integrated actions on tables that speed performance for common tasks.

INTERACTIVE VISUAL DISPLAYS FOR RESULTS MANAGEMENT IN
COMPLEX MEDICAL WORKFLOWS

by

Sureyya Tarkan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:

Professor Ben Shneiderman, Chair
Associate Professor Linda Aldoory
Professor Ben Bederson
Professor Rance Cleaveland
Dr. Catherine Plaisant

© Copyright by
Sureyya Tarkan
2013

To my family,
For making all of this possible.

Acknowledgments

I am grateful to all the people who have helped me in completing my Ph.D. studies and dissertation.

First and foremost I would like to thank my advisor, Ben Shneiderman, for giving me an opportunity to work on challenging and interesting projects over the past three years. He has been very approachable for help and advice and easily available for discussions, and I have always felt that he has my best interest in mind. It has been a pleasure to work with and learn from such an outstanding individual. He is the reason why I became a better researcher and person. He has taught me the importance of setting high standards for my work. He has helped me put structure in my ideas and work as a whole. He has been an invaluable resource as an individual and a professional. His enthusiasm and encouragement constantly inspires people around him. He is a great professor with creativity, depth and breadth of knowledge, and vision. I cannot imagine a better advisor to work with.

I would also like to thank Catherine Plaisant. Without her extraordinary ideas and strict criticisms, this thesis would have been a distant dream. From her, I have learned many user interface design principles and lessons about how to work with users to solve their real problems. The work in this dissertation could not have been completed without her invaluable feedback and suggestions.

Thanks are due to Rance Cleaveland, Ben Bederson and Linda Aldoory for agreeing to serve on my thesis committee and for sparing their time reviewing the manuscript. I am also thankful to Mike Hicks for his guidance and encouragements in times of uncertainty and despair. I have learned from all of you, and it has been

a pleasure working with you.

I owe special thanks to Kostas Pantazos and Lyndsey Franklin. Kostas helped me with the iterative development and testing of MSProVis. I appreciate his determination to push this idea further for a publication. While I was away for a summer internship, Lyndsey updated the design of my controlled experiment by running pilot studies. Thanks to her hard work, I was able to finish the full experiment rapidly.

My friends and fellow human-computer interaction lab members have enriched my graduate life in many ways and deserve a special mention. I will never forget our brown bag lunches, annual symposiums, practice talks, discussions around the cubicles, and other activities. I am thankful for having friends whose emotional support has provided me with the comfort of a social environment. I am fortunate to have close friends from the Computer Science department, Darya Filippova, Awalin Shopan, Nicholas Chen, Vanina Martinez, Alexandros Tzannes, Florin Chelaru, Shashvat Thakor, and Anand Bahety; my other friends and colleagues, Eva Sitaridi, Helmut Veith, Susanne Glissmann, Sarah Knoop, and Adriana Malheiro; my friends for 15 years, Hanife Demir, Yasemin Rencher, Kamil Erbek, and Samet Demir; and many more.

I would like to acknowledge the financial support from the Office of the National Coordinator for Health Information Technology Grant No. 10510592 for Patient-Centered Cognitive Support under the Strategic Health IT Advanced Research Projects Program (SHARP). I appreciate the participation of the project members, Jorge Herskovic, Elmer Bernstam, Chitra Shriram, Todd Johnson, and Daniel Murphy. I thank A. Zach Hettinger from the Washington Hospital and MedStar Institute for Innovation for his feedback on the early prototypes and his collaboration on the papers. In addition, the medical insights, encouragement, and partnership of Seth M. Powsner (Yale School of Medicine) were greatly helpful. Many thanks to the reviewers of the design guidelines: Stephen Few, Jeff Belden, and Janey Barnes. Without their comments, the design guidelines would

not have been useful.

I am grateful to my family – Mom, Hilal, Nihal, Alparslan, Dad – for encouraging me to pursue the career I had always desired and supporting me unconditionally throughout this process. I am especially thankful to my mom, Didar Tarkan – without you, this journey would have not been possible or worthwhile.

Lastly, I thank God for giving me the strength to fight all the evil that has befallen on me. In grad school, He taught me there are too many innately bad people, who purposefully mislead, prevent others' dreams, and cause harm in life. Whenever I prayed to God, He has helped me move far away from the bad, surround myself with the good ones (who cherish me), and work really hard to accomplish my goals.

Contents

Acknowledgments	iii
Contents	vi
1 Introduction	1
1.1 Dissertation Overview	2
1.2 Dissertation Contributions	4
1.3 Dissertation Organization	5
2 Background and Related Work	6
2.1 Background	6
2.1.1 Issues in Result Management	6
2.1.2 Solutions for Results Management	10
2.2 Related Work	17
2.2.1 Alerts and Reminders	18
2.2.2 Information Visualization	21
3 Timely Management of Medical Results: MStart	38
3.1 Requirements	38
3.2 The User Interface of MStart	40
3.2.1 Order	42
3.2.2 Track	47
3.2.3 Complete	54
3.3 Summary	54

4	Evaluation	57
4.1	Iterative Design Reviews with Medical Professionals	58
4.1.1	Method	58
4.1.2	Results	59
4.1.3	Discussion	61
4.2	Controlled Experiment for Awareness of Order Timeliness	62
4.2.1	Question	63
4.2.2	Design	63
4.2.3	Variables	63
4.2.4	Protocol	67
4.2.5	Experimental Setup	71
4.2.6	Results	74
4.2.7	Predictive Model for Results Management	79
4.2.8	Discussion	83
4.3	Evaluations of Actions for Rapid Completion	86
4.3.1	Evaluating Time to Execute Actions	86
4.3.2	Evaluating the Number of Steps	89
4.3.3	Discussion	100
4.4	Summary	104
5	Design Guidelines	106
5.1	Results Management Design Guidelines	109
5.1.1	Show pending results (Evidence: High; Conformance: Mandatory)	110
5.1.2	Prioritize by late and lost status (Evidence: High; Conformance: Mandatory)	111
5.1.3	Indicate physician acknowledgment and timeliness (Evidence: Medium; Conformance: Mandatory)	113

5.1.4	Embed actions when appropriate (Evidence: Medium; Conformance: Recommended)	114
5.1.5	Provide retrospective analysis (Evidence: Medium; Conformance: Recommended)	117
5.1.6	Distinguish preliminary and final results (Evidence: Low; Conformance: Mandatory)	119
5.1.7	Support views for different clinician roles (Evidence: Low; Conformance: Recommended)	119
5.1.8	Clarify responsibility (Evidence: Low; Conformance: Recommended)	120
5.2	Actions for Rapid Completion (ARC) Design Guidelines	121
5.2.1	Allow In-place Editing for Cells (Evidence: High; Conformance: Mandatory)	122
5.2.2	Reveal a panel on hover or click of a row for more complex actions (Evidence: Medium; Conformance: Mandatory) . . .	122
5.2.3	Make the entire row clickable if there is one action set on a row (Evidence: Medium; Conformance: Mandatory)	123
5.2.4	Support key accelerators to speed up action taking (Evidence: Medium; Conformance: Mandatory)	124
5.2.5	Spell everything out while keeping the content compact for visual scanning (Evidence: Medium; Conformance: Mandatory)	124
5.2.6	Support undo/redo of actions (Evidence: Medium; Conformance: Mandatory)	125
5.2.7	Avoid drop-downs that necessitate multiple clicks to access and select (Evidence: Medium; Conformance: Recommended)	125
5.2.8	Provide sounds for feedback (Evidence: Medium; Conformance: Recommended)	126

5.2.9	Keep the panels and popups on the screen until the user intentionally dismisses them (Evidence: Low; Conformance: Mandatory)	126
5.2.10	Keep the row visible when the panels are shown (Evidence: Low; Conformance: Mandatory)	126
5.2.11	Highlight the selected row while unhighlighting the rest of the data in all tables (Evidence: Low; Conformance: Mandatory)	127
5.2.12	Rows that appear in more than one table should be linked (Evidence: Low; Conformance: Mandatory)	127
5.2.13	Open panels as close to the mouse click location as possible (Evidence: Low; Conformance: Mandatory)	128
5.2.14	Make sure the panels and popups are visible at all times (Evidence: Low; Conformance: Mandatory)	129
5.2.15	Save and allow access to completed actions (Evidence: Low; Conformance: Mandatory)	130
5.2.16	Present a popup of row information along with the actions (Evidence: Low; Conformance: Recommended)	130
5.2.17	Color the panel and popup the same as the color coding of the row (Evidence: Low; Conformance: Recommended) . . .	131
5.3	Summary	131
6	Retrospective Analysis of Medical Orders: MSProVis	140
6.1	Introduction	141
6.2	The Process Completion Diagram (PCD)	143
6.2.1	Layout	144
6.2.2	Color Encoding	146
6.2.3	Detailed Information	146
6.3	Multi-Step Process Visualization (MSProVis)	147

6.3.1	Process Overview	148
6.3.2	Steps in Details	148
6.3.3	Actors in Details	150
6.4	MSProVis for Software Development Cycle	151
6.5	Implementation	152
6.6	Usability Studies	152
6.6.1	Study with Physicians	154
6.6.2	Studies with Novice Users	156
6.6.3	Discussion	160
6.7	Surveys	161
6.7.1	Survey 1	162
6.7.2	Survey 2	164
6.7.3	Discussion	168
6.8	Summary	170
7	Conclusion and Future Directions	172
7.1	Conclusion	172
7.2	Future Directions	176
7.2.1	Workflow Management Systems	176
7.2.2	Design Guidelines	177
7.2.3	Longitudinal Studies	179
7.2.4	Visualization	179
7.3	Summary	180
A	Requirements Analysis	181
A.1	Meeting with a Primary Care Physician	181
A.1.1	Observations	182
A.1.2	Discussion of Findings	183
A.2	University Health Center Visit	185
A.2.1	Observations	186

A.2.2	Discussion of Findings	187
A.3	Second & Third University Health Center Visit	189
A.3.1	Observations	189
A.3.2	Discussion of Findings	191
A.4	Emergency Department Visit	192
A.4.1	Observations	192
A.4.2	Discussion of Findings	194
A.5	Meeting with an Internist	195
A.5.1	Observations	195
A.5.2	Discussion of Findings	197
A.6	Visit to Outpatient Clinic	201
A.6.1	Observations	201
A.6.2	Discussion of Findings	203
A.7	Medical Center Visit	205
A.7.1	Observations	205
A.7.2	Discussion of Findings	208
B	Process Model	211
B.1	Processes	211
B.1.1	Simple Linear Process	212
B.1.2	Extended Linear Process	213
B.1.3	Parallel Process	215
B.2	Actor Actions	216
B.2.1	Actions during Process Result Review	216
B.2.2	Other Actor Actions	218
B.3	Summary	219
C	Implementation	220
C.1	Workflow Element Relationships	220
C.2	Process Hierarchy	221

C.3	Predicting Process Times	223
C.4	Data Objects	225
C.5	Interface Generator	226
C.6	MStart Simulator	229
C.7	Summary	230
D	Table Design Guidelines	231
D.1	Data Arrangement	232
D.1.1	Columns/Rows:	232
D.1.2	Row Sequence:	237
D.1.3	Column Sequence:	238
D.1.4	Related Tables:	239
D.2	Labeling	241
D.2.1	Style headers differently but ensure consistency (Evidence: High; Conformance: Mandatory)	241
D.2.2	Give the table a descriptive title with a total row count (Ev- idence: High; Conformance: Mandatory)	242
D.2.3	Keep the headers visible in the window at all times (Evi- dence: High; Conformance: Mandatory)	242
D.2.4	Align column headers with their associated data (Evidence: High; Conformance: Mandatory)	243
D.2.5	Avoid a header that is significantly wider than the data it is indicating by spreading such headers into two or more lines (Evidence: Medium; Conformance: Mandatory)	244
D.2.6	Indicate editable columns (Evidence: Low; Conformance: Mandatory)	244
D.2.7	Show a tooltip for the title that describes the table's function (Evidence: Low; Conformance: Recommended)	245

D.2.8	Do not truncate column headers; break long headers by full words whenever possible, otherwise split in the middle with a hyphen (Evidence: Low; Conformance: Recommended) . . .	245
D.3	Settings & Help	246
D.3.1	Provide custom filtering on-demand (Evidence: High; Conformance: Mandatory)	246
D.3.2	Allow settings to be saved and loaded (Evidence: Medium; Conformance: Recommended)	247
D.3.3	Do not let tooltips go outside of screen space (Evidence: Low; Conformance: Mandatory)	248
D.3.4	Provide help in a separate window (Evidence: Low; Conformance: Mandatory)	248
D.3.5	Derive filter values from current table entries, not all database entries (Evidence: Low; Conformance: Recommended) . . .	249
D.3.6	Group filter values by range if possible (Evidence: Low; Conformance: Recommended)	249
D.3.7	Support search for large tables (Evidence: Low; Conformance: Recommended)	249
D.3.8	Show a description for columns in a tooltip (Evidence: Low; Conformance: Recommended)	250
D.4	Delineation	250
D.4.1	Feature light white space between the rows, no heavy gridlines (Evidence: High; Conformance: Mandatory)	250
D.4.2	Allow sufficient space between columns to clearly separate them but no more (Evidence: High; Conformance: Mandatory)	251
D.4.3	Calculate initial column widths from data but offer resizing of columns (Evidence: Medium; Conformance: Recommended)	252

D.4.4	Define a min and max width for each column depending on the data it presents; these widths should be user changeable (Evidence: Medium; Conformance: Recommended)	252
D.4.5	Do not stretch tables to fill available space; only the last column may be stretched, if not right-aligned, to align with other tables (Evidence: Medium; Conformance: Recommended)	253
D.4.6	Add some padding to columns if there is room; allow user-defined values (Evidence: Medium; Conformance: Recommended)	253
D.4.7	Add some padding to rows for easy visual scanning; allow this to be set by the user (Evidence: Medium; Conformance: Recommended)	254
D.5	Formatting	255
D.5.1	Orientation	255
D.5.2	Alignment	255
D.5.3	Number and Date Precision	258
D.5.4	Font	261
D.5.5	Color and Icons	262
	Bibliography	265

List of Figures

2.1	Results Manager front screen	27
2.2	Little-JIL Chemotherapy Preparation and Administration	28
2.3	Radiology interpretation process modeling	29
2.4	A Personal Task List Manager Component	30
2.5	RADAR – A Multitask Coordination Assistant	31
2.6	Approaches for displaying multiple system alarms	32
2.7	BizView	33
2.8	Sunfall Data Taking window	34
2.9	TimeSearcher	35
2.10	Lifelines2	36
2.11	LifeFlow	37
3.1	User interface of MStart (all orders by a clinician)	41
3.2	User interface of MStart (orders per patient by a clinician)	43
3.3	User interface of MStart (orders for patients with visits today)	44
3.4	User interface of MStart (orders for all patients by all clinicians)	45
3.5	Order screen of MStart	46
3.6	Planned orders	48
3.7	Changing Default Colors	50
3.8	Custom filters on demand	51
3.9	The trace of responsibility	52
3.10	ARCs have interactive components to facilitate completion.	53
3.11	Complete screen of MStart	55

4.1	Baseline Interface	64
4.2	Second Level Interface	65
4.3	Third Level Interface	66
4.4	Training Session	72
4.5	Experiment results regarding the time to correctly mark order timeliness	77
4.6	Experiment results regarding the number of corrections made in identifying order timeliness	78
4.7	ARCs have interactive components to facilitate completion.	87
4.8	The state-of-the-art follow-up screen in MStart.	88
4.9	Informing patients of their results.	92
4.9	Informing patients of their results.	93
4.10	To follow-up without a follow-up, a result's right-click menu has QVerify, which stands for quick verify.	94
4.11	Confirming a result that requires a repeat order for follow-up.	96
4.11	Confirming a result that requires a repeat order for follow-up.	97
4.11	Confirming a result that requires a repeat order for follow-up.	98
4.12	Closing the patient record is via the File menu.	99
4.13	Confirming a result that requires a new visit for follow-up.	100
4.13	Confirming a result that requires a new visit for follow-up.	101
4.13	Confirming a result that requires a new visit for follow-up.	102
5.1	Veterans Administration View Alert window.	111
5.2	Rich tabular displays as seen by a physician	112
5.3	ARCs have interactive components to facilitate completion.	113
5.4	Taking an action.	115
5.4	Taking an action.	116
5.4	Taking a follow-up action in OpenVista.	117
5.5	Retrospective Analysis Visualization	118

5.6	Views for different clinician roles	119
5.7	Trace of responsibility	121
5.8	In-place edit for a table cell value.	122
5.9	A dialog box to take actions.	123
5.10	Keyboard shortcuts revealed.	124
5.11	Drop-downs on popups	125
5.12	The panel (left) and popup (right) stay visible until the mouse moves outside of their borders.	127
5.13	Selecting one item in the pending table highlights the same item in the results table.	128
5.14	Popups are visible.	129
6.1	MSProVis Interactive Visualization	141
6.2	The Process Completion Diagram	144
6.3	Applying MSProVis to Software Development Cycle	153
6.4	Survey 1 explores three different designs for the PCD	162
6.5	Chart 1	165
6.6	Chart 2	165
6.7	Chart 3	165
B.1	XML definition of a simple linear process	213
B.2	XML definition of an extended linear process	214
B.3	XML definition of a parallel process	215
B.4	XML definition of process result review actions	217
B.5	XML definition of adding process-specific actions	218
B.6	XML definition of laboratory technician actions during preliminary result recording	219
B.7	XML definition of a TSH test process	219
C.1	Class hierarchy of workflow element relationships built by MStart	221

C.2	Class diagram of laboratory test processes	222
C.3	XML file to predict process times	224
C.4	Entity-Relationship diagram of database Actor and Organization objects	225
C.5	Relationships between events, orders, and logs	226
C.6	MStart interface as seen by a clinic assistant	227
C.7	Pending blood test visualized in MStart	228
C.8	ARC for a laboratory technician recording preliminary results . . .	228
C.9	MStart accepts an order file generated by <code>OrderWriter</code> class	229
D.1	Rich tabular displays as seen by a physician	232
D.2	Extra columns are indicated with a double-arrow.	234
D.3	Excel defaults to same styles.	234
D.4	Sort icons and column sequence.	235
D.5	An example table that requires users to compute.	235
D.6	Buttons that show more items in a list.	236
D.7	Combined and abbreviated columns	237
D.8	Row sequence	238
D.9	User-settable table sizes	240
D.10	Planned table is collapsed.	240
D.11	Table title contains row count (after filters are applied).	242
D.12	Column headers are still visible when the table is scrolled down. . .	243
D.13	Aligning headers with data	243
D.14	Long headers are split into two or more lines	244
D.15	Editable column	245
D.16	A tooltip for the table title describes the table's function.	245
D.17	Column header is split with a hyphen when the data itself takes little space or the column is dragged.	246
D.18	Custom filters on-demand.	247

D.19 Large tables need a search option.	250
D.20 Tooltips for headers provide more information about the column.	250
D.21 Heavy gridlines versus light white space between the rows.	251
D.22 Column widths that are equal versus calculated from data.	252
D.23 An example of a too short and a too long column.	253
D.24 Columns without and with padding.	254
D.25 Rows without and with padding.	255
D.26 Consistent alignment of similar data.	256
D.27 Double-aligned column	256
D.28 Headers are bottom-aligned when they are split in two or more lines.	257
D.29 A center-aligned column with an icon impedes horizontal continuity.	257
D.30 A tooltip shows the full table cell value.	259
D.31 Column data shown changes as the column widths change.	260
D.32 Consistent precisions	260
D.33 Cell split into two lines	261
D.34 Saturated color-coding	263
D.35 Color legend shows and filters based on color coding.	263

List of Tables

4.1	Orders for Interface 1	74
4.2	Orders for Interface 2	75
4.3	Orders for Interface 3	76
4.4	Expected Durations for Orders	76
4.5	Results of evaluating time in representative actions.	89
4.6	Results of evaluating steps in representative actions.	99
4.7	Summary of evaluations, results, and insights.	105
5.1	Summary of Results Management design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).	133
5.2	Summary of table design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).	133
5.3	Summary of Actions for Rapid Completion (ARC) design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).	138
6.1	Results of Survey 1	164
6.2	Results of Survey 2	167

Chapter 1

Introduction

“Jane hurts her leg in a bad fall. Her primary care physician orders an X-Ray, one of the dozens of orders he writes that day. Jane chooses to schedule her appointment at an independent radiology center near her home. She remembers to go the next day, and a radiology technician takes the images. Next, the radiologist is supposed to review the images and write a report, which will be faxed back to the primary care physician, who should review the result and follow-up with Jane, if needed. However, something goes wrong and the physician never sees the results. He has an Electronic Health Record (EHR) system¹ but he does not notice that he never received Jane’s results. Jane loses her leg since the fracture was never treated.”

Unfortunately, this is a real story. Many things can go wrong in this complex process with multiple steps, actors, and responsible parties, and patients are physically harmed when results of medical orders (e.g. lab tests, imaging studies, referrals to specialists) are mishandled or lost [53]. Patients can become sicker or even die. In fact, failure to follow up on abnormal results is one of the most frequent causes of malpractice litigation in outpatient medicine [104]. It impacts efficiency and effectiveness of treatment, patient safety, and overall satisfaction [90].

¹An EHR system is composed of electronic health records of patients, computerized physician order entry, and results management.

1.1 Dissertation Overview

There is a rich history of research on the issues of outpatient settings [6,15,27,36,40,53,60,90,108,115,116,124,132,149,159]. Researchers exploring common problems in ambulatory care mainly highlight the importance of timely results management in primary care [54,107,123,148]. They indicate that medical orders of patients were missed by the ordering primary care physician [27,38,60,126,150]. By a series of case studies, researchers report the number, reason, location, categories, and steps of missed reports [36,38,89,124,159]. Some researchers analyze the consequences of missing or delayed results [53,104]. Not only do researchers find out the current methods for coping with errors and clinical personnel's satisfaction with these methods, but researchers also suggest that well-designed user interfaces could help medical staff during this complex process [6,39,73,109,148].

Prior research reports in great details about the challenges and the importance of timely results in primary care, therefore, one might expect advanced visual displays to be generally available to help medical staff during this complex process. On the contrary, other studies show that there is no standard, widely applied method for handling results [15,39,89,149]. A study of current EHR systems used for tracking discovered there were interface and logic errors in results routing, physician records, system setting interfaces, and system maintenance tools [159]. During my interviews, I saw environments where needed results are received in a timely and reliable fashion (e.g. emergency rooms in a hospital with all test facilities in house); while others reported very high rates (20%) of late or lost results. I observed physicians keeping paper cheat sheets, while staff – and patients – reported spending hours on the phone tracking what went wrong. Even when physicians have reliable systems, medical staff do not routinely check the status

of all pending orders [40]. Better designs to help medical staff such as physicians or clinic managers who are responsible for timely results management could bring large benefits.

To support clinicians in this complex task, this dissertation aims to use human-computer interaction and information visualization techniques to ensure timely results management. The goal of human-computer interaction is to improve the interactions between humans and computers by designing systems that minimize the barrier between the human's cognitive model of what they want to accomplish and the computer's understanding of the human's task [114]. Therefore, I believe that by designing suitable visual representations and interaction techniques for EHR systems, medical professionals can track their clinical results rapidly and with fewer errors. I worked on two ways to ensure timely management of medical results:

1. **Interactive visual display with action support to ensure timely completion of results:** I visited hospitals and clinics to study the issues and define the results management workflow. While result correctness is of paramount concern, it is important to expedite this process by explicitly accounting for timing. As lists remain the most common way to manage daily work, my focus is on interactive tabular displays to show the progress of orders. Since commercial systems do not currently design the interactive features well, my research aims to improve actions so that results are promptly followed up.
2. **Visualization support for clinic managers to summarize and compare performance:** Although the tabular display helps users for now and later, in safety research it is important not only to be alerted by errors oc-

curing at the moment, but also to identify issues likely to happen in the far future and to know how to prevent them [122]. By recording all results, retrospective analyses enable users to understand the issues aroused before and their root causes to allow for continuous improvement. Clinic managers can learn from past data, classify general issues, and prevent prior problems from reoccurring. For instance, a visualization could illustrate the percentage of medical results that were delayed permitting clinic managers to inspect bottlenecks. Such analyses are beneficial for the management of guidelines, performance measurements, and quality control.

This dissertation describes in detail how to solve these two research questions and includes several user studies, ranging from usability studies, a controlled experiment to a predictive model that prove the benefits of my approaches. A list of guidelines have been developed, which can be generalized to tracking interfaces built for other processes where various responsible parties collaboratively handle different steps, being separated by time and distance. Examples include but are not limited to the software development cycle, paper-review in academic journals, and business processes (online shopping, personal financial planning, the return merchandise authorization in the electronics industry, the internal supplier agreement in IT companies, sales within a huge corporation).

1.2 Dissertation Contributions

While there many limitations in current systems, there is little evidence in the literature that provides support for using visual displays to ensure timely results management. There is a need to develop an interactive display that can address drawbacks. It is fundamental to design richer interfaces for clinicians with follow-up

action controls embedded. A user interface and an information visualization tool automatically-generated from a process specification seems promising solutions to the problem. My contributions are the following:

1. Clinician observations and interviews in clinics and hospitals, with discussions of findings that can improve results management.
2. Comparative evaluations of the displays and action techniques for use in time-critical applications where distractions undermine users' ability to track and react to orders.
3. Design guidelines for an interactive visual display to overview the dynamics of ongoing orders and to act on the results rapidly.
4. A novel retrospective analysis visualization and usability studies for this tool.

1.3 Dissertation Organization

The next chapter will introduce the reader to the background on medical informatics applications and relevant work in the area of human-computer interaction and information visualization. The rest of this dissertation is organized as follows. Chapter 3 describes my observational work for the design and implementation of an interactive tabular display. Chapter 4 reports on the evaluations performed to prove the effectiveness of this research. Chapter 5 explains the design guidelines developed based on literature reviews, analyses of existing systems, and empirical results. Chapter 6 presents the visualization to retrospectively analyze medical orders and evaluations of this approach. Finally, Chapter 7 concludes with a discussion of future work.

Chapter 2

Background and Related Work

2.1 Background

This section will introduce the issues in primary care offices and suggested solutions of medical researchers.

2.1.1 Issues in Result Management

Wahls [148] defined missed results as “mishandling of abnormal results”, which are then lost to follow-up. In a provider survey, 47% of respondents reported encountering one or more patients (a total of 312 cases in approximately 20,000 visits) with a missed result in the previous two weeks [149]. Clinical laboratory tests and imaging studies were most commonly reported as having missed results. A study by Tang [132] confirmed that in 81% of cases, physicians could not locate all the information to make informed patient care decisions during a visit. Among the unavailable data, the most common type of missing information was related to the result of laboratory tests/procedures (15-54%) [36]. Researchers also found that 37% of providers indicated at least one patient (a total of 276 cases in approximately 20,000 visits) who experienced delays in diagnosis or treatment because of missed results [149].

Besides self-response survey data from care providers, recent studies based on an examination of patient medical records confirmed that errors and delays oc-

curred during screenings for osteoporosis, breast cancer, and colorectal cancer. These results initiated a widespread discussion about the lack of adequate results management systems. In one study, Cram et al. [27] reviewed bone density scans of 428 patients for five months. Of the 48 patients who were newly diagnosed with osteoporosis, 16 received no treatment recommendation. In 11 out of 16, the medical record showed no result review by providers. In another study, Poon et al. [108] showed that of 126 women with abnormal mammograms, 45 did not receive appropriate and/or timely follow-up care within seven months. In addition, the physician did not adequately document discussions with patients in 29% of the cases with abnormal results, and did not document the follow-up plan and 27% of those cases. Furthermore, during a four-month period using screening cards mailed by a managed care organization, Baig et al. [6] identified 544 patients with abnormal fecal occult blood testing, a screening test used for reducing colorectal cancer mortality. A total of 248 of those patients did not undergo a complete diagnostic evaluation after the positive test. Only 50% of physicians were able to provide reasons, the remainder had no follow-up due to a variety of causes, including physician, specialist, or practice-related decisions.

Although often laboratories may cause the delay, as can be seen from the previous paragraph, there are cases when they are not the root cause of the delay. Steindel and Howanitz [129] investigated the result turnaround time in the emergency department (ED). Their results showed that delayed results lengthened the treatment time and the duration of stay in the ED. Therefore, there is a need for identifying the bottlenecks in different situations [129] and support decision-making, which can improve healthcare delivery [145].

Other researchers analyzed the occurrence of problems related to results man-

agement in a broader setting. For example, Singh et al. [123] studied all critical imaging alert notifications in an outpatient setting with an advanced electronic medical record system for eight months. Of the 123,638 imaging studies, 1,196 images generated alerts via the “View Alert” notification window; 217 (18.1%) of these were unacknowledged within two weeks of transmission. Timely follow-up at four weeks was lacking in 92 alerts (7.7%) and the occurrence was similar for both acknowledged and unacknowledged alerts. Nearly all missed abnormal results had a measurable clinical impact in terms of further diagnostic testing or treatment. In addition, Roy et al. [115] examined 2,644 patient discharges during five months. Of these, 1,095 patients collectively had 2,033 results arrive after discharge. Of these results, 191 were potentially actionable, and surveys were sent to 155 primary care physicians. Of the 105 survey responses, physicians indicated they were unaware of 65 results, although 24 of these were actionable and eight were urgent. Finally, Schiff et al. [116] linked laboratory and pharmacy databases over a two-year period to determine if patients with elevated thyroid-stimulating hormone (TSH) received appropriate treatment (the drug levothyroxine). Out of 36,760 unique patients tested for TSH levels, 982 had high levels, including 177 patients who had no associated recorded prescriptions. While 54 patients were lost to follow-up, the researchers contacted 123 patients and found 23 who were unaware of their abnormal results.

According to data from the National Ambulatory Medical Care Survey of 2002, family physicians and general internists order lab tests in 29-38% of patient encounters on average, and imaging studies in 10-12% [54]. A physician in an internal medicine practice weekly reviews a mean of 930 individual chemistry/hematology tests and 60 pathology/radiology reports [110]. These orders are for screening or

diagnostic purposes, or to manage and monitor medications and/or chronic health problems [38]. Although a few of the orders are performed in the office while patients wait, most patients or their samples are sent to outside locations such as standalone testing facilities or hospitals [38]. Results may be available during the same office visit or take weeks to obtain and do not have a universal format [107].

It is not surprising that multiple testing locations, large number and variety of orders, as well as variable reporting processes would lead to delays and errors [54]. One study found that 83% of primary care physicians reported at least one delay in reviewing results in the prior two months [107]. As can be seen in the studies cited above, patients can be – and sometimes were – seriously harmed by such errors in results management [53, 123]. According to a study by Hickner and colleagues [53], adverse consequences of missing results include delay in care (24%), time/financial loss (22%), pain/suffering (11%), adverse clinical consequence (2%). Patients were harmed in 18% of these instances while in 28% harm status was unknown. As a consequence, failure to follow up on abnormal results is in fact a frequent cause of medical malpractice litigation in outpatient medicine [104]. This, in turn impacts efficiency and effectiveness of treatment, patient safety, and overall satisfaction [90].

These studies all validate that delays and errors happen during the management of results. This shows that there is room for improvement in the process of ordering tests [127, 145, 146, 151]. It is not the lack of effort that causes many institutions and physicians to lose results as significant time was spent searching and managing results [126]. Hickner et al. [54] propose that the complexity of the process causes such issues. This includes physicians ordering and reviewing a large number and variety of laboratory tests and imaging studies [27, 60, 109, 123], as well as the variable result arrival times and reporting processes associated with

multiple testing locations [39, 54]. Wahls and Cram [150] pointed out that there are no established standards as how best to manage results. Yackel et al. [159] studied current information systems used for tracking and found there were interface and logic errors in results routing, physician records, system setting interfaces, and system maintenance tools. Even when physicians did have reliable methods, medical staff did not routinely check for pending orders without results [40]. Thus, less than one-third of physicians were satisfied with how they managed results [90].

Research studies also revealed the desired features of an electronic results management system. Most of them found out that a tool to generate and send result letters with predefined texts to patients via email is the highest-rated feature of a potential results management system [107, 150]. Physicians wanted to acknowledge all results electronically through an “in-box” function [107]. Another most desired capability concerned the prioritization of diagnostic results such that abnormal ones are shown before normal results and the built-in review prompts support the physician to make a decision and take further action [107, 150]. Physicians also suggested tracking their orders to completion, i.e. a warning mechanism to detect whether orders have not been completed [107]. Finally, physicians requested delegation of responsibility to other staff [150]. More specifically, they desired a forwarding capability to allow the use of surrogates during planned absences and a consistent process for designating proxies when they are unavailable [107, 150].

2.1.2 Solutions for Results Management

To formalize the clinical results management in a family practice setting, Mold [89] described four steps during which laboratory errors were detected: 1) Test Tracking: 15% of ordered laboratory tests were not recorded in the logbook and

were missing in patients' charts, 2) Patient Notification: 92% of patients received their test results, 3) Documentation of Notification: 40% of charts were lacking sufficient documentation of patient notification (half were not initialed and half were not dated), 4) Follow-up Tracking: 40% of charts had poor documentation of follow-up tracking while 35% of patients did not follow up in 3 months, and 10% followed up late.

More recent research studies by Hickner and his group used the process steps depicted below for the management of results in outpatient settings [54]. There are multiple steps involved in the management of test results and this workflow captures the fundamental steps. The workflow begins when a medical provider places an order and ends when all the results are acted upon. According to this workflow, the overall process is grouped into three major phases, which also include several sub-steps. Moreover, the processing is defined sequentially, i.e each step in this process is triggered when the previous is completed. For example, a laboratory test cannot be performed if the patient's blood or urine is not collected first. A delay in one step may cause a delay in the whole process. The performance of an actor affects the duration of the step and the entire process.

1. Pre-analytic phase
 - (a) Ordering the test
 - (b) Implementing the test
2. Analytic phase
 - (a) Performing the test
3. Post-analytic phase
 - (a) Reporting results to the clinician

- (b) Responding to the results
- (c) Notifying the patient of the results
- (d) Following-up to ensure the patient took the appropriate action based on results

Using the model above, medical researchers, clinicians, and healthcare system developers can group the laboratory tests in two to four categories using a combination of imaging studies (radiology) and laboratory tests (chemistry, hematology, and pathology) [27,39,54,149]. In the simple case, the patient is asked to get a test done, and the physician analyzes the results that arrive back in her/his office. Further details are added based on test type (a radiology study includes a radiologist report, while a urinalysis will require a lab technician report). Other variations may occur depending on the test conducted; for example, a lipid panel order in blood work might necessitate the patient fasting overnight. Multiple errors can and do occur at each step [39,54,89]. For instance, in the pre-analytic phase an order could be lost, or a specimen never drawn or lost or damaged during transport. During the post-analytic phase, the results may be misplaced, not documented in the patient's record, or not followed up with the patient.

Carraro and Plebani [21] found that mistakes during laboratory testing occur primarily during the pre-analytic phase (68.2%), with 18.5% and 13.3% during the post-analytic and analytic phases, respectively. They suggest the improvement of the total testing process, including pre- and post-analytic phases as a solution to the problem. Other authors [54] suggest that standardization of collection, preparation, and delivery of specimens could solve ordering and implementation problems. To help with tracking and returning issues, they recommend increasing the quality of communication via dual tracking between the lab and physician's

office. Specifically, they describe a backup fail-safe system that records all orders requested, sent, received, and completed. In addition, a computerized tracking system integrated into an electronic health record (EHR) is suggested for response and documentation mistakes [54]. To avoid patient notification errors [54], auto-generated letters and voice systems may aid in communicating results. This research confirms the need for a complete tracking interface that provides users continuous feedback on order progress and results.

Current methods that clinical personnel use to track their test results are still through paper charts [38]. Most research reports only suggest general-purpose alternatives such as logbooks [150] or checklists [107] rather than an application tied to the task. In accordance, researchers have approached the problem of missed results by implementing non-electronic solutions. Marcus et al. [80] evaluated two interventions for women with abnormal Pap smears: (i) an intensive follow-up protocol that depends on numerous attempts to contact the patient via mail or phone, (ii) economic vouchers to compensate for the expenses of follow-up visits. Their study showed that both conditions improved the rate of follow-up as compared to a control group. Sung et al. [130] distributed a survey to physicians, to learn their interest in direct reporting of laboratory test results to patients by mail. This research demonstrated that care providers preferred direct reporting of normal, rather than abnormal, results. Secondly, physicians were more supportive in direct reporting of results deemed to have less emotional impact. Both of these solutions suggest that including the patient in the process will facilitate results management. While patient involvement is an important layer of safety, patients are not always able to interpret results and a robust system should not rely on human vigilance alone to manage results.

While there is some user interface work in showing numerical laboratory test results [74], to the best of my knowledge, the most comprehensive test tracking system is Results Manager (Figure 2.1) by Partners Healthcare [110]. It organizes laboratory tests around patient visits and does not let individual tests to be listed. Users can see all (currently open and closed) visits, flag them, and add to their watchlist. Chemistry, hematology, radiology, and pathology tests are the only supported types of tests so the addition of new tests or the modification of existing ones is not supported. Abnormality has three degrees: critical, abnormal, and normal. Through patient charts, physicians may acknowledge a result, add a visit note, or generate patient result letters although researchers reported that automatically generated papers sometimes make no sense and are typically not useful [5]. This tool has a complex user interface since the user has to switch between the Manager screen where all visits are listed and patient charts where the result details are. It is not easy to learn either because all numerical results are listed in a table format that could easily be overlooked even with color options supported (red for high, blue for low abnormal result). Although clinicians may mark results with reminders for future follow-up, follow-up decisions are left to the provider. It is not obvious where to see individual late results immediately and act upon that. It does not support interfaces for medical staff that are actively involved in the testing process, other than the physician.

A clinical event monitor [59] is a system in which database operations and external events trigger the evaluation of a condition. The monitor determines whether an action should be performed based on the conditions, consisting of events and patient data. When certain conditions occur, the monitor generates alerts for clinician, patients, or other organizations. LabCheck [35], CLEM [147], ReNAP [109],

and many other clinical event notification systems [73,93,99,118] remind a provider about follow-up. Since these systems are implemented as rule-based engines accessing a database, they have various complex mechanisms. More importantly, such systems usually do not show the progress of the event, so the clinician has less control and understanding over the entire process. In fact, researchers indicated that they can generate many undesired alerts and cause clinician alarm fatigue; this may result in the clinician either bypassing or missing the truly important alarms [70,123].

As opposed to rule-based architectures, some researchers have applied workflow management techniques to clinical situations. Instead of looking at specific events, these techniques seek to model the general case. For instance, Ling and Schmidt [77] applied time workflow (Petri) nets to an example of a Patient Workflow Management System [31]. They define firing time intervals, and show that some transitions are reachable only after certain time periods have passed. In their example, some places are labeled as actors while transitions, which are assigned durations, represent task steps. Because some places are not labeled, it is not clear who is responsible for the tasks executed. This model is good for understanding workflow, however it does not constrain the total time. Little-JIL [23] is a process definition language used to model medical processes. Christov et al. first create a detailed and precisely defined model of a medical chemotherapy process with Little-JIL (Figure 2.2). Then, they identify process defects and vulnerabilities that pose safety risks. Applying rigorous automated analysis techniques to the original model leads to an improved process. Finally, they reanalyze the improved process to show that the original defects are no longer present. The final model serves as a proof but is no longer used.

The workflow management techniques were borrowed from the business process modeling domain, whose primary interest is in developing advanced modeling languages to formally represent business processes. While BPMN [156] uses graphical representations, the specifications written can be converted to BPEL [45] or XPDL [155] using XML. Later on, these specifications could be fed into a workflow engine that generates a running application. One such tool referred to as YAWL [143] is based on Petri nets [102] that offer a mathematical theory for process analysis. Its Resource Service is responsible for the (i) allocation of tasks to users through its Manager, (ii) embedment of web forms to administer the engine and data, and to provide a UI to process management via its Worklist Handler, (iii) creation of web forms on-the-fly for the presentation and completion of work items with its Dynamic Forms Generator, (iv) execution of codelets for automated tasks using its Coordinator. A key difference between this work and mine comes from the difference in goals. These models assist in optimizing or testing workflows in the context of an ideal environment, which may fall far from reality. Process management models do not provide information about what is actually happening or happened in practice, because these models are not bound to data [79]. Poorly designed models can affect the process and then lead to incorrect decisions [144]. My work focuses on what can go wrong in the real world.

Considering the limitations of process modeling, a new research discipline, named process mining, has emerged. The foundations of this discipline are machine learning, data mining, process modeling and analysis [144]. Process mining techniques are applied to integrate business processes and event-logs aiming at discovering, monitoring and improving real processes. Mans et al. [79] conducted a case study where they applied process mining techniques to analyze a healthcare

process. In that study, the authors used a dot-chart, which can be considered as a visual analytics technique, to gain a better understanding of the event-logs in the process. Users were able to obtain new insights and identify patterns in the data. Their study showed that process mining techniques assist in improving a healthcare process, and visual analytics techniques enhance users' capability in exploring the data.

Recently, several computer-interpretable clinical-guideline modeling languages have been developed including machine-executable ones that support authoring, editing, and enactment [48, 100, 119]. A medical guideline is a document with the aim of guiding decisions and criteria regarding diagnosis, management, and treatment. Medical guidelines do not always stipulate a specific process or schedule for performing medical services. Researchers proposed adopting formal methods to verify these protocols [8, 16, 101, 136, 137]. Clinical pathways, also called care pathways, have been introduced to overcome these drawbacks. They use medical guidelines to define and sequence different tasks of healthcare professionals. For instance, Noumeir describes the model of a radiology interpretation process [92] (Figure 2.3) but this specification does not have the necessary elements to generalize to interpretation processes, or imaging study processes.

2.2 Related Work

Two relevant areas of research inspired my work: alarm and alert systems and information visualization.

2.2.1 Alerts and Reminders

Alerts and reminders have been extensively studied in the human-computer interaction domain.

One approach is to support organization of users' daily tasks and roles. These are called Personal Information Management (PIM) systems [11]. Some focused on organizing documents, files, and notes for the purpose of reminding and efficient retrieval [7, 18, 43], others studied calendars and schedulers for time management [14, 95], another group of researchers looked at email inboxes [157] and to-do lists [49] for task management. The earliest system, called Lookout [57], predicts an event date and time on a communication message based on an analysis of its content. This prediction is then used to schedule delivery of automated assistance. The most notable is a Task List Manager (TLM) system named TaskVista [9] (see Figure 2.4) that helps users manage and execute their to-dos with its Gantt Chart based temporal view. To-dos are assigned priorities (importance), time constraints (durations and deadlines), participants, and abstraction levels (simple task vs. project). Another system is a Multitask Coordination Assistant (MCA) referred to as Reflective Agents with Distributed Adaptive Reasoning (RADAR) [42] (Figure 2.5). Its Action List provides a task-centric view of an email inbox by listing the tasks contained within email messages under the following categories: incomplete, overflow, and completed actions. The user can inspect the tasks that RADAR created, add new ones, delete the wrongly created ones, and launch web pages to perform some of the tasks. Its Progress Bar visualization suggests a schedule, which specifies an order in which to perform outstanding tasks. It shows completed and deleted tasks to the left of the current time, and the suggested schedule to the right. These systems generally display tasks in a list [9, 42], which

corresponds to the serial display in Endsley’s definition (see Figure 2.6a) [41]. Since tables are the most common way of organizing users’ work, it may lead to promising research to improve this familiar design. They all remind users about their pending tasks but their design does not assist user decision-making and actions, i.e. they do not permit the user to perform their tasks. Therefore, the user is expected to switch to another screen to act on their tasks. However, this causes issues since interruptions have been found to reduce awareness as explained in the next paragraph.

Another approach was to study how users react to notifications and interruptions. An in situ diary study [29] characterized the amount of task switching and interruptions experienced by typical knowledge workers over the course of a work week. The results showed that task complexity, task duration, length of absence, number of interruptions, and task type influence the perceived difficulty of switching back to tasks. Specifically, complex tasks comprise a significant portion of workload, but reacquiring such tasks is considered difficult. The findings suggest that methods for capturing and remembering representations of tasks may be valuable in both reminding users about suspended tasks, and in assisting users to switch among the tasks. Examples of such methods include time-centric visualizations and tools that can record – and reconfigure upon demand – the layout of multiple windows of content and applications that comprise a task. Another study [28] probed the cost of interrupting users with instant messages during different phases of a computing task. The researchers found that (i) interrupting users during the “evaluation phase” of the task resulted in significantly longer completion times than interruptions in other phases, (ii) interruptions that were irrelevant to the task resulted in longer times to process the message and longer task resumption times

than relevant messages. Researchers conducted a field study [62] of the computing activities of 27 users over a two-week period, exploring the suspension, recovery, and resumption of tasks in participants' natural work settings. They discovered that (i) participants spent on average nearly 10 minutes on switches caused by alerts, and spent on average another 15 minutes before returning to focused activity on the disrupted task, (ii) following an alert-based suspension, subjects would often visit several applications in addition to the notifying application, (iii) 27% of task suspensions resulted in more than two hours of time until resumption. In interviews, users attributed long delays to the loss of context associated with the task switch. In addition, findings about the association between greater visibility of windows of suspended applications and faster resumption of tasks suggest that visual cues may serve as reminders to return to suspended applications. Bounded deferral [58] study allowed users to wait some pre-specified maximum amount of time before being informed about an alert, so as to minimize interruptions in return for a relatively small cost of delayed awareness. One study [61] explored effects of managing notifications with a novel defer-to-breakpoint policies system on users and their tasks. Results showed that (i) scheduling notifications at breakpoints reduces frustration and reaction time relative to delivering them immediately, (ii) the relevance of notification content determines the type of breakpoint at which it should be delivered. This indicates that users would likely adopt the use of notification management systems in practice. A later study [63] investigated the effects of email notifications and their imposed absence on users' task-execution patterns. Results showed that users react to only about a quarter of all notifications, and that user focus on primary tasks is largely unaffected if notifications are disabled. Moreover, users view notifications as a mechanism to provide passive awareness

rather than a trigger to switch tasks and are willing to incur some disruption to maintain that awareness. Although these studies inform the design and the guidelines developed can readily be used in future systems, none of them addresses the issue of time-critical awareness.

2.2.2 Information Visualization

Existing literature on designs for situation awareness (SA) and temporal event data inform about related work in information visualization and visual analytics. It has been established that interactive visual displays let domain experts freely explore large data sets, quickly identify important information, and consequently generate new questions [120].

2.2.2.1 Situation Awareness

Designs for situation awareness were mostly domain-dependent so this section will explain the systems by domain.

The most common tools are in the domain of network security alerts for intrusion detection tasks and emergency response situations. VisAlert [47] is a visual correlation tool facilitates SA in complex network environments by providing a holistic view to help detect, diagnose, and treat malicious activities. An inventory management system such as BizView (Figure 2.7) [106] is similar in that it allows an overview of all alert on the map, giving user ability to filter and group alerts by their priority. Comparative evaluations and field studies [69, 138] revealed that the textual interface allows users to better control the analysis of details of the data through the use of rich, powerful, and flexible commands while the visual interface allows better discovery of new attacks by offering an overview of the current state

of the network, thus, hybrid interfaces were recommended. Two experiments with the decision-space visualization and analysis tool [103] explored weighing strategies of decision makers for the options presented to them. The majority of participants did not agree with top-ranked options and only when participants were given control over the rankings, they selected the most top-ranked option. A simulation of an emergency scenario run with students, half using a visualization system and half text-based system, found that a visualization system aids comprehension and projection, but has no effect on perception [33]. Human Supervisory Control (HSC) system for power grid supervision [88] was used in field studies which discovered using parallel coordinates for representing alarms and providing a visual way to filter the coordinated alarm list outperformed the alarm list alone in terms of execution time, correct answers to tasks, and tasks rated as difficult. Results of case studies with a novel collaborative visual analytics application called Sunfall Data Taking (Figure 2.8) [4] for scientists in the astrophysics demonstrated its effectiveness in situation awareness. In the maritime surveillance domain, users are involved in the anomaly detection process. A usability assessment was carried out for an overlay visualization of normal behavioral models built from maritime traffic [112]. Participants were divided into two groups, with/without visualizations, to establish a normal situational picture of vessel traffic and the results unveiled that visualizations helped participants perform better but no significant differences regarding the time to complete the tasks were noted. Within the same domain, given that uncertainty exists in the data that populates the maritime surface picture, experiments comparing different iconic representations demonstrated how uncertainty could be brought to the attention of an operator [81]. The only solution regarding a domain-independent approach was a Decision-Centered Visualization (DCV) sys-

tem for time-critical applications [67, 68]. Researchers integrated a domain knowledge ontology and database with an interactive visualization architecture. These tools include simple visual displays for monitoring and more complex, linked, multiple displays for supporting diagnosis and analysis that allow analysts to drill down and examine in more detail and from several different views. Current tools focus on the monitoring phase with limited support for analysis and diagnosis of alerts. Dynamic interaction and exploration capabilities in these tools are usually missing or limited, although they are crucial for the tools' successful application. New tools incorporating these guidelines should be developed to support the entire process of detecting, monitoring, and follow-up analysis activities.

2.2.2.2 Temporal Event Data

The complexity of analyzing a multi-step process and deriving insights from retrospective data produces a need for a simple and intuitive presentation integrated in a visual analytics approach. The purpose of a visual analytics approach is to allow clinic managers to gain an overview of the results management process, interact with the result data, identify the critical tasks, highlight the root causes of delays, and make better decisions.

Previous works [25, 52, 55, 72, 76, 83, 111, 154, 160] have shown the effectiveness of visual approach in analyzing temporal data from different domains. Correll [25] and Heer [52] examine how different visual effects can support better graphical perception of time-series data. TimeSearcher (Figure 2.9) [55] is an information visualization tool that combines query-by-example and time-box queries that allow users to draw regions on a two-dimensional display to specify constraints on time series datasets. CloudLines [72] deals with the need to keep an eye on recent

events together with providing context on the past by using a logarithmic time scale that lets the items fade away according to their relevance. VizTree/Diff-Tree [76], which is a time-series pattern discovery and visualization framework based on augmenting suffix trees, offers motif discovery, anomaly detection, and query by content. LiveRAC [83] is a visualization system for browsing and correlating large collections of system management time-series data (consisting of hundreds of parameters across thousands of network devices) with high information density using a re-orderable matrix of charts, with semantic zooming (adapting each chart's visual representation to the available space). KronoMiner [160] is a multipurpose time-series exploration tool based on a radial display that can be drilled into details by facilitating the identification and manipulation of multiple sub-pieces of interest for finer analysis, rather than providing a single global view of numerous long time-series to be analyzed as a whole. Weber et al. [154] presented a new approach for the visualization of time-series data based on spirals as opposed to classical bar charts and line graphs with the goals of detecting and confirming periodic behaviors. All these systems indicate the need for new information visualization techniques to analyze time-series data. Instead of a new visualization metaphor, Process Completion Diagram (PCD) employs an advanced version of the more familiar bar chart representation for the same purpose. Pretorius and Wijk [111] present a new method that combines a schematic diagram, time series plots, and a state transition graph to provide the user with a powerful analysis tool. On the contrary, Multi-Step Process Visualization (MSProVis) makes use of only PCDs for consistency.

A relevant area of research is software visualizations for event traces. While making use of multiple coordinated views, AllocRay [113], Zinsight [32], and Car-

diagram [117] all show transition sequences via lines or in a node-link diagrams. Several interactive visualizations have been developed to present event-log data. LifeLines [105] is an interactive visualization that presents an overview of a patient's medical record. This visualization was the inspiration for LifeLines2 (Figure 2.10) [152], but it visualizes a collection of medical records, where users can explore the event logs for temporal patterns. LifeLines2 is limited in showing an overview of event sequences. This limitation was addressed in LifeFlow (Figure 2.11) [158]. LifeFlow provides an overview of event sequences by summarizing all possible sequences and visualizing the temporal space of events within sequences. ProcessLine [78] visualizes time series data of a process combining several visualization techniques in one. Their evaluation using time-series data from a beer industry showed the effectiveness of the ProcessLine. McNames et al. [84] developed a graphical display of semiconductor manufacturing processes to monitor tools performance in the process. This approach uses color encoding to represent tools and transition usage. However, all the aforementioned visualizations share the same limitation: they do not categorize and provide a summary of event-log data. This means aggregating event-logs durations into in-time, late, and not-completed completions, and then providing an overview to users who can identify new insights regarding the process, step, and actors' performances. A recent study [71] proposes visualization requirements for business processes, where users can gain a better understanding of the changes in a process. These requirements can be used as guidelines for investigating further how to visualize changes in business processes. Although considerable work [105, 152, 158] has been done in visualizing event-log data, when it comes to visualizing multi-step processes, there is room for improvement. To the best of our knowledge, there is no previous

study, which addresses the issue of summarizing event-log data, and promoting timely completion of multi-step processes.

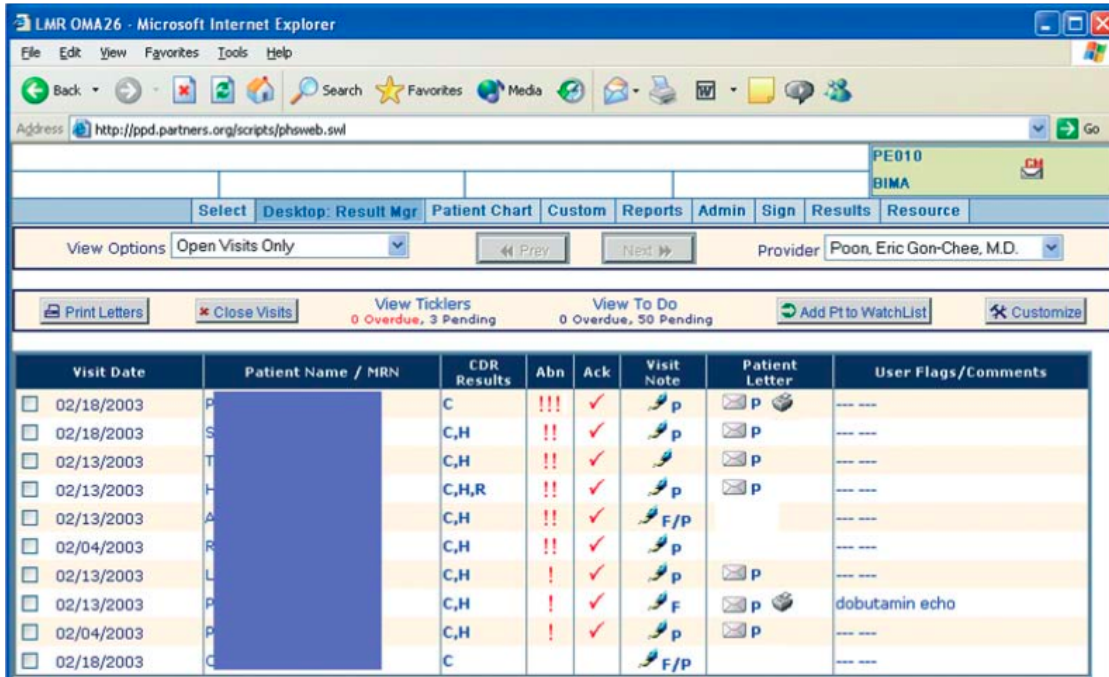
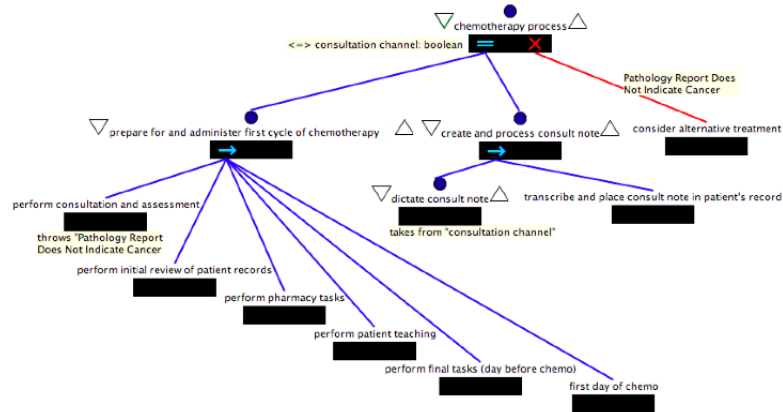
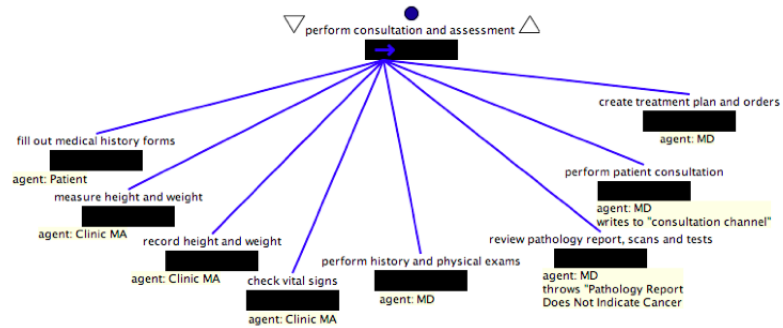


Figure 2.1: Results Manager is a tracking system. *View options:* (1) *Open Visit Only* display all open visits with results that require review. Clinicians may ‘close’ a visit after all follow-up actions for that visit have been completed. (2) *Open and Closed Visits* display all visits with results even if they have been previously ‘closed’ by a clinician. (3) *Visits with User flags* display all visits for which a user had previously placed free-text flags or comments. (4) *Patient Watchlist* displays the panel of patients for which a clinician wants to receive all outpatient results, regardless of who ordered them. (5) *Schedule View* displays the list of patients on a clinician’s schedule for a particular day. *CDR Results* denote that results of various types are available for review. The actual results can be reviewed by clicking the letters in this column. (1) C – Chemistry; (2) H – Hematology; (3) R – Radiology; (4) P – Pathology. *Abn* indicates the degree of abnormality for the most abnormal result associated with the visit. *Ack* is displayed with a checkmark if all results associated with the visit have been explicitly acknowledged by a clinician. *Visit Note* is displayed with a ‘fountain pen’ icon if clinic notes dated around the time of the visit is present. Notes that have been finalized are denoted by an ‘F’. Notes that are still in preliminary form are denoted by a ‘P’. An ‘F/P’ is displayed if both finalized and preliminary notes are present. *Patient Letter* is displayed with a ‘letter’ icon if a result letter has been written through RM. *User Flags/Comments* is a free-text field which gives the clinician the opportunity to annotate visits.

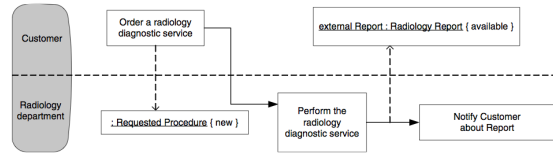


(a) A coordination diagram of chemotherapy process

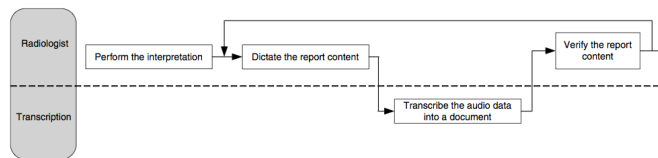


(b) The task decomposition of *perform consultation and assessment*

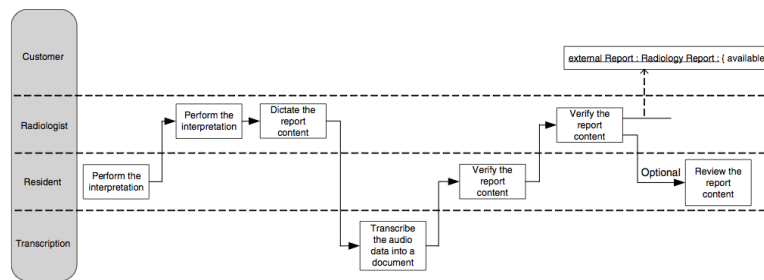
Figure 2.2: The process definition is decomposed into two sub-steps that can be executed in parallel. The root step ‘chemotherapy process’ has a sub-step ‘consider alternative treatment’ that is an exception handler. The first sub-step, ‘prepare for and administer first cycle of chemotherapy’ is decomposed into six sub-steps to be executed in sequence. ‘Perform consultation and assessment’ is a sequential step. It captures information about the agents who execute the tasks in a process. The step ‘perform patient consultation’ writes a parameter to the “consultation channel” and thus it needs to execute before the step ‘dictate consult note’, which reads from the “consultation channel”, can start execution.



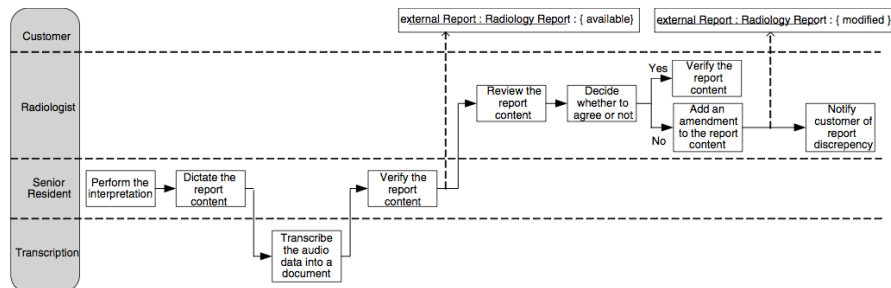
(a) A radiology process



(b) Common workflow (interpretation, dictation, transcription, verification)



(c) Workflow involving a resident, who may verify the content and review the final result



(d) Workflow involving a senior resident, who interprets, dictates and verifies

Figure 2.3: The trigger is an order that is composed of radiology procedure(s). Each radiology procedure results in an external diagnostic report. The radiology customer may be notified at the end of the process about the result availability.

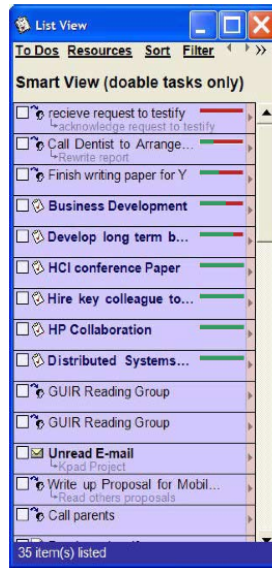


Figure 2.4: TaskVista is a lightweight resource for collecting and listing to-dos and conveniently launching tasks from them. It is a comprehensive to-do list that easily handles a realistic number of active to-dos. Users can easily create a new to-do by typing one in, or, dragging an item (e.g. a file or email) into the list. The (editable) title defaults to the subject or title of a dragged-in item. Additional items such as notes, documents, etc., can be dragged in to a to-do, so the to-do becomes resource for saving content and launching activity on the task, like a pile or folder. But unlike a pile or folder, a to-do has computational properties that support task management. Users can change importance easily by dragging a to-do up and down the list. Old to-dos are filtered out of sight when they become defunct or are done, to avoid clutter. A to-do can have time constraints, other properties that, for example, show location, task or participant dependencies, and whether it is a project. Green ‘warning bars’ turning red are a salient visualization to cue users of the urgency of approaching deadlines.

Incomplete Actions (11)						
Order	Description	Subject	Sender	Created	Modified	Creator
1	Modify Event: Demo M1: Driver Monitoring Systems	Attendance figures and new #	Amy Lim <lim12@ardra.org>	Today, 3:32 PM		RADAR
2	Modify Event	note schedule changes	Spence Pierre <spiero@ardra.org>	Today, 3:54 PM		RADAR
3	Modify Room: Flagstaff: Sternwheeler	Sternwheeler Capacity	Meredith Lorenz <lorenze@pittsburgh-flagstaff.com>	Today, 4:07 PM		RADAR
4	Modify Room: Flagstaff: Sternwheeler	Sternwheeler Capacity	Meredith Lorenz <lorenze@pittsburgh-flagstaff.com>	Today, 4:10 PM		USER
5	Optimize the Schedule	no email		Today, 3:45 PM		RADAR
6	Website Update [VDO]: Modify Person: Austin Florton	Webpage	Austin Florton <aflorton@ardra.org>	Today, 3:37 PM		RADAR
7	Website Update [VDO]: Modify Person	Attendance figures and new #	Amy Lim <lim12@ardra.org>	Today, 3:32 PM		RADAR
8	Website Update [VDO]	Organization Wrong	Sonal Malhotra <small@ardra.org>	Today, 4:32 PM		RADAR
9	Website Update [WEB]	change phone numbers	Emily Halwizer <halwizer@ardra.org>	Today, 4:17 PM		RADAR
10	Place a Vendor Order	Tech. Request - flip charts	Maggie Foxenreiter <mfox@ardra.org>	Today, 3:33 PM		RADAR
11	Send a Briefing	Brief me, please	Jonathon Robertson <jrobertson@ardra.org>	Today, 4:42 PM		RADAR

Overflow Actions (1)						
Order	Description	Subject	Sender	Created	Modified	Creator
1	Reply to Question	Vegetarian options?	Sandra Nubanks <snubanks@ardra.org>	Today, 4:02 PM		RADAR

Completed Actions (1)						
Order	Description	Subject	Sender	Created	Modified	Creator
1	Modify Event: Workshop 14: Intermodal Passenger Screening	Attendance figures	Amy Lim <lim12@ardra.org>	Today, 3:21 PM	Today, 3:45 PM	RADAR

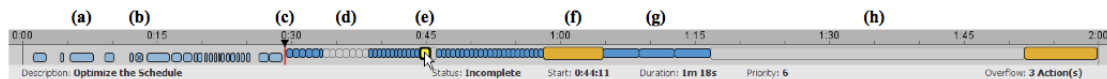
Deleted Actions (1)						
Order	Description	Subject	Sender	Created	Modified	Creator
1	Modify Speaker's Availability	Planning for History Week	Michelle Randal <mich-randal@gmail.com>	Today, 4:20 PM	Today, 4:34 PM	RADAR

Possibly Conference-Related Emails (1)						
Read	Subject	Sender	Date			
<input checked="" type="checkbox"/>	for my presentation	Laura Tindale <laurat2@ardra.org>	Today, 3:24 PM	Add an Action		
Blake, I didnt know who to contact about making sure to have a laptop available, and connected to teh AV equipment - ie projector. I want all that ready on the ...						

Other Emails (1)						
Read	Subject	Sender	Date			
<input checked="" type="checkbox"/>	car arrangements	Angie Randal <angiednacer6@gmail.com>	Today, 3:23 PM	Add an Action		
No K is counting on me to help out with the kids' dance class. The car is still in the shp. Can you drop me off over there? thanks :-)						

Deleted Emails (1)						
Read	Subject	Sender	Date			
<input checked="" type="checkbox"/>	Precipitation Update	Weather Alerts <weather@weather.gov>	Today, 3:56 PM	Add an Action		
There is a 70% probability for thunderstorms with heavy rain in ALLEGHENY COUNTY this evening through tomorrow. Plan accordingly and be safe! Go to www.weather.gov ...						

(a) The Action List provides a task-centric view of an email inbox. The “Incomplete Actions” (a), “Overflow Actions” (b), and “Completed Actions” (c) tables list the tasks contained within email messages, allowing the user to sort by task-centric properties. The three email tables contain emails for which no tasks have been created (e, f, and g).

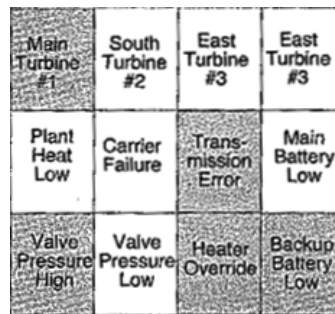


(b) The Progress Bar shows completed (a) and deleted (b) tasks to the left of the current time (c), and the suggested schedule to the right. Noncritical tasks are blue (a, b, and g), critical tasks are orange (f), and expected tasks are gray (d and h). Details about the highlighted task (e) are shown in the status bar at the bottom.

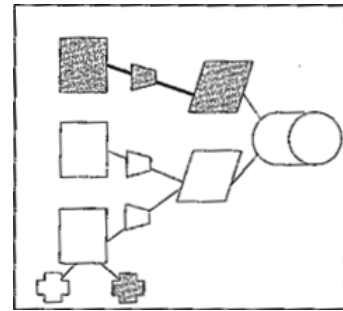
Figure 2.5: The Action List contains seven tables divided into two groups: tasks and emails. The task group contains four tables that list “Incomplete” (a), “Overflow” (b), “Completed” (c), and “Deleted” (d) tasks. Tasks that users have yet to perform are split between the Incomplete and Overflow table, with the latter table containing tasks that users should skip due to time constraints. Tasks completed by users appear in the Completed table, which provides users with a record of their progress and allows them to go back and revisit previous tasks. “Possibly Conference-Related Emails” table contains emails that RADAR thinks may contain tasks but for which it could not confidently identify the exact task type (e). The second table contains other emails that RADAR did not identify as task-related (f). The third table contains emails that users deleted (g).



(a) Serial display



(b) Annunciator-based display



(c) Mimic display

Figure 2.6: *Serial displays*, which list multiple alarms, can be useful in retaining information about the order in which the alarms occurred, but are problematic for operators due to the large number of alarm messages that can build up on the list, hence, impeding to find related alarms and sort through relationships. *Annunciator-based displays* show multiple alarms for a system all at once and help people assess multiple alarms at the same time. As the same alarm is usually displayed in the same location, they can also support human pattern matching. As a downside, however, they are very poor at helping for diagnosis. *Mimic displays* attempt to display alarm signals in a pattern that corresponds to some pictorial representation of the underlying system.

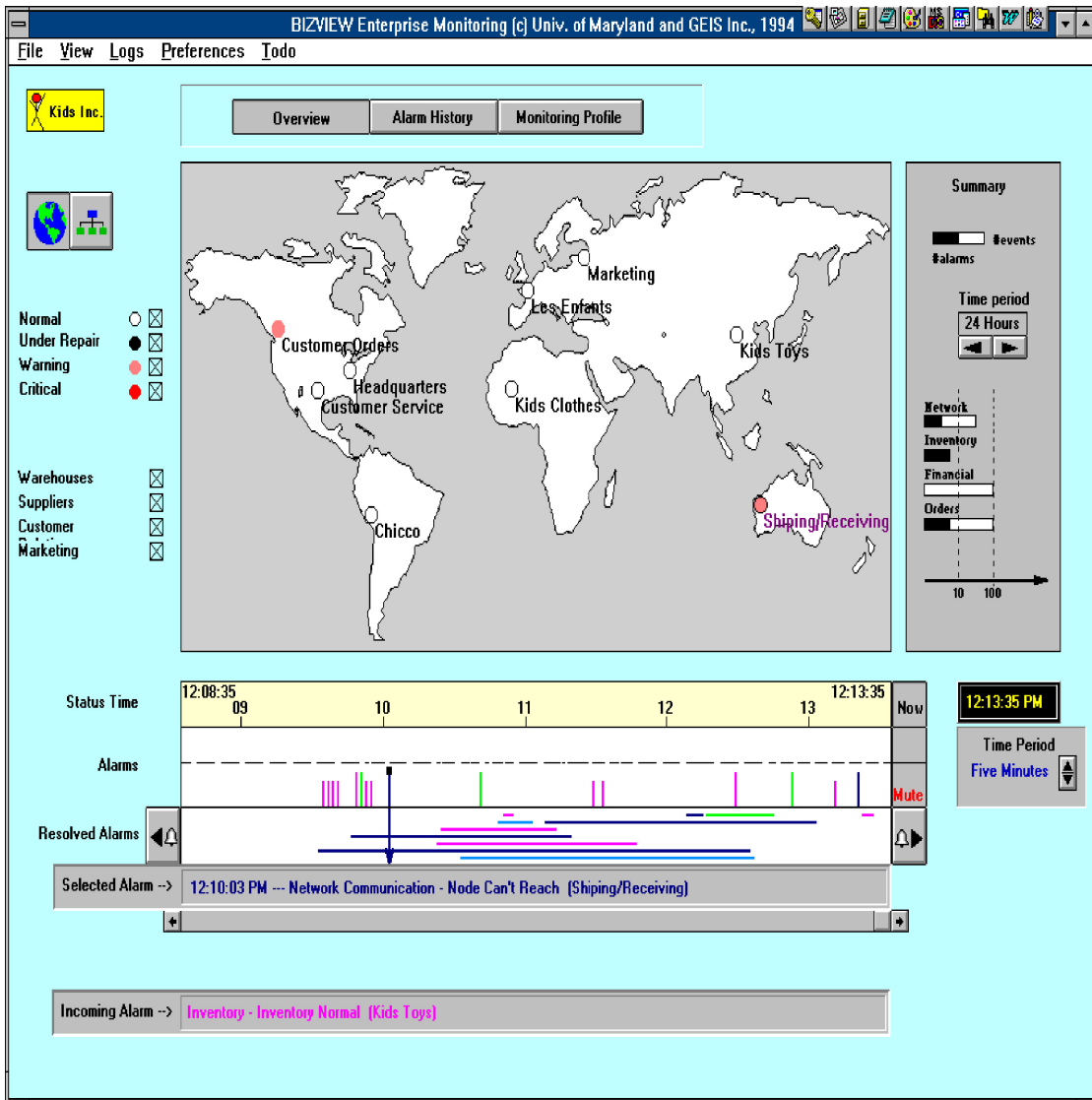


Figure 2.7: The Overview screen includes a status map, node filters (left), and a timeline with a flag for each alarm. A brief description of the latest incoming alarm appears at the bottom. On the status map, bright red dots indicate nodes in critical condition and light red dots represent warnings. Using the filters buttons, nodes can be grayed out or hidden according to desired categories or status. On the timeline, each vertical flag represents an alarm. New alarms first appear on the right side and slide to the left as time passes. The color of the flag identifies the category of the alarm and its height tells the severity. A click on an alarm displays the corresponding text message.

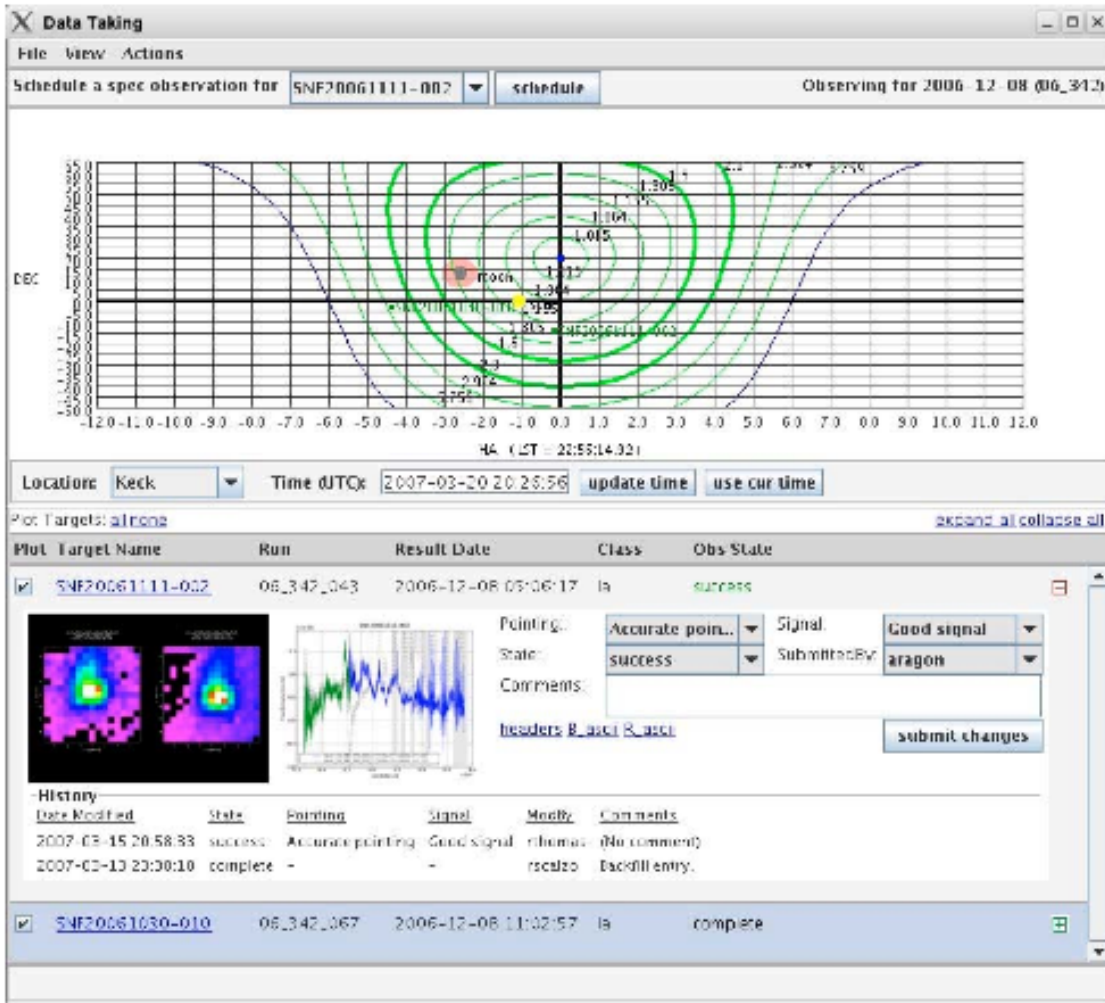


Figure 2.8: The Sky visualization depicts the positions of targets in the sky at a given time and ground location. The green lines represent airmass. The blue line represents the horizon, and the red halo around the moon is the area where light cast by the moon makes it difficult or impossible to view a faint target. The yellow circle represents the sun. Major telescope names and corresponding latitudes and longitudes are displayed on a drop-down menu. The two small images on the lower left-hand side of each observation sub-window are custom visualizations. Color-coding and position indicate the accuracy of the telescope pointing at a target and the signal strength of the received data by the spectrograph. Spectral data are plotted in green and blue. The spiky grey lines depict the spectrum of the background sky. The broad grey bands represent areas of atmospheric absorption.

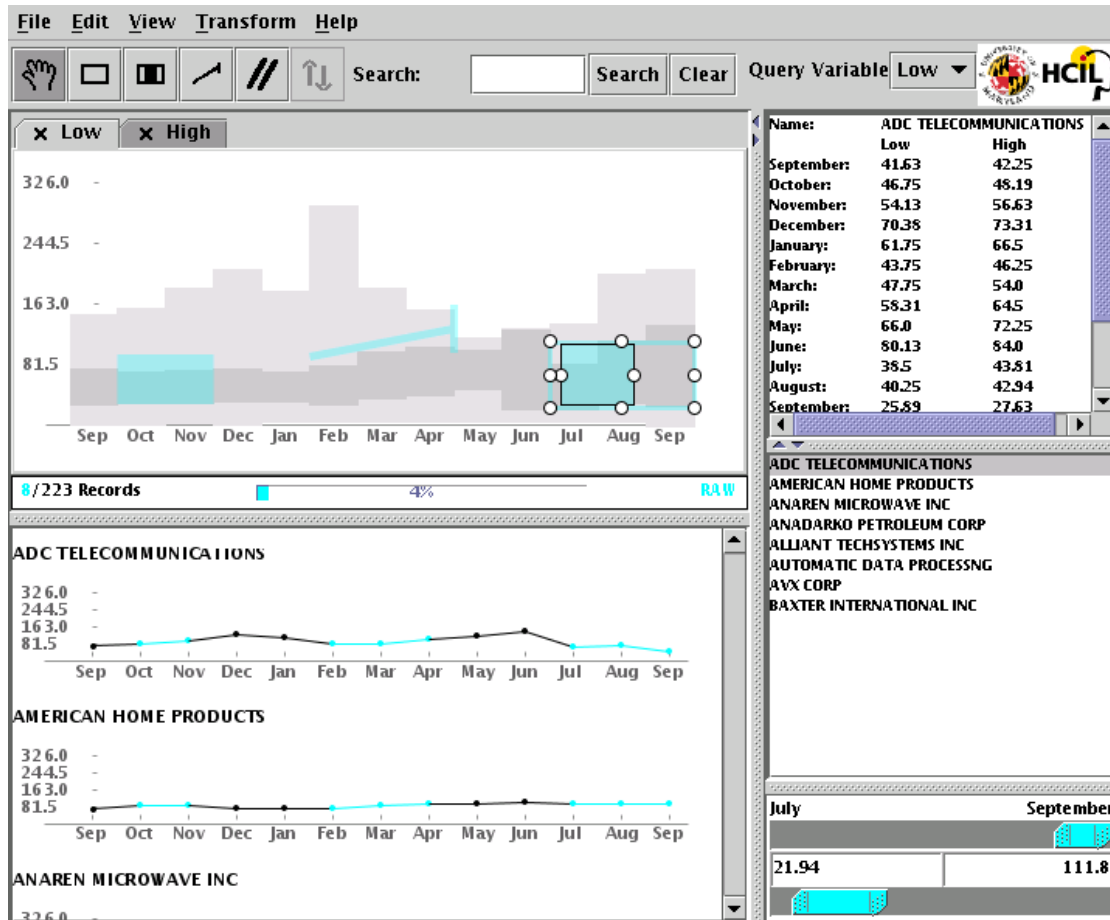


Figure 2.9: A prototype environment for interactive querying and exploration of time-series data. Clockwise from upper-left: query space where timeboxes are drawn (with data envelope, query envelope, and graph overview), details-on-demand for selected items, list of items by name, range sliders for query adjustment, and individual time-series display list, containing the graph display for each item in the data set. These timeboxes support interactive formulation and modification of queries, thus speeding the process of exploring time-series data sets and guiding data mining. Other features include drag-and-drop support for query-by-example and graphical envelopes for displaying the extent of the entire data set and result set from a given query.

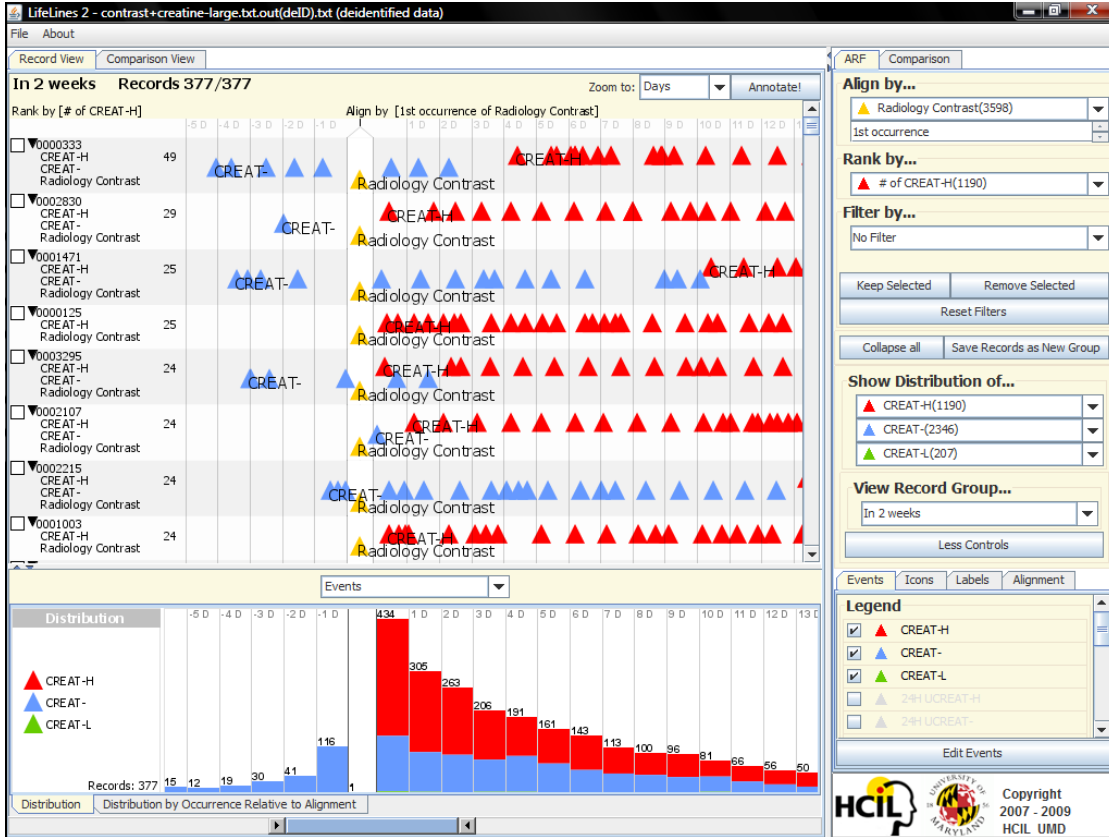


Figure 2.10: Each row represents a record, showing its ID on the left. Each record contains several types of events, listed below its ID. Each type is color-coded, and each instance of event is represented by a colored triangle on the timeline. The time increases from left to right. The white vertical band represents the alignment line (aligned by Radiology Contrast in orange triangles). The combo box to the top right indicates that each finer tick represents a day. A temporal summary is shown in the bottom. The label on the left indicates that it is showing the distribution of the event types CREAT-H, CREAT, and CREAT-L. Each bar indicates how many events of that type there are in that day. The combo box indicates that we are aggregating events. Below is a range slider that can be used to pan and zoom. To the right is the control panel. It shows the controls and the current state of Align, Rank, and Filter (ARF). Lower part of the right panel has additional controls for temporal summary, and also for navigating groups.

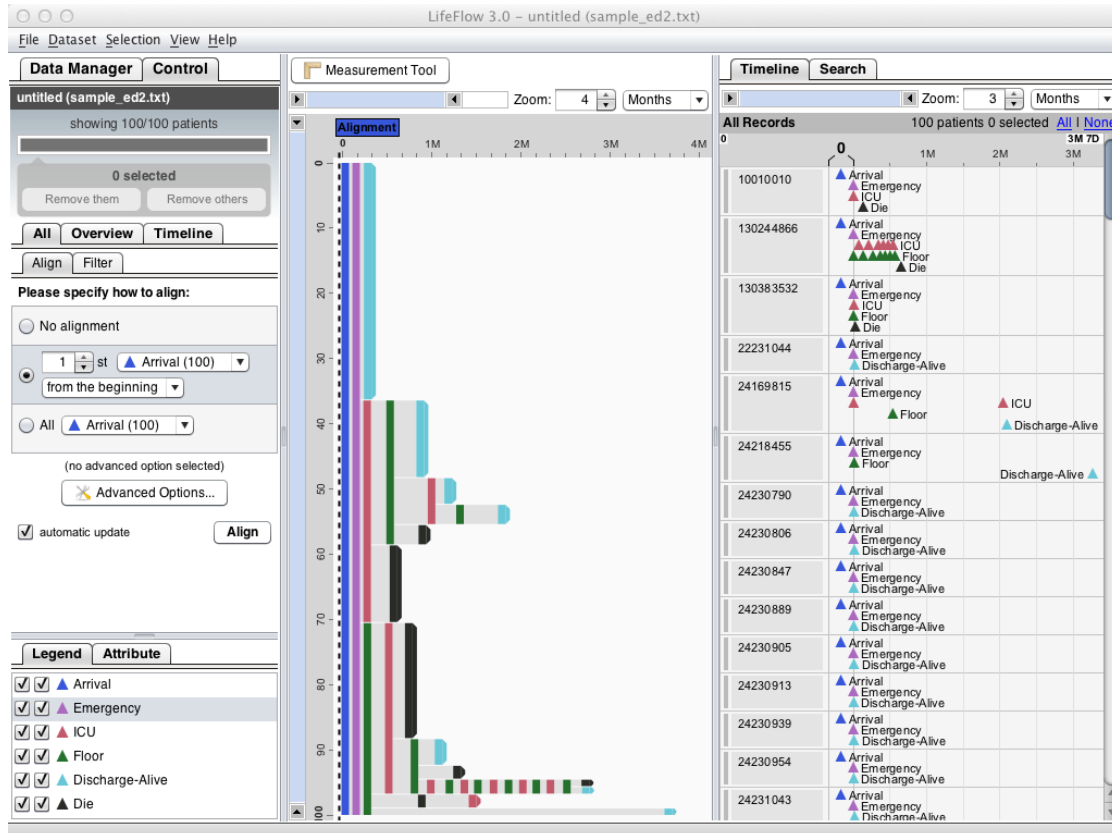


Figure 2.11: This screenshot of LifeFlow shows a random sample of patient transfer data based on real de-identified data. Here, LifeFlow is used side-by-side with LifeLines2 so that individual records can be reviewed by scrolling. When a user clicks on a sequence in LifeFlow, the sequence is highlighted and all corresponding records are also highlighted and moved to the top in LifeLines2, allowing the user to examine them in more details. The way to read sequences in LifeFlow is to read the colors (using the legend). The horizontal gap between colored bars represents the average time between events. The height of the bars is proportional to the number of records, therefore, showing the relative frequency of that sequence. The bars with same parent are ordered by frequency (tallest bar on top). To find the most frequent pattern is to find the tallest bar at the end. Here it shows that the most common sequence is Arrival, Emergency then Discharge alive. Surprisingly, two patients were reported dead before transferred to ICU, which indicates a data entry problem.

Chapter 3

Timely Management of Medical Results: MStart

“Because technology is continually changing, we will never have a stable and complete set of guidelines, but scientific studies will have enormous benefits in terms of reliability and the quality of decision making about user interfaces. Design processes, ethnographic methods, participatory design activities, scenario writing, and social impact statements are evolving.” – Shneiderman (2010)

The first problem I tackled was learning the challenges in medical results management and devising ideas to reduce the time needed for clinicians to track and act on their medical orders. My observations and interviews lead to the discovery that while result correctness is of paramount concern, delays can be equally dire; delays can result in late or incorrect diagnoses, which may yield repeat orders, dissatisfied patients, or death. MStart (Multi-Step Task Analyzing, Reporting, and Tracking) is a system I built to allow professionals, who actively handle results, to promptly track and act upon their orders.

3.1 Requirements

I visited 7 hospitals and clinics in the United States to understand the medical workflow and to learn about the challenges clinicians face while using current Electronic Health Record (EHR) systems for results management. Throughout

two years, I interviewed primary care physicians, internists, assistants, nurses, residents, and epidemiologists in outpatient clinics, health centers, emergency departments, and medical centers using electronic systems MicroMD, Application-Manager, Azyxxi, AllScripts, Epic, and Cerner. This section summarizes my observations and findings. More details can be found in Appendix A.

There were many observations about the problem areas and possible improvements to system designs. The major findings were as follows. I used my observations to devise use case scenarios for results management. It was confirmed that certain results do not return back (20% of X-Rays in one case) and the consequences of failures are heartbreaking (as patients get physically harmed). Handling results takes a significant amount of time. Tables are commonly used for results management but they are not designed well. Fixing table designs can reduce the issues in results management significantly. Most tables are filled with too much and different types of information. Especially when pending orders are not shown, they increase the chance of missing truly important data. There is a need to implement more than one table that are linked and optimized for results management tasks. Deadlines and delays are usually not captured in current systems. Because lateness is not electronically available, clinicians report spending hours tracking the results down. Most systems do not readily provide relevant information as they are used for recording (mostly text-based) but not for automatic computations. Clinicians have to do many calculations in their head, which leads to mistakes especially by novice users. Personal checklists and patients serve as order reminders. Taking follow-up actions in EHR systems is cumbersome. The interactive features need new mechanisms for ensuring that the follow-up is complete in a timely fashion.

3.2 The User Interface of MStart

Based on the observations and interviews, MStart (Multi-Step Task Analyzing, Reporting, and Tracking) was designed and built to help medical staff, such as physicians, residents, or clinic assistants who are responsible for timely results management. MStart system could eventually be integrated into EHR systems. This section explains the user interface of MStart, while its implementation details, regarding the workflow model and the underlying system architecture will be discussed in Appendices B and C, respectively.

MStart is a domain-independent system. There are many other application domains that multi-step processes are tracked. For example, in academia, potential applications include monitoring the progress of undergraduate and graduate students in different departments, or reminding researchers of scientific publication acceptances. Business processes of enterprises can be tracked in a similar manner. To support the complexity of the results management process with multiple people, there were four scenarios built in to MStart. This section will explain the first scenario in detail and will briefly describe other scenarios. MStart requires users to login and depending on their role, they see a different screen and data pertaining to their responsibilities. By default users only see the work assigned to them.

1. *Physician, Joe Brown, reviews results from his inbox several times a day.* In the observations, clinicians preferred to see current orders placed for all of their patients. The user interface of MStart has three tabs for three screens (1) order, (2) track, and (3) complete as can be seen in Figure 3.1, with the tab “Track” being selected.
2. *Resident, Bob Green, looks at the list of orders made during Isabel Bailey’s*

MSTART File

MSTART Simulator Physician Joe Brown reviews results from his inbox several times a day Current Time: Friday, March 11, 2011 4:46 PM Pause Continue

Riverside Clinic Provider: Brown, Joe Sign out

Order Track Complete

Results to Review (15) Search All New Late Not Completed

Patient	Test	Order Date	Review By	Abnormality
Parker, Amy	MRI / P	Mar 2, 2011 11:42 PM	Mar 9, 2011 11:42 PM	
Reed, Megan	Mammogram / P	Mar 3, 2011 9:55 AM	Mar 10, 2011 9:55 AM	
Stewart, Kim	Mammogram / F	Mar 3, 2011 10:09 AM	Mar 10, 2011 10:09 AM	
Richardson, Allison	Blood (CBC) / F	Mar 7, 2011 11:07 PM	Mar 15, 2011 12:07 AM	⚠
Richardson, Allison	Blood (BMP) / P	Mar 7, 2011 11:07 PM	Mar 15, 2011 12:07 AM	⚠
Bennett, Claire	Pap Smear / F	Mar 7, 2011 8:01 PM	Mar 14, 2011 9:01 PM	
Reed, Megan	Dietetics / P	Mar 7, 2011 8:43 AM	Mar 14, 2011 9:43 AM	
Howard, Amanda	Blood (BMP) / P	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	⚠
Howard, Amanda	Blood (CBC) / P	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	⚠
Phillips, Sarah	Pap Smear / P	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Cox, Pamela	Blood (BMP) / F	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	
Cox, Pamela	Blood (CBC) / F	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	

Pending Test Results (6) All Late Not Completed

Patient	Test	Order Date	Result Due	Status
Morris, Monica	TSH	February 15, 2011 2:51 PM	February 21, 2011 2:51 PM	Sample drawn and sent
Parker, Amy	MRI	March 2, 2011 11:42 PM	March 8, 2011 11:42 PM	Image(s) interpreted and reported
Peterson, Audrey	Blood (BMP)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered
Peterson, Audrey	Blood (CBC)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered
Reed, Megan	Mammogram	March 3, 2011 9:55 AM	March 9, 2011 9:55 AM	Image(s) interpreted and reported
Sanders, Chelsey	Pap Smear	March 3, 2011 4:54 AM	March 9, 2011 4:54 AM	Specimen collected and sent

Planned Tests (10)

Figure 3.1: Tracking screen of the user interface of MStart as seen by a care provider, Joe Brown, in Riverside Clinic (order and complete tabs not shown). The tables show arrived, pending, and planned orders. Results that have returned to the physician’s office are listed at the top in “Results to Review”, while orders that are in progress and not returned to the physician are shown under “Pending Test Results”. Orders that are placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red. Abnormal results are indicated with a warning sign.

last visit in the patient record before seeing her. Some clinicians, who handle results, may like to see current orders per patient. They still see the same user interface as in the previous scenario (Figure 3.2) and can perform the same follow-up actions. However, the lists are less populated as they belong to one patient only.

3. *Assistant, Jennifer Young, checks if all the recent orders for patients coming today are back or not to track them down.* Assistants are responsible for checking if all current orders for patients, who have visits today, are returned. The pending list is more important for them (see Figure 3.3). Although assistants can not order, follow-up on, or complete orders, they still need to look at the pending orders and call the people, who are processing the orders, to speed up the progress.
4. *Clinic manager, Debra Barnes, oversees all the orders from the clinic.* Clinic managers oversee all current orders for patients treated in the clinic. The lists include too many orders that might be difficult to understand from the tracking screen. Thus, managers are provided with a retrospective analysis screen that facilitates analysis of all past orders.

3.2.1 Order

In the order screen (Figure 3.5), clinicians place orders, see the estimated normal and max duration of a test. They may set a later date for their order if the test is planned in advance or change the expected date of orders, if desired. Note that this is a much simplified order screen since ordering is not the focus of this work. They can also indicate which laboratory facility they want the patient to visit.

MSTART File

MSTART Simulator Resident Bob Green looks at the list of orders made during Isabel Bailey's last visit in the patient chart before seeing her. Current Time: Sunday, March 13, 2011 3:06 AM [Pause] [Continue]

Riverside Clinic Resident: Green, Bob Patient: Bailey, Isabel [Sign out]

[Order] [Track] [Complete]

Results to Review (7) Search [All] [New] [Late] [Not Completed]

Patient	Test	Order Date	Review By	Abnormality
Bailey, Isabel	Dietetics / P	February 12, 2011 3:52 PM	February 19, 2011 3:52 PM	
Bailey, Isabel	MRI / F	March 8, 2011 9:03 AM	March 15, 2011 10:03 AM	
Bailey, Isabel	Mammogram / F	March 10, 2011 9:46 PM	March 17, 2011 10:46 PM	
Bailey, Isabel	Pap Smear / P	March 10, 2011 9:46 PM	March 17, 2011 10:46 PM	
Bailey, Isabel	TSH / F	March 8, 2011 6:44 AM	March 15, 2011 7:44 AM	
Bailey, Isabel	Blood (BMP) / F	March 10, 2011 2:41 PM	March 17, 2011 3:41 PM	
Bailey, Isabel	Blood (CBC) / F	March 10, 2011 2:41 PM	March 17, 2011 3:41 PM	

Pending Test Results (1) [All] [Late] [Not Completed]

Patient	Test	Order Date	Result Due	Status
Bailey, Isabel	Dietetics	February 12, 2011 3:52 PM	February 18, 2011 3:52 PM	Preliminary Report prepared

Planned Tests (0)

Figure 3.2: Tracking screen of the user interface of MStart as seen by a resident, Bob Green, in Riverside Clinic (order and complete tabs not shown). The tables show arrived, pending, and planned orders for the patient, Isabel Bailey. Results that have returned to the clinic are listed at the top in “Results to Review”, while orders that are in progress and not returned yet are shown under “Pending Test Results”. Orders that are placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red. Abnormal results are indicated with a warning sign and moved to the top of their section.

MSTART File

Assistant Jennifer Young checks if all the recent orders for patients coming today are back or not to track them down

Current Time: Sunday, March 13, 2011 11:26 AM

Pause Continue

Riverside Clinic Assistant: Young, Jennifer Patient: Phillips, Sarah; Reed, Megan; Wood, Courtney Sign out

Track

Results to Review (9) Search

Patient	Test	Ordered By	Order Date	Review By	Abnormality
Reed, Megan	Mammogram / P	Brown, Joe (P)	Mar 3, 2011 9:35 AM	Mar 10, 2011 9:35 AM	
Phillips, Sarah	TSH / F	Brown, Joe (P)	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Phillips, Sarah	Pap Smear / F	Brown, Joe (P)	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Phillips, Sarah	MRI / F	Green, Bob (Res)	Mar 6, 2011 9:31 PM	Mar 13, 2011 10:31 PM	⚠
Wood, Courtney	MRI / F	Green, Bob (Res)	Mar 6, 2011 3:50 PM	Mar 13, 2011 4:50 PM	⚠
Wood, Courtney	X-Ray / F	Green, Bob (Res)	Mar 7, 2011 4:30 PM	Mar 14, 2011 5:30 PM	⚠
Reed, Megan	Pap Smear / F	Green, Bob (Res)	Mar 6, 2011 9:46 PM	Mar 13, 2011 10:46 PM	
Reed, Megan	Dietetics / P	Brown, Joe (P)	Mar 7, 2011 8:43 AM	Mar 14, 2011 9:43 AM	
Wood, Courtney	Blood (CBC) / P	Green, Bob (Res)	Mar 7, 2011 7:45 AM	Mar 14, 2011 8:45 AM	

Pending Test Results (4)

Patient	Test	Ordered By	Order Date	Result Due	Status
Phillips, Sarah	X-Ray	Green, Bob (Res)	3/7/11	3/13/11	Exam scheduled
Reed, Megan	Mammogram	Brown, Joe (P)	3/3/11	3/9/11	Image(s) interpreted and reported
Reed, Megan	Dietetics	Brown, Joe (P)	3/7/11	3/13/11	Preliminary Report prepared
Wood, Courtney	Blood (CBC)	Green, Bob (Res)	3/7/11	3/13/11	Preliminary Results recorded

Planned Tests (1)

Figure 3.3: Tracking screen of the user interface of MStart as seen by an assistant, Jennifer Young, in Riverside Clinic. The tables show arrived, pending, and planned orders for patients, Sarah Phillips, Megan Reed, and Courtney Wood, who have visits today. Results that have returned to the clinic are listed at the top in “Results to Review”, while orders that are in progress and not returned yet are shown under “Pending Test Results”. Orders that are placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red. Abnormal results are indicated with a warning sign.

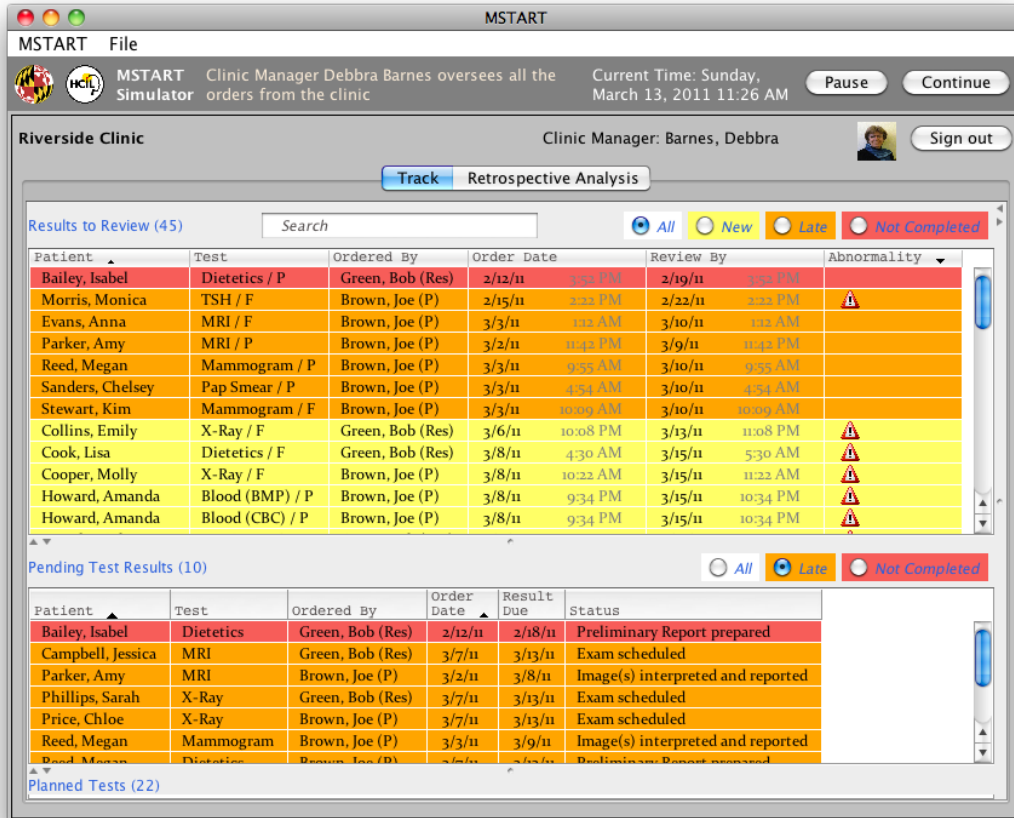


Figure 3.4: Tracking screen of the user interface of MStart as seen by a clinic manager, Debbra Barnes, in Riverside Clinic (retrospective analysis tab not shown). The tables show arrived, pending, and planned orders for all patients treated in the clinic. Results that have returned to the clinic are listed at the top in “Results to Review”, while orders that are in progress and not returned yet are shown under “Pending Test Results”. Orders that are placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red. Abnormal results are indicated with a warning sign.

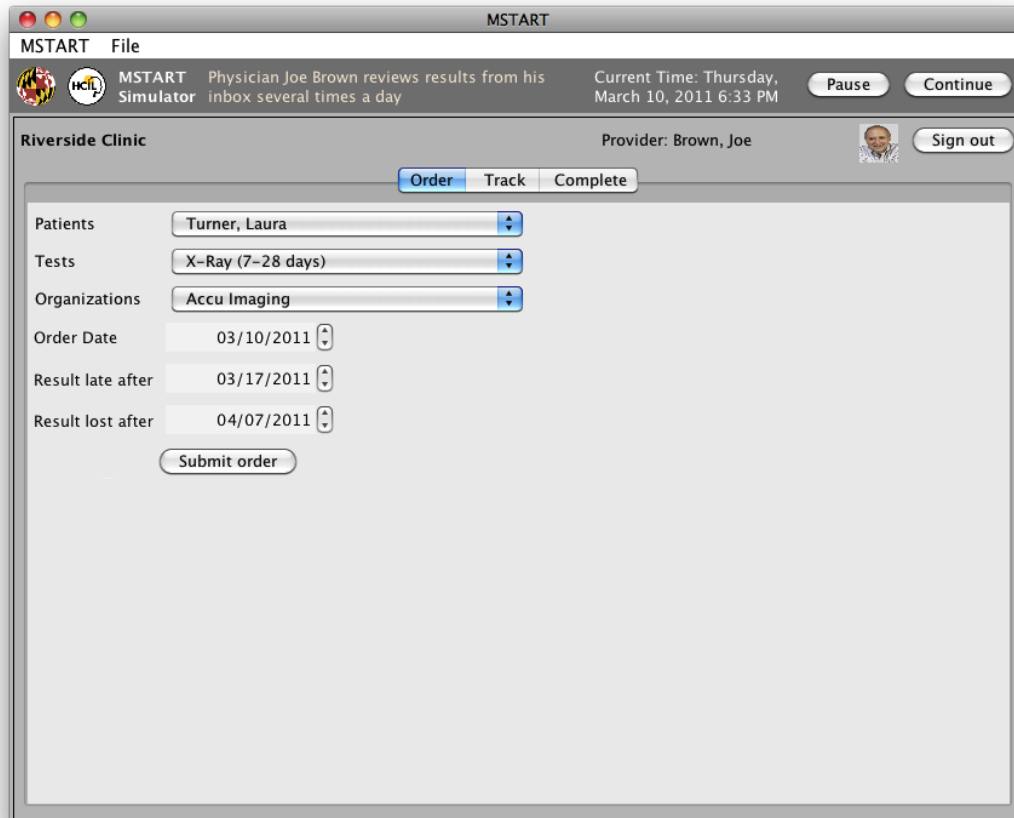


Figure 3.5: Care Provider, Joe Brown, in Riverside Clinic orders an X-Ray from Accu Imaging for patient, Laura Turner, with the default (7-28 days) estimated normal (March 17th) and maximum (April 7th) dates.

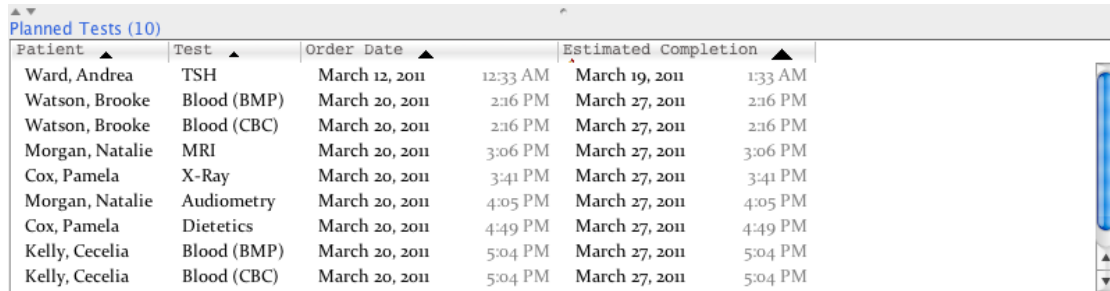
3.2.2 Track

This section presents rich tabular displays and ARC (Actions for Rapid Completion). Both of these techniques when combined help overwhelmed and distracted clinicians manage their results rapidly. Since the requirements analysis proved that lists are the most common format for managing clinical data, I focused on rich tables to show the progress of orders. Rich tables are generated by MStart (Multi-Step Task Analyzing, Reporting, and Tracking) [135] from a workflow model that defines processes, composed of multiple tasks that are instrumented with appropriate user actions.

- *Results, Pending, and Planned Tables:* Figure 3.1 shows the “Track” screen of an ordering medical provider (Joe Brown, as shown on the upper right corner, in Riverside Clinic, indicated on the upper left). Results that have returned to the physician’s office are listed at the top in “Results to Review”, while orders that are in progress and not returned to the physician are shown under “Pending Test Results”. Orders that have been placed already but will take effect in the future (e.g. a mammogram at patient’s fiftieth birthday) can be accessed in the “Planned Tests” (collapsed here).

The results table shows patient name, test name, order date, review by date, and abnormality information. Abnormal results are indicated with a warning sign. The columns for pending orders are patient name, test name, order date, result date, and status. The planned orders (Figure 3.6) have the patient name, test name, order date, and estimated completion date. On the date for when the planned orders were placed, orders automatically move from planned to the pending table. Similarly, as the results of pending orders arrive, the entry in the pending table moves up to the topmost table. Results

are removed once the physician has reviewed the report and confirmed that the follow-up is complete.



Patient	Test	Order Date		Estimated Completion	
Ward, Andrea	TSH	March 12, 2011	12:33 AM	March 19, 2011	1:33 AM
Watson, Brooke	Blood (BMP)	March 20, 2011	2:16 PM	March 27, 2011	2:16 PM
Watson, Brooke	Blood (CBC)	March 20, 2011	2:16 PM	March 27, 2011	2:16 PM
Morgan, Natalie	MRI	March 20, 2011	3:06 PM	March 27, 2011	3:06 PM
Cox, Pamela	X-Ray	March 20, 2011	3:41 PM	March 27, 2011	3:41 PM
Morgan, Natalie	Audiometry	March 20, 2011	4:05 PM	March 27, 2011	4:05 PM
Cox, Pamela	Dietetics	March 20, 2011	4:49 PM	March 27, 2011	4:49 PM
Kelly, Cecelia	Blood (BMP)	March 20, 2011	5:04 PM	March 27, 2011	5:04 PM
Kelly, Cecelia	Blood (CBC)	March 20, 2011	5:04 PM	March 27, 2011	5:04 PM

Figure 3.6: Planned table has orders written for a future date. They will be started processing on that date.

Some results appear in both pending and results tables such as Amy Parker’s preliminary MRI. Although the outside facility generates an initial report for physicians to see, the order has not been finished processing by the facility and the outside facility is still responsible for finalizing the report. The letter ‘P’ or ‘F’ in the results table distinguishes preliminary from final results (Figure 3.1).

- *Result Management Workflow:* MStart employs an underlying result management workflow model that assigns a normal and a maximum duration to each step in an order. For example, a patient might be given three to seven days to schedule and go to get an X-Ray. MStart can then calculate a normal and maximum expected duration along each step of the process e.g. one to three weeks (and physicians can overwrite the normal duration at order time if they want rapid returns). After the normal expected duration – reflected in the “Result Due” date – the order is determined late and shown orange (e.g. Amy Parker’s MRI is still being processed in Figure 3.1). After the maximum time has passed, the order is considered lost or not completed (shown

as red) and the order may have to be repeated or cancelled (e.g. the TSH of Monica Morris in Figure 3.1). The time to complete each step is logged for later retrospective analysis to inspect bottlenecks as well as the best and worst performers [135], and adjust normal and maximum expected durations. The current implementation simulates how the system would work in a real setting as orders become active by advancing the time rapidly. On the top right of Figure 3.1, there are pause and continue buttons to stop and start the simulation.

The step to be completed by physicians (i.e. review and follow-up) also has a normal and maximum duration. The result list at the top has a “Review By” date and physicians’ work might be late (orange) or not completed (red). For example, Kim Stewart’s Mammogram is late (Figure 3.1). Orders that have not been looked at yet are yellow (e.g. Allison Richardson’s blood tests), while the white ones have been reviewed but the follow-up is not complete yet (e.g. Amanda Howard’s blood tests).

The color legend helps users learn the color-coding and acts as a filter, e.g. the pending table is pre-filtered to show only severe (late and lost) cases. These colors are customizable through a preference dialog (Figure 3.7). Extra filters are available on the right of the screen by expanding the split pane (Figure 3.8). Users might be interested in seeing orders per patient, test, or date and extra filters provide such functionality. When his colleague, Bob Green, goes on vacation, Joe Brown can see his results by checking from ‘Ordered By’ filter. Due dates can be modified directly in the table, if appropriate.

All tables are sorted by default so as to visually aid users in seeing important

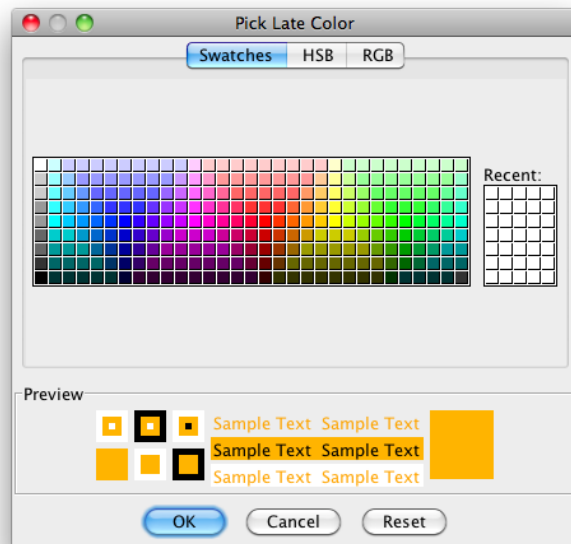


Figure 3.7: By clicking on the late label in the track screen, users can pick a new late color with this dialog.

results at the top: (i) results are sorted first by lateness, secondly by whether they have been seen, thirdly by abnormality status, and finally by patient name to group results of the same patient together. (ii) Pending orders are sorted by lateness first, then by patient name. (iii) Planned tests are shown with first priority given to estimated completion time (sooner ones go up in the table). Arrows on column headers indicate sort direction and priority, which is customizable with a click on the header. In Figure 3.1, Kim Stewart's late mammogram is close to the top of the table, while Allison Richardson's unviewed blood results come next, then comes unviewed normal tests, after that viewed abnormal blood tests of Amanda Howard appear, and finally viewed normal blood tests of Pamela Cox are listed.

- *Responsibility*: Pending orders have a column for order status (see Figure 3.1)

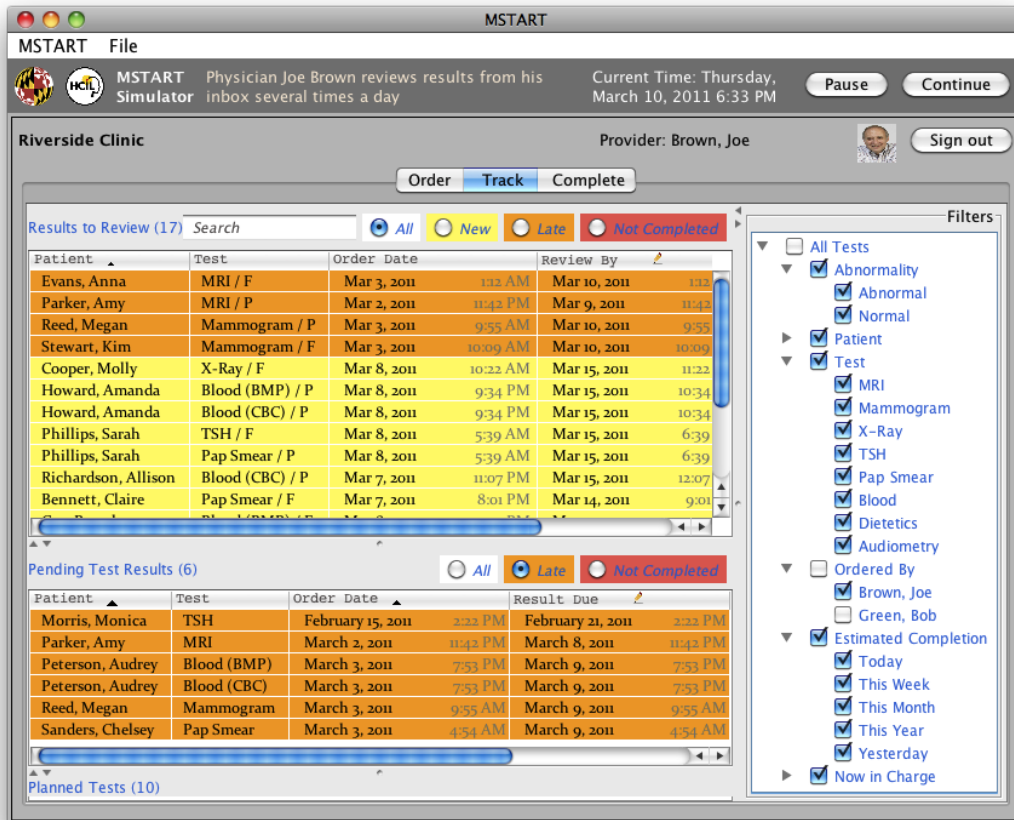


Figure 3.8: Filters pane open on demand and present users with options to widen or narrow their criteria to see more and less data, respectively.

indicating the last completed step. A click on the row brings a popup menu (Figure 3.9) showing who is currently responsible for the progress of the order and the expected completion time. For late orders, there is information on who to call to speed up the process. When the order is considered lost, a reorder button appears¹. The border color matches row status. In Figure 3.9, we observe that the dietitian consultation of Isabel Bailey is lost. We can see that Eric Robinson is responsible for finalizing the report. The menu shows that he is a dietitian in Monroe Health center, whose manager is Cynthia Long with contact number, ext.384. This allows physicians (or their assistant) to see who is responsible for the lateness and to take action by calling their supervisor or reordering this consultation.

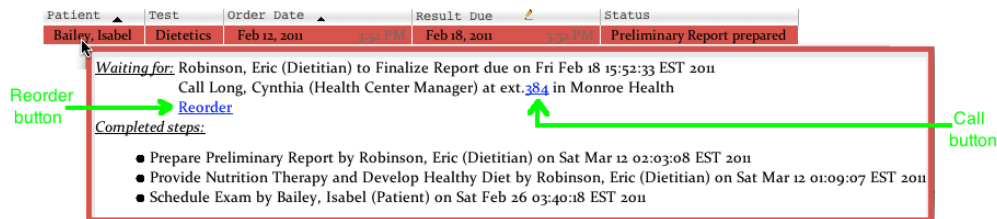


Figure 3.9: Popup menu for pending orders identifies the responsible person (dietitian, Eric Robinson) who is handling the current step of the order (finalizing the report) along with their deadline (February 18th, 2011) and manager’s contact information (Cynthia Long’s phone extension is 384). Below, it has a reorder button for the lost consultation and illustrates the progress of the order step by step.

In addition, the popup enumerates the completed steps in chronologically descending order. The first step of each pending order is the patient scheduling the exam so the patient is the first responsible person (see the last item of “Completed steps” in Figure 3.9). The next steps involve the outside facility processing the order (see the first two bullets under “Completed steps” in

¹Similarly, a cancel button could be provided.

Figure 3.9).

- *Actions for Rapid Completion (ARC)*: When physicians or residents click on a result, the result report and simple follow-up actions appear side-by-side, below the row (Figure 3.10). This is called Actions for Rapid Completion (ARC). If more information is needed to deal with a complex case, a double-click will open the result in the patient record. In other cases, the panel of common simple actions is easily accessible and stays on the screen until users indicate either they need to come back and further “Review Later” (the result remains in the list and is colored white), or that the follow-up is “Complete” (the result is removed from the list and goes to a separate “Complete” panel in Figure 3.11).

Patient	Test	Order Date	Review By	Abnormality
Evans, Anna	MRI / F	Mar 3, 2011 1:12 AM	Mar 10, 2011 1:12 AM	

No Follow-up
 Ask Nurse to Inform Patient (Evans, Anna)
 Schedule Visit
 Repeat Test

Order: 1 days weeks months years

Comments:

Review Later Complete

NAME: Evans, Anna
STUDY DATE: 03/09/2011 4:28:29 PM
COMPARISON: None available
HISTORY: Pain
MRI OF THE RIGHT ANKLE (WITHOUT IV CONTRAST)
FINDINGS:
 The bones demonstrate no fracture or gross marrow abnormality. Articular surfaces are within normal limits.
 The spring ligament complex appears to be intact.
 The tibialis posterior, flexor digitorum, and flexor hallucis longus tendons are intact. The peroneus brevis and longus tendons are intact. The Achilles tendon is intact.
 The tarsal sinus is unremarkable. There are no significant fluid collections to suggest ligamentous injury.
 The remaining muscles and soft tissues are unremarkable.

Figure 3.10: Actions for Rapid Completion (ARC) list the most common actions for routine and normal results. When the user left clicks on the rows in any of the rich tables, ARCs appear. In the context of rich tables, ARCs are close to the mouse-click location and results are on the right. “Complete” finalizes follow-up so the result could be removed from “Results to Review” table. On the other hand, “Review Later” button means the result was viewed and/or some actions may have been taken but the follow-up is not complete yet. Such a result is still kept in “Results to Review” table for further processing but is marked as “viewed”.

This prototype integrates follow-up actions with result review. Possible ac-

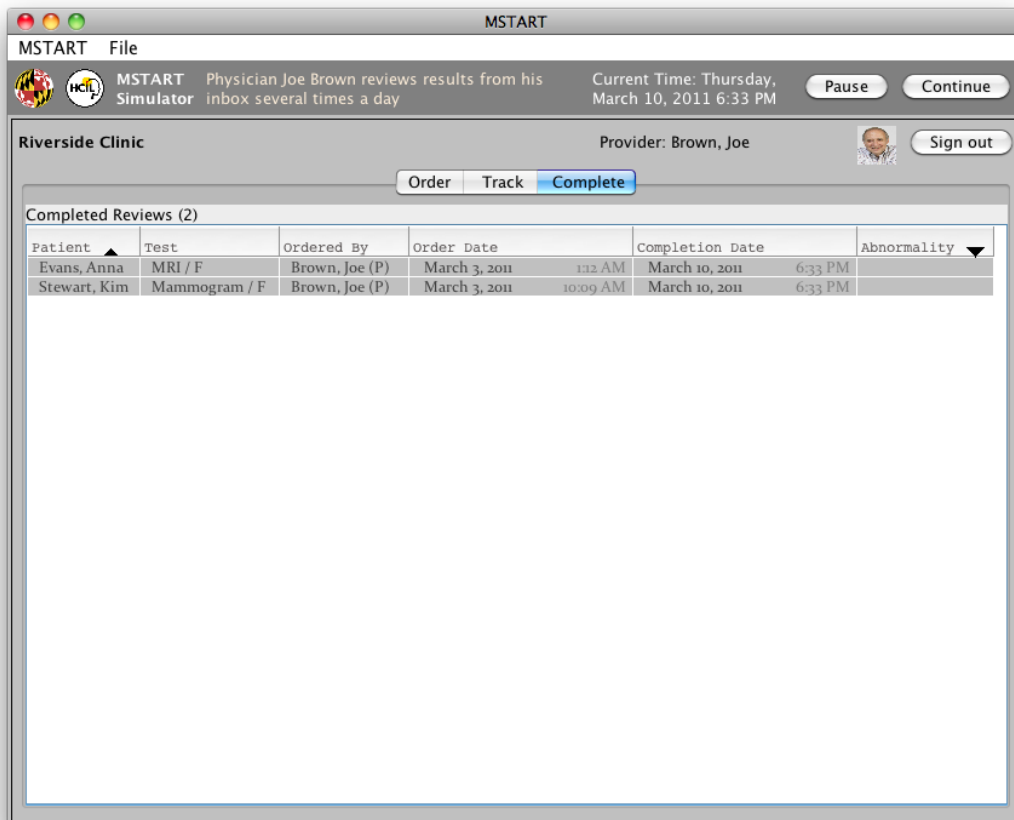
tions depend on the role of the logged in user. Figure 3.10 introduces a list of most common actions for normal and routine cases. It is expected to improve efficiency and reduce memory load to have the most frequent actions as one-click buttons. To dispose the panel, users click elsewhere on the screen or hit ‘esc’ key. Disposing the panel does not affect anything in cases when the physician unintentionally opened a result by mistake or wants it to remain as unviewed. From Figure 3.10, physicians can pick a predefined reason from a list of options if no follow-up is necessary. Physicians can ask their assistant to inform the patient and/or schedule a visit. Another possible action is to simply repeat the current test. In Figure 3.10, only informing the patient, Anna Evans, is chosen.

3.2.3 Complete

Once physicians click “Complete” in ARCs, a sound confirms the completion and the result moves to the “Complete” tab (Figure 3.11). This panel displays all recently completed reviews. “Completed Reviews” table indicates the details of the result (i.e. patient name, test name, ordered by, order date, completion date, abnormality information). Completed reviews are sorted by abnormality and patient name.

3.3 Summary

This chapter presented the lessons learned in observations and interviews conducted with medical professionals at seven hospitals and clinics. The main results of this requirements gathering were that there is a workflow around results management, lists are the conventional widgets that clinicians interact with to manage



The screenshot displays the MSTART application window. The title bar reads 'MSTART'. Below the title bar, the application name 'MSTART Simulator' is shown alongside a logo. The status bar indicates 'Physician Joe Brown reviews results from his inbox several times a day' and 'Current Time: Thursday, March 10, 2011 6:33 PM'. There are 'Pause' and 'Continue' buttons. The main content area is titled 'Riverside Clinic' and shows 'Provider: Brown, Joe' with a 'Sign out' button. Below this, there are three tabs: 'Order', 'Track', and 'Complete', with 'Complete' being the active tab. The 'Completed Reviews (2)' table lists the following data:

Patient	Test	Ordered By	Order Date	Completion Date	Abnormality
Evans, Anna	MRI / F	Brown, Joe (P)	March 3, 2011 1:12 AM	March 10, 2011 6:33 PM	
Stewart, Kim	Mammogram / F	Brown, Joe (P)	March 3, 2011 10:09 AM	March 10, 2011 6:33 PM	

Figure 3.11: Completed reviews table lists all orders that are reviewed and followed up from the results table in Track screen.

their results, and follow-up is not ensured. Employing an underlying workflow model, improving current table designs and interactions can help resolve the issues significantly.

The user interface of MStart incorporated those findings into design. MStart helps clinicians, who are responsible for medical results management, track their orders in a timely manner. It relies on a workflow model that exposes responsible parties, utilizes a table design that attempts to reduce delayed orders, and integrates actions into the list to facilitate rapid follow-up.

My requirements analysis and the design ideas used in MStart are helpful to academics, researchers, and vendors, who work on medical informatics and especially, results management. The design ideas can be applied to other domains where a workflow is present, lists are commonly used, and actions have to be taken frequently. Examples include but are not limited to academic, software, and business processes.

Chapter 4

Evaluation

“Each experiment has two parents: the practical problems facing designers, and the fundamental theories based on principles of human behavior and interface design. Each experiment also has three children: specific recommendations for the practical problem, refinements of theories, and guidance for future experimenters.” – Shneiderman (2010)

My research is driven by the needs of my target users to support their daily work. Therefore, I followed a user-centered approach of iterative design and evaluation for MStart (Multi-Step Task Analyzing, Reporting, and Tracking), which is built to help medical staff manage their results in a timely fashion. Accordingly, iterative design reviews with medical professionals, a controlled experiment, and two other user studies were conducted. Iterative design reviews allowed to learn medical expert suggestions and comments on my prototypes and general design principles. A controlled experiment evaluated two of the interface design choices, showing pending orders and prioritizing by lateness information. Finally, a series of evaluations compared ARC (Actions for Rapid Completion) to current systems to discover differences in terms of the number of steps and interactions, and the time to execute representative tasks.

4.1 Iterative Design Reviews with Medical Professionals

Following a user-centered approach of iterative design and evaluation, I often engaged with clinical staff and system developers as they became available. I conducted regular meetings with clinicians who provided feedback on MStart from November 2010 to June 2011 and from January to May, 2012. Approximately forty medical experts provided feedback at fifteen different events for an estimated total of about twenty hours of review and discussions over eight months.

4.1.1 Method

Two iterative design reviews were employed using the participatory design [50] method. To obtain rapid feedback on the basic principles during the initial stages of the project, presentations were prepared for a group of people (thirty or more professionals), who are knowledgeable in medicine, human-computer interaction, or software engineering areas, followed by question/answer sessions.

As the project evolved and the prototypes became more stable, the second reviews happened. These were more structured design reviews with the project collaborators in Houston, Texas. There were six such meetings – one onsite and five teleconference style – between January and May 2012:

- January 25th, 2012 (onsite)
- March 13th, 2012
- March 27th, 2012
- April 10th, 2012
- April 24th, 2012
- May 29th, 2012

Each of them lasted between 30 minutes to two hours. There were three to ten people in each meeting and at least one physician (up to four physicians) present. These meetings started with a live demonstration of my prototype (with an interactive executable distributed). Then, there were questions about recommendations to improve the interface design and interactions, as well as my evaluation plans. Comments, emails, screenshots, or reports were recorded. Moreover, a graphic designer helped improve the aesthetics of my tool. Four emails (on 09/29/11, 03/20/12, 03/29/12, and 04/11/12-04/12/12) included mockups and screenshots from other systems regarding the buttons, tooltips, fonts, panels, interactions, etc. These guided my development to refine the prototype.

4.1.2 Results

All experts approved my process model and the calculation of expected durations, giving strong support to the explicit statement of responsible agents. Most comments were related to the improvement to the model itself: (i) initially nurses in the clinic receive the results and if necessary, they are in charge of distributing results to physicians, (ii) an ordering physician can send the report to his resident or trainee within the clinic to check first, (iii) a clinic manager can check whether arrived results are acted upon by care providers in a timely manner to ensure quality control. These all account for the routing within the clinic after the result has arrived. By adding a couple of lines to the model and setting the preferences of the user interface, these suggested improvements were accommodated.

During a discussion with a medical doctor, a draft of improvements to the workflow model was prepared and the medical examples were elaborated. These yielded extended linear and parallel processes. The physician requested improve-

ments to the model to capture reflex tests, where one test could automatically lead to a second test. In addition, the distinction between preliminary and final results were discussed in detail. Finally, it was explained that abnormal results sometimes take longer to process, and the physician discussed how results were not processed at the same speed every weekday so calculations of temporal aspects should take this into account. Different facilities may have different factors that influence how fast they return results, and this value might be adjusted over time to offer better predictions. Users could even adjust the factors through retrospective analysis. These were all accommodated in the second version.

There were many suggestions about the interface and interactions. One reviewer encouraged to improve the aesthetics of the interface so a graphic designer helped develop a better interface and interaction designs. Terminology used on screens has been updated several times based on comments from medical experts. Medical experts were able to point me to resources to obtain real reports of results. There were also refinements to the interface such as adding or removing functionality to make its intended use clear. For instance, after presenting my idea of responsible agents, it did not make sense to see only the name of a person. Discussions lead to assigning a manager in the model that gets notified and displaying the manager's contact information along with the responsible person's role. While the initial prototype contained only the name of the person, medical experts recommended that it would be more useful to see the current step of the order and the steps already completed for the order. The incorporation of another table for planned orders was completed after another physician's question.

The idea of a retrospective analysis for managers was well received and found to be helpful to quickly quantify general problems. However, domain experts

expected to see how late the results were besides the fact that some were late. This led to the design of a visualization, which shows the number of results and their lateness information together. To see more information was useful and most domain experts were interested in seeing more data to track (as physicians are responsible for hundreds of orders daily) and analyze. As a result, an anonymized dataset for a medical process was used later on.

4.1.3 Discussion

These reviews showed the problem was better understood from the point of view of my target user population. Approval or revisions of my design ideas by the clinicians led to the refinements of my prototype to address more issues the clinicians faced with their current systems. Also, my project borrowed ideas from computer science, such as human-computer interaction and software engineering. By advice of researchers in these domains, up-to-date standards and technologies were used.

As reported in the previous section, most recommendations were incorporated. While this section listed the main suggestions that made their way into MStart, there were some challenging ones that were considered but left out. For example, one interesting proposal was to have quality attributes built into the system so as to compute some metrics for later analyses. Another idea was to use machine learning techniques (by taking into account any attribute associated with an order) to adjust result time constraints accordingly. With this advanced configuration, the system continuously learns new factors from past behavior, e.g. work hours (morning hours may get less work done compared to afternoon hours) or seasons (summer months may be slower because staff is on vacation). Finally, asking clinicians to enter they have completed each step of the process can eventually

become a burden. It was requested to implement automated tracking (by barcodes and RFID tags) to further reduce the clinical staff workload, such as those used in commercial tracking (e.g. Fedex).

The implication of the results is creating a system flexible enough to accommodate different needs of clinicians. These meetings allowed me to brainstorm about the ideas not yet considered in the current version and expand the prototype to allow more functionality that could be useful in medical results management. Between each meeting, new functionality was added based on the previous meeting. Some suggestions were more difficult to implement than others and took longer. Seeing their comments in MStart encouraged clinicians to more openly provide feedback and sometimes even alter their prior suggestions.

One limitation of these reviews is that the clinicians merely helped design MStart, but did not test the system. Their feedback was invaluable to build a system close to reality, however, they did not participate in a formal evaluation where they used MStart for their results management. This is because MStart was built as an inspirational prototype for vendors, who develop full electronic health record systems, to adopt some of these features, which will then be used by clinicians.

4.2 Controlled Experiment for Awareness of Order Timeliness

In this experiment¹, participants took on the role of physicians and answered questions about the timeliness of orders using three interface variations. The goal of

¹This work was done in partnership with Computer Science Master's student, Lyndsey Franklin.

the experiment was to quantify the benefit of the first two medical guidelines, i.e. showing pending orders and prioritization by late and lost status. The experimenter recorded the time to arrive at the correct answer and the number of corrections they had to make. In addition to a \$10 compensation, a bonus \$10 was offered to the best performers in each interface. The experiment focused on participant awareness of late and lost orders and not on participants' ability to review the results of orders that have been returned. Two pilot experiments were run with 4 participants each and a full experiment was run with 18 participants. The subjects were recruited by verbal advertisements, through emails to mailing lists and by paper fliers.

4.2.1 Question

The experiment is designed to answer the following question: With respect to the awareness of orders placed, are there statistically significant differences in the time needed by participants to correctly answer questions about order statuses?

4.2.2 Design

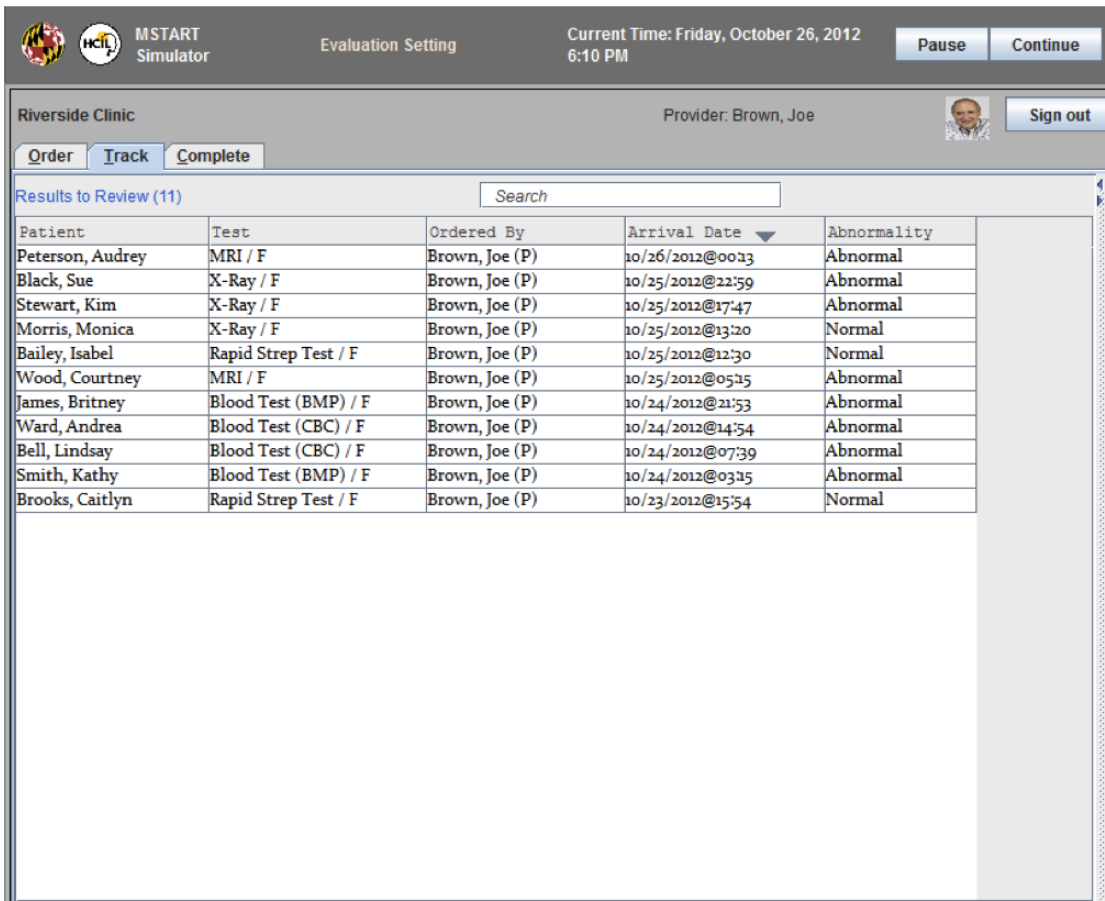
This experiment follows a 1×3 within-subjects design with a baseline interface and two incrementally elaborate interfaces.

4.2.3 Variables

The independent variable is the support for awareness of the status of results available through the prototype interfaces. The dependent variables are the time it took for participants to correctly answer the questions and the number of corrections they had to make. This section describes the user interfaces in more detail and

how specific features distinguish the level of the independent variable.

1. **Interface 1 (Baseline):** Interface 1 (Figure 4.1) represents a control condition. With Interface 1, participants are presented with a single list of traditional chronologically-ordered results. Orders with no results are not represented on this baseline form but the complete list of patient orders are available on the orders list. This baseline/control condition represents the situation where participants have no additional support for awareness of the orders beyond a single sorted list.



The screenshot shows the MSTART Simulator interface for Riverside Clinic. The top bar includes the MSTART Simulator logo, 'Evaluation Setting', and the current time: Friday, October 26, 2012, 6:10 PM. There are 'Pause' and 'Continue' buttons. The main area shows 'Riverside Clinic' and 'Provider: Brown, Joe' with a 'Sign out' button. Below this are tabs for 'Order', 'Track', and 'Complete'. A search bar is present above a table titled 'Results to Review (11)'. The table lists 11 patient records with columns for Patient, Test, Ordered By, Arrival Date, and Abnormality.

Patient	Test	Ordered By	Arrival Date	Abnormality
Peterson, Audrey	MRI / F	Brown, Joe (P)	10/26/2012@00:13	Abnormal
Black, Sue	X-Ray / F	Brown, Joe (P)	10/25/2012@22:59	Abnormal
Stewart, Kim	X-Ray / F	Brown, Joe (P)	10/25/2012@17:47	Abnormal
Morris, Monica	X-Ray / F	Brown, Joe (P)	10/25/2012@13:20	Normal
Bailey, Isabel	Rapid Strep Test / F	Brown, Joe (P)	10/25/2012@12:30	Normal
Wood, Courtney	MRI / F	Brown, Joe (P)	10/25/2012@05:25	Abnormal
James, Britney	Blood Test (BMP) / F	Brown, Joe (P)	10/24/2012@21:53	Abnormal
Ward, Andrea	Blood Test (CBC) / F	Brown, Joe (P)	10/24/2012@14:54	Abnormal
Bell, Lindsay	Blood Test (CBC) / F	Brown, Joe (P)	10/24/2012@07:39	Abnormal
Smith, Kathy	Blood Test (BMP) / F	Brown, Joe (P)	10/24/2012@03:25	Abnormal
Brooks, Caitlyn	Rapid Strep Test / F	Brown, Joe (P)	10/23/2012@15:54	Normal

Figure 4.1: Baseline interface presents a single list of traditional chronologically-ordered results.

2. **Interface 2 (Baseline + Pending):** Interface 2 (Figure 4.2) represents the second level of the independent variable and adds an additional list to the baseline interface. This additional list contains orders that have been placed but have not yet returned any results. It does not contain any additional information beyond the order submitted. This condition represents the situation where participants have minimal support for awareness of the orders they have placed.

The screenshot shows the MSTART Simulator interface for 'Riverside Clinic'. The top bar includes the MSTART Simulator logo, 'Evaluation Setting', 'Current Time: Friday, October 26, 2012 6:10 PM', and 'Pause' and 'Continue' buttons. The main header shows 'Riverside Clinic', 'Provider: Brown, Joe', and a 'Sign out' button. Below the header are tabs for 'Order', 'Track', and 'Complete'. The main content area is divided into two sections: 'Results to Review (11)' and 'Pending Test Results (9)'. The 'Results to Review' section contains a table with columns for Patient, Test, Ordered By, Arrival Date, and Abnormality. The 'Pending Test Results' section contains a table with columns for Patient, Test, and Order Date.

Patient	Test	Ordered By	Arrival Date	Abnormality
Green, Bobbie	Blood Test (CBC) / F	Brown, Joe (P)	10/25/2012	Normal
Parker, Amy	Rapid Strep Test / F	Brown, Joe (P)	10/25/2012	Abnormal
Wood, Courtney	X-Ray / F	Brown, Joe (P)	10/25/2012	Abnormal
Bell, Lindsay	Rapid Strep Test / F	Brown, Joe (P)	10/25/2012	Normal
Edwards, Rachel	Blood Test (BMP) / F	Brown, Joe (P)	10/24/2012	Abnormal
Smith, Kathy	Blood Test (BMP) / F	Brown, Joe (P)	10/24/2012	Normal
Campbell, Jessica	Blood Test (CBC) / F	Brown, Joe (P)	10/24/2012	Abnormal
Murphy, Sabrina	MRI / F	Brown, Joe (P)	10/24/2012	Abnormal
White, Jane	Blood Test (BMP) / F	Brown, Joe (P)	10/24/2012	Abnormal
Richardson, Allison	Blood Test (CBC) / F	Brown, Joe (P)	10/24/2012	Abnormal
James, Britney	Rapid Strep Test / F	Brown, Joe (P)	10/24/2012	Abnormal

Patient	Test	Order Date
Bell, Lindsay	Dietetics	10/24/2012
Cook, Lisa	Dietetics	10/24/2012
Watson, Brooke	Blood Test (CBC)	10/23/2012
Ward, Andrea	Rapid Strep Test	10/23/2012
Parker, Amy	Dietetics	10/23/2012
Richardson, Allison	Rapid Strep Test	10/22/2012
Doe, Jo	Blood Test (BMP)	10/22/2012
Edwards, Rachel	X-Ray	10/22/2012
Edwards, Rachel	Dietetics	10/22/2012

Figure 4.2: Baseline + Pending interface has a pending orders list in addition to the results list.

3. **Interface 3 (Baseline + Pending + Timeliness):** Interface 3 (Figure 4.3) represents the third level of the independent variable and is an interface with two lists of orders, one for results and one for pending, prioritized by the timeliness of orders rather than the arrival or submission date. Late orders appear at the top of each list followed by orders still on schedule. This condition represents a situation where awareness of orders is directly supported by the interface.

The screenshot shows the MSTART Simulator interface for Riverside Clinic, Provider: Brown, Joe. The interface is divided into two main sections: 'Results to Review' and 'Pending Test Results'.

Results to Review (12)

Patient	Test	Order Date	Review By	Abnormality
Kelly, Cecelia	MRI / F	October 22, 2012	October 23, 2012	Abnormal
Evans, Anna	Blood Test (CBC) / F	October 23, 2012	October 24, 2012	Abnormal
Black, Sue	Blood Test (BMP) / F	October 23, 2012	October 24, 2012	Abnormal
Rogers, Emma	Blood Test (BMP) / F	October 23, 2012	October 23, 2012	Abnormal
Bailey, Isabel	Blood Test (BMP) / F	October 23, 2012	October 24, 2012	Abnormal
Cooper, Molly	MRI / F	October 23, 2012	October 24, 2012	Abnormal
Collins, Emily	Rapid Strep Test / F	October 24, 2012	October 24, 2012	Abnormal
Campbell, Jessica	Blood Test (CBC) / F	October 23, 2012	October 24, 2012	Abnormal
Richardson, Allison	Rapid Strep Test / F	October 23, 2012	October 23, 2012	Normal
Edwards, Rachel	MRI / F	October 23, 2012	October 24, 2012	Normal
Richardson, Allison	Blood Test (BMP) / F	October 22, 2012	October 23, 2012	Normal
Collins, Emily	Blood Test (BMP) / F	October 23, 2012	October 24, 2012	Normal

Pending Test Results (8)

Filter: All Late Lost

Patient	Test	Order Date	Status
Bennett, Claire	Blood Test (CBC)	October 23, 2012	Test ordered
Cook, Lisa	X-Ray	October 22, 2012	Test ordered
Sanders, Chelsey	Blood Test (CBC)	October 22, 2012	Test ordered
Bell, Lindsay	Audiometry	October 23, 2012	Test ordered
James, Britney	Audiometry	October 22, 2012	Exam scheduled
Watson, Brooke	Audiometry	October 22, 2012	Exam scheduled
Collins, Emily	Dietetics	October 24, 2012	Test ordered
Phillips, Sarah	Dietetics	October 24, 2012	Exam scheduled

Figure 4.3: Baseline + Pending + Timeliness interface prioritizes orders by timeliness in addition to the lists of pending orders and results.

4.2.4 Protocol

This 1×3 within-subjects study employs the following basic protocol discussed in greater detail in this section. The primary steps are listed here (steps 3-6 are repeated for each of the remaining two interfaces):

1. Participants arrive and sign a consent form.
 2. Participants read a short description of their task.
 3. Participants read and remember a list of orders and normal and maximum durations for up to 5 minutes.
 4. Participants take a break and complete a distraction task for 5 minutes (so that they did not remember the details).
 5. Participants are shown the first interface empty with no orders reflected. There is a brief explanation of this empty interface so that participants have a chance to ask questions.
 6. Participants use a populated interface to identify which orders are late (i.e. longer than normal) and which are lost (i.e. exceeding procedure time limits).
- **Scenario:** “Imagine that you are a primary care physician and that you have to order tests (blood work, MRI, X-Ray, etc.) for your patients. Patients go to outside facilities to get their tests performed. These outside facilities send back test results to you that can be normal or abnormal. Sometimes test results are delayed and arrive late or tests are lost and do not return at all. To properly care for your patients, you need to keep track of test results and quickly determine which tests have had results returned and which results are late or lost. In addition to ordering tests, you also do your daily work such as seeing patients which is represented as a brief intermission in this study.

We will give you a list of tests ordered for your patients and you will have 5 minutes to read the list and ask questions. For the next 5 minutes, we will ask you to help us design a search for medical records by drawing a representation of the search. We will then show you some screen prints and ask you two questions about the tests ordered for your patients: 1) which tests are late, and 2) which tests have been lost. All the information you need to answer these questions will be available on the interfaces you work with and you will have as much time as you need to arrive at a correct answer. If you make a mistake, we will tell you and you will have more time to correct your answer. You may not take notes of any kind during this study. There are three interfaces so we will repeat this process three times. The whole study should take about 1 hour.

For your participation in this study, you will receive \$10. Participants with the fastest time for a given interface will receive an additional \$10 prize, up to \$30. Please be aware: if you make a mistake and we have to ask you to correct an answer there will be a time penalty which will increase your time.”

- **Reviewing Orders & Durations:** Participants were given a chance to examine a list of orders, ask questions about the list, and if they were able, remember details from the list for answering questions later. Participants were informed that they would be answering questions about the orders on the list but given no guidance as to what details would be helpful to remember. This was intended to motivate participants to retain some information from the orders in memory. Participants were provided with a new list of patient orders for each interface to prevent accidental learning of patient orders as the study progressed. In addition, participants were presented with a

list of expected durations for the order types in the list. This reference time requirements sheet was the same throughout the study. No note-taking was allowed during the experiment.

In contrast, during the first pilot experiment, participants were given a list of orders to submit for their patients and asked to use a working order-interface to place orders. However, the ordering process took a great deal of time and that participants consistently felt that they did not understand the list of patient orders any better after having completed the ordering process. Further, participants did not develop strategies for remembering details of the orders by ordering themselves. The expectations regarding participants would develop familiarity and remember details of the list of patient orders for each interface were not met. Additionally, the ordering process required more time to complete than the time needed by participants to answer the study questions. There was no benefit to the ordering process so it was removed from the study. In its place, participants were given a list of orders placed for their patients and a maximum of 5 minutes to read the list and familiarize themselves with it.

- **Distraction Task:** After reviewing the list of orders, participants had an intermission during which they performed a different activity. This intermission period was intended to provide participants with time to forget the details of the patient orders they had just reviewed so that their reliance on the interfaces would be more pronounced as they answered questions. Participants were monitored to assure they did not use this time for taking notes on the patient orders they had reviewed or rehearsed any details of the orders in an attempt to remember the details better.

- **Timed Session:** Questions about the status of orders were the same for each participant and assessed the participants' ability to determine the status of orders from the information available through the different interfaces. Questions about the status of orders included:

- Which in-process tests are overdue/late for results?
- Which tests have been lost?

Participants were given as much time as they needed to completely and correctly answer each question. The experimenter checked the answers of participants as they worked and noted any errors made by participants. If a participant made a mistake, it was noted. After the participant arrived at an answer, the experimenter alerted the participant how many and what types of errors there were. The participant continued working until there were no errors in their answers. Timing for each question began when the question had been asked and stopped when the participant had provided a fully correct answer.

There were some other questions considered and one of them was about abnormality. The second pilot experiment showed that there really was no variation in people's ability to answer "Which are abnormal?" so questions about the abnormal results were removed.

Another difference regarded the number of times the questions were directed. In the first pilot experiment, participants were asked these questions twice. The first time, they were given an extremely brief time limit of 10 seconds. Participants were then asked how confident they felt about their answers. After this, they were given as much time as they required to correctly answer the question. However, it was determined that the brief limit of 10 seconds

for the first questions was too short to show significant differences between interfaces 1 and 2. For the second pilot experiment, this time was increased to 15 seconds. With this second iteration, participants were continuing to have difficulties quickly answering the study questions in the brief 15 seconds allotted. Two participants indicated that they could not answer the question in the time allowed and did not try. Additionally, the confidence reported by participants for late and lost tests were grouped together, low, and with little significant differences. Since the process was frustrating to participants, the 15 second quick answers and the subjective confidence ratings were removed. Instead, participants for the full study only indicated which orders from their list were late and which were lost.

Because the brief questions from the study were removed, participants would be instructed to continue working on their answers until completely correct in future iterations. When participants provided an answer, the study team reviewed the participant's answer and indicated how many orders have been incorrectly marked as late or lost. If participants have unmarked late or lost answers, they were told and asked to continue until all late or lost tests have been marked. The time required for participants to correctly answer and the number of corrections required from the study team were tracked in order to answer the research questions.

4.2.5 Experimental Setup

- **Training:** Training was not required because all interactions were removed. The simplifying assumption was that extensive training was not needed in order for participants to learn the interfaces. Instead of a full tutorial with

sample tasks for each interface, participants were instead given an opportunity to ask questions about an empty representation of an interface (Figure 4.4) after a brief description of the interface’s features. It was anticipated that participant performance would speed up as the study progressed due to learning effects, i.e. questions carried over from one interface to the next in the study session. The order of interface appearance was counterbalanced and assigned randomly such that each interface appeared an equal number of times as first, second, and last to participants.

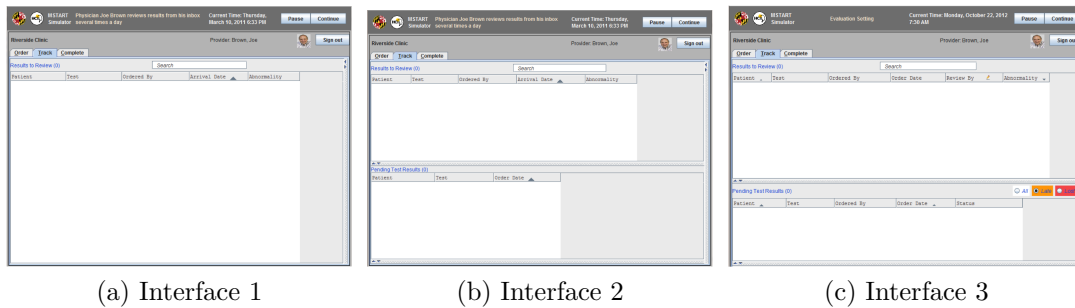


Figure 4.4: During training, participants were given empty interfaces with no orders to explain interface features.

- **Interface Session:** Each participant performed the same order review and question answering phases for each of the three interface conditions. The order of appearance for each interface was randomized to counter-balance for learning effects of both interface features and question answering experience. Participants first reviewed the orders for each interface. After the distraction task, participants were presented with one of the interface conditions and the order status questions. Participants were allowed to ask general questions about the interface but were not provided with detailed explanations of interface contents. Each interface session was repeated with the same protocol

for each of the three interface conditions.

Study scenarios were sufficiently complex with only the initial order list reflected in the interfaces. Orders and results appearing in each interface when participants answered questions were limited to only the orders that the participants saw prior to the question answering phase. There were no extra orders or results appearing in the interfaces to act as distractions.

In the second pilot experiment, the orders, which were scheduled over the course of three days in the study, were tracked by the hour in the interfaces given to participants. However, the instructions told participants to only use days when determining whether or not an order was late or lost. Both the lists of orders and the printed interfaces were revised to remove any resulting ambiguities and to ensure that questions to participants would have clear answers. Any indication of hour from the interfaces was also removed.

Adjustments to time limitations and the number or order types were made as a result of pilot study iterations. In order for participants to be able to answer questions, the time limits and size and complexity of patient-orders used were adjusted so that participants were able to answer all questions in all interface conditions.

- **Additional Materials:** Participants were provided with a printed list of orders (Table 4.1, 4.2, 4.3) while they answered questions. They were also given a reference (Table 4.4) which provided typical time requirements for the variety of orders placed. This time requirements reference was to be used by the participants when determining whether or not tests orders were still in process, in process but late, or lost. Participants were not allowed any other reference materials and no-note taking of any kind was allowed during

the study.

Table 4.1: The list of orders given to participants for Interface 1.

10/22/2012

Bell, Lindsay	Blood Test (CBC)
Brooks, Caitlyn	Rapid Strep Test
Evans, Anna	Blood Test (BMP)
Smith, Kathy	Blood Test (BMP)

10/23/2012

Bailey, Isabel	Dietetics
Bailey, Isabel	Rapid Strep Test
Black, Sue	X-Ray
Brooks, Caitlyn	X-Ray
Doe, Jo	MRI
James, Britney	Blood Test (BMP)
Parker, Amy	Audiometry
Peterson, Audrey	MRI
Richardson, Allison	Dietetics
Stewart, Kim	X-Ray
Ward, Andrea	Blood Test (CBC)
Wood, Courtney	Blood Test (CBC)
Wood, Courtney	MRI

10/24/2012

Bell, Lindsay	Blood Test (BMP)
Brooks, Caitlyn	Audiometry
Morris, Monica	X-Ray

4.2.6 Results

The results (Figure 4.5) suggest that showing pending orders reduces the time to correctly mark all late orders by more than a half, and that with the prioritization of orders, participants are 88% faster. A repeated measures one-way ANOVA (three treatment levels) with pairwise comparisons using the Holm adjustment method was run. Differences are statistically significant ($p < 0.001$, $F(2, 34) = 39.01$) and post-hoc paired t-tests establish differences between the interfaces:

Table 4.2: The list of orders given to participants for Interface 2.

10/22/2012

Doe, Jo	Blood Test (BMP)
Edwards, Rachel	Dietetics
Edwards, Rachel	X-Ray
James, Britney	Rapid Strep Test
Richardson, Allison	Rapid Strep Test

10/23/2012

Bell, Lindsay	Rapid Strep Test
Campbell, Jessica	Blood Test (CBC)
Edwards, Rachel	Blood Test (BMP)
Murphey, Sabrina	MRI
Parker, Amy	Dietetics
Parker, Amy	Rapid Strep Test
Richardson, Allison	Blood Test (CBC)
Ward, Andrea	Rapid Strep Test
Watson, Brooke	Blood Test (CBC)
White, Jane	Blood Test (BMP)
Wood, Courtney	X-Ray

10/24/2012

Bell, Lindsay	Dietetics
Cook, Lisa	Dietetics
Green, Bobbie	Blood Test (CBC)
Smith, Kathy	Blood Test (BMP)

Table 4.3: The list of orders given to participants for Interface 3.

10/22/2012

Cook, Lisa	X-Ray
James, Britney	Audiometry
Kelly, Cecelia	MRI
Richardson, Allison	Blood Test (BMP)
Richardson, Allison	Rapid Strep Test
Sanders, Chelsey	Blood Test (CBC)
Watson, Brooke	Audiometry

10/23/2012

Bailey, Isabel	Blood Test (BMP)
Bell, Lindsay	Audiometry
Bennett, Claire	Blood Test (CBC)
Black, Sue	Blood Test (BMP)
Campbell, Jessica	Blood Test (CBC)
Collins, Emily	Blood Test (BMP)
Cooper, Molly	MRI
Edwards, Rachel	MRI
Evans, Anna	Blood Test (CBC)
Rogers, Emma	Blood Test (BMP)

10/24/2012

Collins, Emily	Dietetics
Collins, Emily	Rapid Strep Test
Phillips, Sarah	Dietetics

Table 4.4: The reference sheet of time requirements for different types of orders used in the experiment.

	Expected Duration (Days)	Considered Lost After (Days)
Imaging and Lab Tests	1	4
Blood Tests [BMP or CBC] Rapid Strep Test MRI X-Ray Mammogram		
Consultations	3	16
Dietetics Audiometry		

results-only to results+pending ($p < 0.001$), results-only to prioritized-pending ($p < 0.001$), results+pending to prioritized-pending ($p < 0.001$).

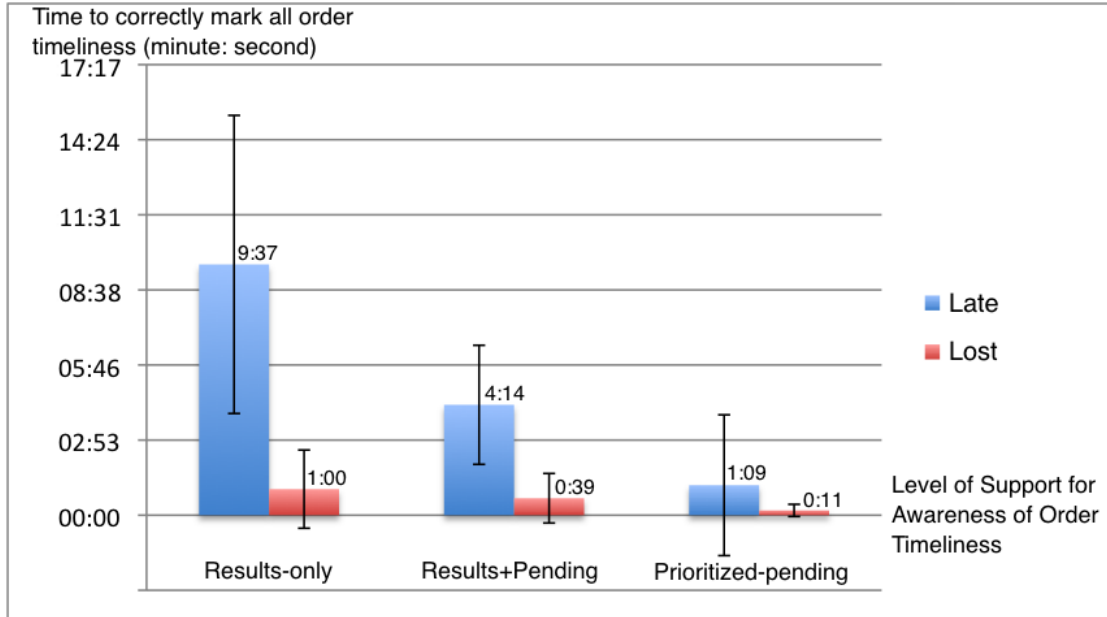


Figure 4.5: The chart shows the average time and standard deviation for 18 participants to correctly mark all late and lost orders using three levels of support for awareness of order timeliness (Results-only, Results+Pending, Prioritized-pending).

Another finding (Figure 4.5) is that showing pending orders reduces the time to correctly mark all lost orders by 35%, and that prioritizing orders has a greater effect (a couple of seconds). A repeated measures one-way ANOVA (with three treatment levels) shows statistically significant differences ($p = 0.038 < 0.05$, $F(2, 34) = 3.601$). With pairwise comparisons using no adjustments, the following differences between the interfaces are obtained: results-only to results+pending ($p = 0.307$), results-only to prioritized-pending ($p = 0.033 < 0.05$), results+pending to prioritized-pending ($p = 0.036 < 0.05$).

The number of mistakes that participants made in identifying all late orders

was reduced by 40% when pending orders were displayed (Figure 4.6). In addition, prioritizing by lateness information causes almost no corrections. A repeated measures one-way ANOVA indicates statistically significant differences between the three treatment levels ($p < 0.001$, $F(2, 34) = 14.79$). Post-hoc paired t-tests with the Holm adjustment method show the differences between the treatment levels: results-only to results+pending ($p = 0.022 < 0.05$), results-only to prioritized-pending ($p = 0.001 < 0.01$), results+pending to prioritized-pending ($p < 0.001$).

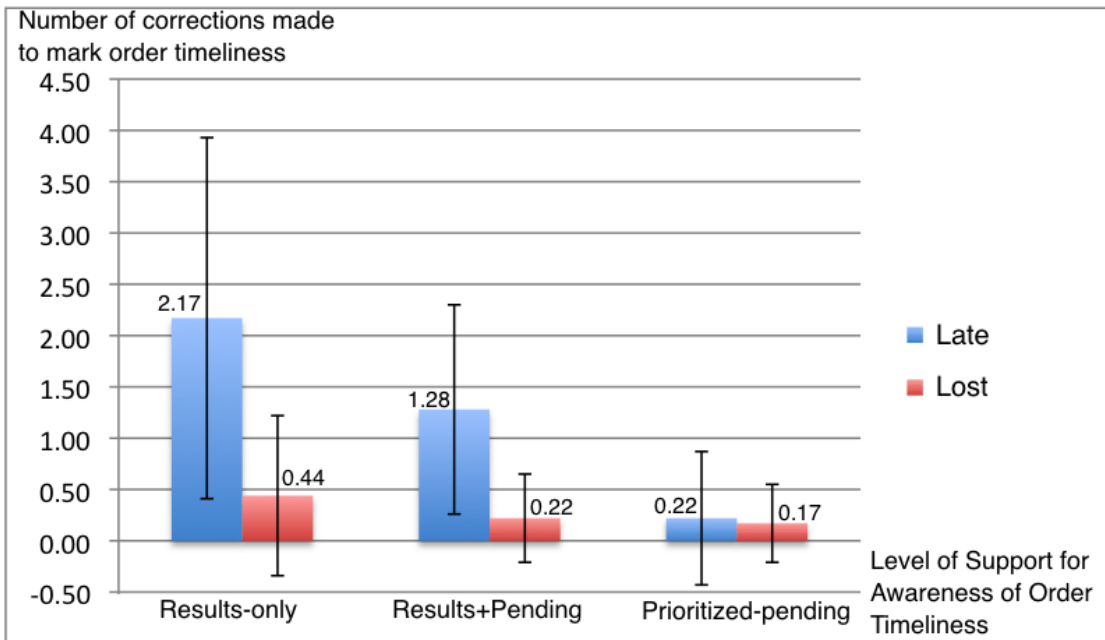


Figure 4.6: The chart shows the average number and standard deviation for 18 participants to correct their answers regarding order late and lost statuses using three levels of support for awareness of order timeliness (Results-only, Results+Pending, Prioritized-pending).

Finally, the experiment also found out that showing pending orders can decrease by half the number of errors made in marking all lost orders, and that prioritizing orders improves a little bit more (Figure 4.6). However, running a repeated measures one-way ANOVA does not show any statistical differences between the three

treatment levels ($p = 0.32$, $F(2, 34) = 1.178$).

4.2.7 Predictive Model for Results Management

After a user interface has been implemented, user tasks can be specified as a list of actions. This list can be used to predict the time required to perform tasks, by adding up the time it takes to complete all the steps. This predictive approach is called Goals, Operators, Methods, and Selection rules (GOMS) [20].

Based on the study results, a predictive model was built. The procedure for answering the questions with results-only level of support for order timeliness awareness is as follows:

```

pending = copy of orders
for each result in results do
  for each order in orders do
    if result.patientName == order.patientName then
      if result.testName == order.testName then
        cross out order from pending
        break
      end if
    end if
  end for
end for
late = empty
lost = empty
for each order in pending do
  if currentDate > normalDuration + order.date then

```

```

if currentDate < maxDuration + order.date then
    add order to late
else
    add order to lost
end if
end if
end for

```

Using this procedure, assume tn is the time to compare names, tl is the time to look up a date, and td is the time to compare dates. The study with 20 orders and 11 results shows that participants on average took 577 seconds and made approximately 2 mistakes for results-only (Figures 4.5 and 4.6). Thus, the below equation is obtained:

$$\begin{aligned}
 & [12 \times tn + 7 \times tn + 14 \times tn + 20 \times tn + 6 \times tn + 17 \times tn + 10 \times tn + 15 \times tn + tn + 4 \times tn + 2 \times tn \\
 & \quad + tn + tn + tn + tn + 2 \times tn + 2 \times tn + tn + tn + tn + tn + tn \\
 & \quad + 8 \times (4 \times tl + 2 \times td) + 2 \times tl + td] \times 3 = 577
 \end{aligned}$$

When I simplify the equation, I have the following:

$$3 \times (121 \times tn + 34 \times tl + 17 \times td) = 577$$

The procedure for answering questions with results+pending level of support for order timeliness awareness is (note that the first part of the algorithm from above is unnecessary):

```

late = empty
lost = empty
for each order in pending do
  if currentDate > normalDuration + order.date then
    if currentDate < maxDuration + order.date then
      add order to late
    else
      add order to lost
    end if
  end if
end for

```

Using the same notation from above and the study with 20 orders and 11 results show that participants on average took 254 seconds and made 1 mistake for results+pending (Figures 4.5 and 4.6). Thus, the following equation is obtained:

$$2 \times [2 \times (2 \times tl + td) + 7 \times (4 \times tl + 2 \times td)] = 254$$

The simplified equation is:

$$2 \times (32 \times tl + 16 \times td) = 254$$

There are three unknowns and two equations so the following estimates are obtained:

$$tn \approx 0.5$$

$$2 \times tl + td \approx 8$$

Professionals, who are familiar with the task, may skip some of the basic steps and do not make any mistakes. In this case, these equations turn into the following:

$$121 \times tn + 8 \times (2 \times td) + td = 120 \text{ (results-only)}$$

$$20 \times td = 60 \text{ (results+pending)}$$

For the parameters to hold, these ranges are necessary:

$$3 \leq td \leq 3.5$$

$$2 < tl \leq 2.5$$

Based on these, with x orders, y results, and z type of orders, it is possible to generalize the time it will take for an average user to answer the question. On average, a user has to check half the orders each time and assuming a patient is given half of the various types of orders, the cost for scanning the orders list is $tn \times y \times (x/2 + z/2)$. In addition, if $1/a$ of the pending orders are late or lost, the cost for determining late/lost status is $[\frac{x-y}{a} \times (4 \times tl + 2 \times td)] + [\frac{a-x+y}{a} \times (2 \times tl + td)]$.

Let's assume ts represents the cost to scroll and calculate the additional cost to the equations in the previous paragraph. According to the first algorithm above, the orders list is scanned $y/2+1$ times (to determine the pending and lastly to mark lateness) while the results list is scanned once (i.e. during marking lateness) from top to bottom. If only $1/b$ of the orders list fits on one page, the cost associated with scrolling the order list is $(y/2 + 1) \times b \times ts$. If the results list has only $1/c$ entries fit on one page, the cost will be $c \times ts$.

For a clinic manager to determine late/lost *results*, the total cost will be (if $1/d$

of the results are late or lost) $[\frac{y}{a} \times (4 \times tl + 2 \times td)] + [\frac{d-y}{a} \times (2 \times tl + td)]$.

Note that in the experiment, there were four durations to lookup (a 2×2 table). Time to look up a date, tl , is dependent on table dimensions. If the lookup table was larger (e.g. blood test and MRI had different expected durations), tl is going to grow proportional to the number of columns \times rows. In addition, if abnormal results have different expected durations than normal ones, the cost of lookup will increase even more because the table will quadruple in size. The cost to determine late/lost status will increase by $(x - y)$ because users are now required to check an additional column for information.

4.2.8 Discussion

It is interesting that the highest level of support for order timeliness awareness reduces the time to 1 minute and 9 seconds on average given that all the necessary information is readily available. The reason was three of the participants did not understand when the experimenter described the interfaces, did not ask questions about the interfaces, and did not pay attention to the features on the populated interfaces. When these outliers are removed from the final computations, the average drops to 12 seconds (standard deviation: 6 seconds).

To ensure that participants did not learn from prior trials, the experiment used different orders for the interfaces. To reduce the confounding effects, the number of orders were kept the same. In addition, the types of orders were similar for each interface. An additional factor for the experiment was the different distraction tasks². However, this was not a confounding factor because the participants did not even attempt to remember the details of orders and relied on the interfaces,

²Distraction tasks included reading an article, performing a drawing task, or playing an online game. This was an unintended and unexpected choice.

as will be discussed later in this section.

One of the participants was a physician and he was the winner of the bonus prize in the lowest level of support for order timeliness awareness. Because he was more familiar with the baseline interface than other participants, he performed the best but he still saved a significant amount of time and made fewer mistakes with the highest level of support for order timeliness awareness. He answered the first question in (i) 2 minutes and 26 seconds (1 correction) with the lowest level of support, (ii) 1 minute and 41 seconds (0 correction) with the middle level of support, and (iii) 13 seconds (0 correction) with the highest level of support.

There were no statistically significant differences found between the treatment levels for the number of errors made in marking all lost orders. Also, the post-hoc paired t-tests did not show statistically significant differences for the time to correctly mark all lost orders between results-only and results+pending. These were due to participants identifying lost orders while they were answering the first question. Most of the time, participants marked an order late although it was lost or vice versa. Because the experimenter pointed out their mistakes for the first question, they already had the correct answer to the question about lost orders.

There were some limitations to this study. It was difficult to simulate a real environment in an hour-long study. The distraction task separated the ordering from the reviewing step but did not include distractions that may happen during order or review time. The study results could have been more pronounced if there were more distractions but it would have overwhelmed the participants. Although a prize was offered to increase motivation, the participants did not develop any strategies to remember the orders and as the study progressed, they realized attempting to memorize orders was useless and they gave up completely. This also

confirmed what was observed in the clinics where physicians made extensive use of cheat sheets to remember their orders.

The implication of the experimental results is that showing pending and prioritizing can provide great help for busy medical workers. As improved systems reliably report delayed orders and effectively guide clinicians to act on them, they will spend less time sorting out the issues in the process. When physicians manage their results more promptly, they will also be able to provide better care for their patients.

The implications of the predictive model is that comparing names is easier than comparing dates, which is not surprising. If interfaces (like my prototype) reduce the number of date comparisons, users can perform their tasks more efficiently. Comparing dates is more costly than looking up a date. It is understandable that comparisons are more computationally intensive than lookups. However, this may be due to the low dimensions of the table and for a higher dimensional table, the time values can change.

These findings are also generalizable to tracking interfaces built for other processes where awareness of timeliness plays an important role. Examples include but are not limited to the software development cycle, paper-review process in academic journals, and business processes such as the return merchandise authorization process in the electronics industry and the internal supplier agreement process in IT companies. Finally, my predictive model serves as a guide to future designers and developers of tracking systems.

4.3 Evaluations of Actions for Rapid Completion

Since the previous experiment tested the non-interactive design principles, the next evaluation concerned the evaluation of interactive features called Actions for Rapid Completion (ARC) (see Figure 4.7). There were two studies that compared my research to other systems in terms of differences in: (i) time to execute representative actions, and (ii) the number of steps. The goal was to quantify the benefit of embedding actions within the list. To test the interaction techniques, five representative actions below were used. According to physicians, these were the most common actions for following up with patients whose results are normal and routine.

1. Review Later
2. Inform Patient + Confirm
3. Inform Patient + Repeat Test (1 month) + Confirm
4. Inform Patient + Schedule Visit (1 week) + Confirm
5. No Follow-up (Other) + Confirm

4.3.1 Evaluating Time to Execute Actions

The first evaluation was a comparison of time to execute the aforementioned representative actions. To be able to formally compare time, a second way to take follow-up actions were built into the prototype tool. This section explains the second version of the prototype, the protocol for the evaluation, and the study results.

- **Setup:** To control for other factors, everything was identical in the two versions of the prototype, except the interaction techniques. With the second

Patient	Test	Order Date	Review By	Abnormality
Evans, Anna	MRI / F	Mar 3, 2011 1:12 AM	Mar 10, 2011 1:12 AM	

No Follow-up

Ask Nurse to

Inform Patient (Evans, Anna)

Schedule Visit

Order

Repeat Test

Comments:

Review Later Complete

NAME: Evans, Anna

STUDY DATE: 03/09/2011 4:28:29 PM

COMPARISON: None available

HISTORY: Pain

MRI OF THE RIGHT ANKLE (WITHOUT IV CONTRAST)

FINDINGS:

The bones demonstrate no fracture or gross marrow abnormality. Articular surfaces are within normal limits.

The spring ligament complex appears to be intact.

The tibialis posterior, flexor digitorum, and flexor hallucis longus tendons are intact. The peroneus brevis and longus tendons are intact. The Achilles tendon is intact.

The tarsal sinus is unremarkable. There are no significant fluid collections to suggest ligamentous injury.

The remaining muscles and soft tissues are unremarkable.

Figure 4.7: Actions for Rapid Completion (ARC) list the most common actions for routine and normal results. When the user left clicks on the rows in any of the rich tables, ARCs appear. In the context of rich tables, ARCs are close to the mouse-click location and results are on the right. “Complete” finalizes follow-up so the result could be removed from “Results to Review” table. On the other hand, “Review Later” button means the result was viewed and/or some actions may have been taken but the follow-up is not complete yet. Such a result is still kept in “Results to Review” table for further processing but is marked as “viewed”.

action representation, the goal was to capture the best practice found in ongoing research [125] rather than the conventional approach. Therefore, some features were implemented from observations of work-in-progress that lead to ARC implementation. Figure 4.8 shows the state-of-the-art follow-up screen. First of all, it lists all possible follow-up actions together. Secondly, it appears at the same screen location and on top of the main screen that contains the orders list. Finally, it integrates actions with result review. Note that conventional follow-up is done through the patient records and separates actions and result review in separate tabs.

- **Protocol:** Each representative action was evaluated using MStart’s interactive features: (i) the state-of-the-art follow-up screen, (ii) ARCs, and (iii) keyboard shortcuts in ARC. Each representative action was performed three

Patient: Evans, Anna

NAME: Evans, Anna
 STUDY DATE: 03/09/2011 4:28:29 PM
 COMPARISON: None available
 HISTORY: Pain
 MRI OF THE RIGHT ANKLE (WITHOUT IV CONTRAST)
 FINDINGS:
 The bones demonstrate no fracture or gross marrow abnormality. Articular surfaces are within normal limits.
 The spring ligament complex appears to be intact.
 The tibialis posterior, flexor digitorum, and flexor hallucis longus tendons are intact. The peroneus brevis and longus tendons are intact. The Achilles tendon is intact.
 The tarsal sinus is unremarkable. There are no significant fluid collections to suggest ligamentous injury.
 The remaining muscles and soft tissues are unremarkable.

Ordered By: Brown, Joe

No Follow-up
 Ask Nurse to
 Inform Patient (Evans, Anna)
 Schedule Visit 1 months
 Order
 Repeat Test 1 months
 New Test Edit New Test
 Medication Edit Medication
 Consultation Edit Consultation
 See History
 Comments:
 Review Later Complete

Patient	Test	Order Date	Result Due	Status
Morris, Monica	TSH	Mar 2, 2011	Mar 8, 2011	Sample drawn and sent
Parker, Amy	MRI	Mar 2, 2011	Mar 8, 2011	Image(s) interpreted and reported
Peterson, Audrey	Blood (BMP)	Mar 3, 2011	Mar 9, 2011	Test ordered
Peterson, Audrey	Blood (CBC)	Mar 3, 2011	Mar 9, 2011	Test ordered
Reed, Megan	Mammogram	Mar 3, 2011	Mar 9, 2011	Image(s) interpreted and reported
Sanders, Chelsey	Pap Smear	Mar 3, 2011	Mar 9, 2011	Specimen collected and sent

Planned Tests (10)

Figure 4.8: Double-click on an order brings up the state-of-the-art follow-up screen. It appears on top of MStart's current screen at the same screen location and disallows interaction with MStart until closed. This screen lists all possible follow-up actions together on the right side while the result report is shown on the left. "Complete" finalizes follow-up so the result could be removed from "Results to Review" table. On the other hand, "Review Later" button means the result was viewed and/or some actions may have been taken but the follow-up is not complete yet. Such a result is still kept in "Results to Review" table for further processing but is marked as "viewed".

times for ten orders. The first trial was discarded and the second and third trials were averaged to come up with a time value. The timing started with a new orders list and ended when ten orders were completed correctly, i.e. using the chosen representative action. The entire study was completed by the same person on the same computer.

- **Results:** The results in Table 4.5 show that ARCs speed up the follow-up process for normal routine results by 25% compared to the state-of-the-art follow-up approach in the case that necessitates informing the patient, a new

visit, and confirmation. The keyboard shortcuts even further (up to 40%) reduce the time to follow-up.

Table 4.5: Results of evaluating representative actions using three different interactive features in MStart. Times are shown in minute:second format while the percentages represent the speed-up compared to the state-of-the-art follow-up.

Representative Action	State-of-the-art Follow-up	ARC	ARC (keyboard)
Review Later	0:30	0:28 (7%)	0:26 (13%)
Inform Patient + Confirm	0:44	0:39 (11%)	0:29 (34%)
Inform Patient + Repeat Test (1 month) + Confirm	0:51	0:42 (18%)	0:34 (33%)
Inform Patient + Schedule Visit (1 week) + Confirm	1:08	0:51 (25%)	0:41 (40%)
No Follow-up (Other) + Confirm	1:02	0:53 (15%)	0:49 (21%)

4.3.2 Evaluating the Number of Steps

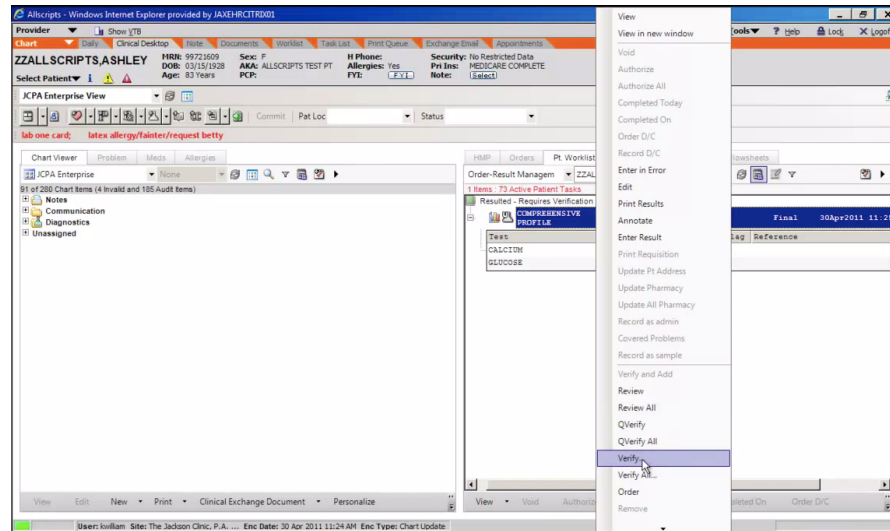
While the previous study compared time, this evaluation consisted of comparing the number of steps needed to execute the aforementioned representative actions. To make the evaluation closely match reality, dominant electronic health record (EHR) systems that have standard functions for follow-up were chosen. While there are many EHR systems (e.g. Epic), Allscripts was chosen because the screenshots were publicly available on Youtube [140] and OpenVista is an open source system. EHR vendors typically do not make screenshots or videos available and do not publish guidelines. In addition, both of these systems still have many clinicians and clinics using them. This section will briefly describe OpenVista and Allscripts. Then, it will explain the necessary steps in one of these EHR systems to accomplish the representative tasks above.

- **OpenVista:** OpenVista Clinical Information System [85] is a cross platform client application for interfacing with the Veteran Affairs' VistA EHR system and systems derived from VistA, like OpenVista Server. It is based on the design of the Veteran Affairs' Computerized Patient Record System and includes image viewing and other commercial enhancements. Many clinics build their own custom implementations on top of OpenVista so it is representative of many EHR systems. For this study, OpenVista Clinical Information System version 0.9.96 was selected and ran using Public Demo Server (OpenVista Server version 1.5 SP2).
- **Allscripts:** Allscripts Enterprise EHR [2] is a commercial clinical and business software for large ambulatory practices (with 25 or more physicians, or for multi-specialty practices).
- **Setup:** Because OpenVista and Allscripts did not support taking actions from the same screen, the steps in each representative action above are broken down to unit actions. Furthermore, "Review Later" representative action was unavailable in both EHR systems, instead the results were removed after they have been opened. To be able to compare the steps, each unit action is supposed to start and end with the results list. Here are the unit actions in ARC:

1. *Inform Patient:* ARCs require three steps to inform patients of their results:
 - (a) Left-click on the result to bring up ARC
 - (b) Check "Inform Patient"
 - (c) Click "Complete"
2. *Confirm:* ARCs require two steps to confirm results:

- (a) Left-click on the result to bring up ARC
 - (b) Click “Complete”
3. *No Follow-up*: ARCs require four steps to confirm results that need no follow-up:
- (a) Left-click on the result to bring up ARC
 - (b) Check “No Follow-up”
 - (c) Select a reason from drop-down list
 - (d) Click “Complete”
4. *Repeat Test*: ARCs require three to five steps to confirm results that require a repeat order for follow-up:
- (a) Left-click on the result to bring up ARC
 - (b) Check “Repeat Test”
 - (c) *Optional*. Choose a number in the numeric stepper
 - (d) *Optional*. Select unit from the radio buttons
 - (e) Click “Complete”
5. *Schedule Visit*: ARCs require three to five steps to confirm results that require a new visit for follow-up:
- (a) Left-click on the result to bring up ARC
 - (b) Check “Schedule Visit”
 - (c) *Optional*. Choose a number in the numeric stepper
 - (d) *Optional*. Select unit from the radio buttons
 - (e) Click “Complete”
- **Procedure:** Below are the necessary steps to accomplish the unit actions in either Allscripts or OpenVista. Only one system was used to evaluate because not both of them support each action.

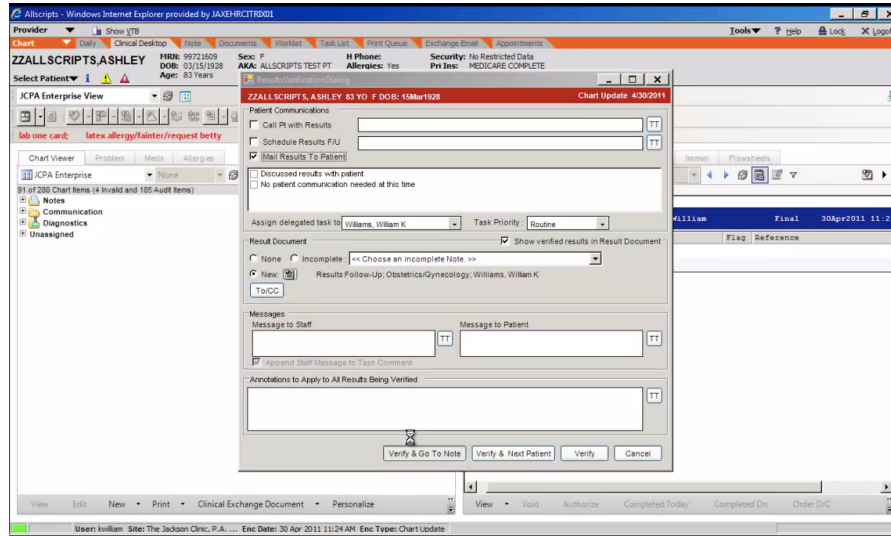
1. *Inform Patient*: Allscripts requires six steps to inform patients:
 - (a) Double-click on the result to open the report in the patient record
 - (b) Right-click on the result (Figure 4.9a)
 - (c) In the list of options, scroll to and select “Verify”
 - (d) Check “Mail Results to Patient” (Figure 4.9b)
 - (e) Click “Verify” button
 - (f) Exit out of the patient record by clicking “Daily” tab (Figure 4.9c)



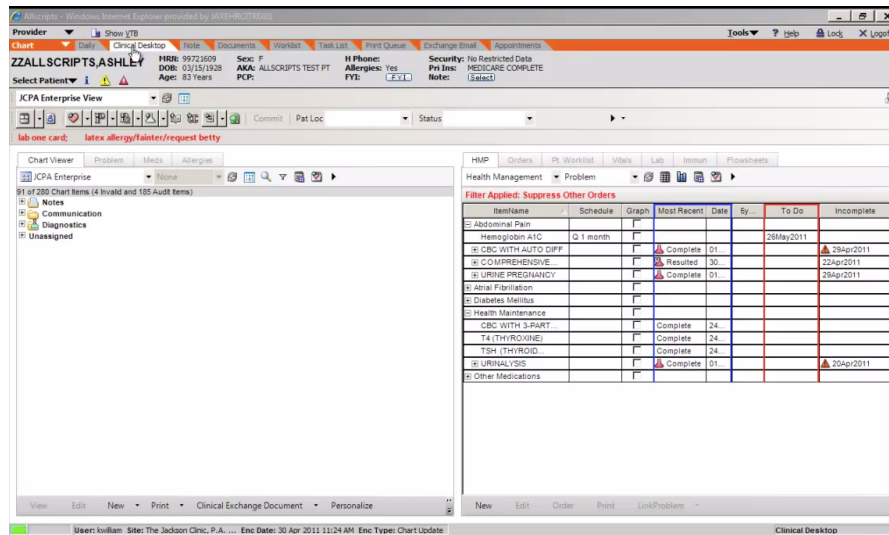
(a) Right-click on a result to bring up the context menu

Figure 4.9: Steps (a)-(c) correspond to the different screens and interactions to inform a patient.

2. *Confirm*: Allscripts requires five steps to confirm results:
 - (a) Double-click on the result to open the report in the patient record
 - (b) Right-click on the result (Figure 4.9a)
 - (c) In the list of options, select “Verify”
 - (d) Click “Verify” button (Figure 4.9b)
 - (e) Exit out of the patient record by clicking “Daily” tab (Figure 4.9c)



(b) Results Verification Dialog appears after clicking Verify.



(c) Switch from Clinical Desktop to Daily tab

Figure 4.9: Steps (a)-(c) correspond to the different screens and interactions to inform a patient.

3. *No Follow-up*: Allscripts requires four steps to confirm results that need no follow-up:
 - (a) Double-click on the result to open the report in the patient record
 - (b) Right-click on the result (Figure 4.9a)
 - (c) In the list of options, select QVerify (Figure 4.10)
 - (d) Exit out of the patient record by clicking “Daily” tab (Figure 4.9c)

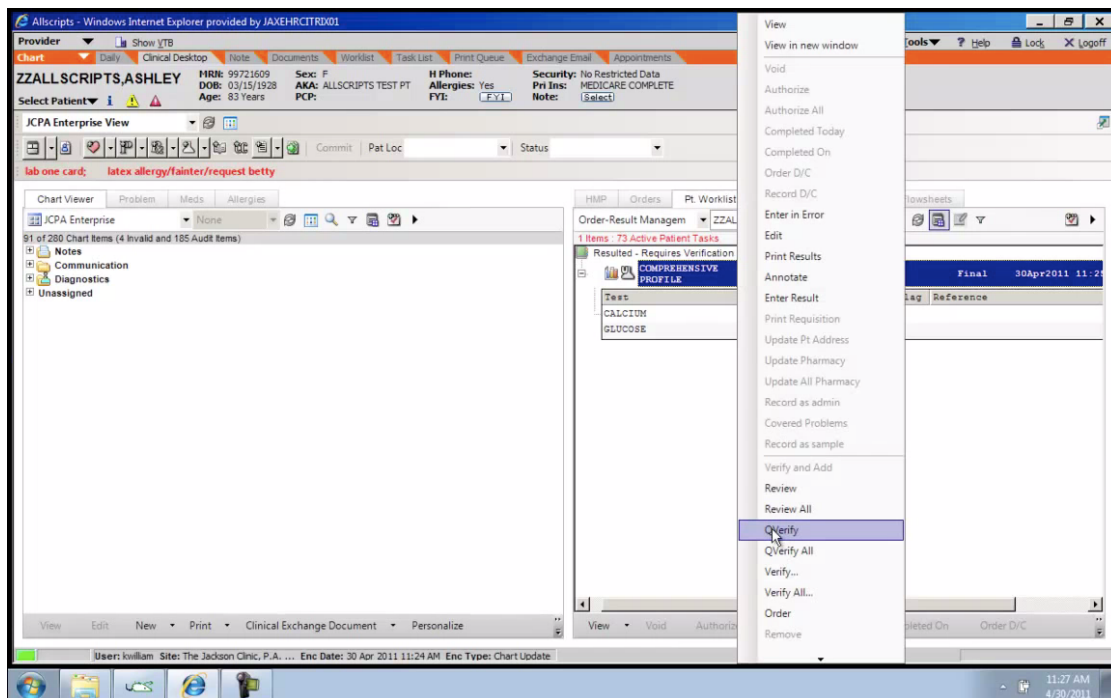
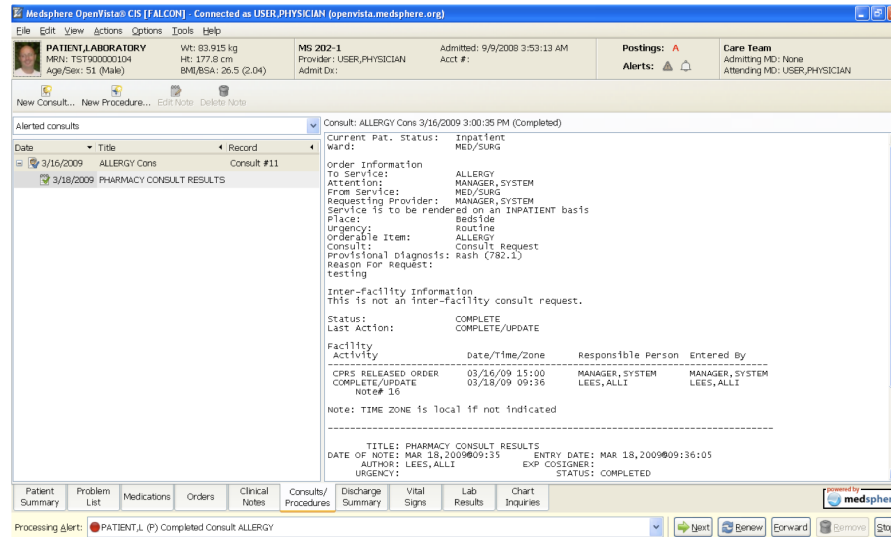


Figure 4.10: To follow-up without a follow-up, a result’s right-click menu has QVerify, which stands for quick verify.

4. *Repeat Test*: With OpenVista, it takes eleven steps and seven sub-steps to confirm results that require a repeat order for follow-up.
 - (a) Double-click on the result to open the report in the patient record (Figure 4.11a)
 - (b) Select the “Order” tab (Figure 4.11b)

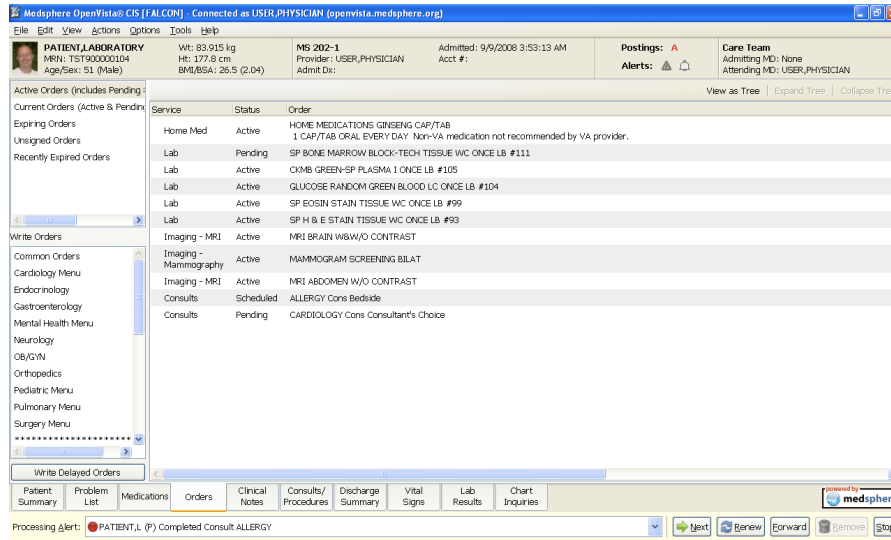
- (c) Click “Write Order” from this screen (Figure 4.11c)
 - (d) In the list of “Common Orders”, scroll and locate the current test type such as lab, imaging, and consultation (Figure 4.11d)
 - (e) Click test
 - (f) In the “Order a test” dialog box (Figure 4.11e),
 - i. From the drop-down list, choose a specific order
 - ii. From Date/Time drop-down list, choose “Future”
 - iii. In the popup menu, go to the month from within the calendar
 - iv. Select the date
 - v. Click “Select” button
 - vi. Fill out “Reason for Request” text area
 - vii. Click “Order” button
 - (g) Close the dialog box
 - (h) In a popup menu, approve by clicking “Stop Ordering” button
 - (i) Close the list of common orders
 - (j) Click “File” menu (Figure 4.12)
 - (k) Choose “Close Patient”
5. *Schedule Visit*: With OpenVista, it takes twelve steps to confirm results that require a new visit for follow-up:
- (a) Double-click on the result to open the report in the patient record
 - (b) Click on the physician name in the toolbar to bring up the “Select Encounter” dialog box (Figure 4.13a)
 - (c) Choose “New Visit” tab (Figure 4.13b)
 - (d) From the scroll-down list, choose “Visit Location” (Figure 4.13c)
 - (e) Click “Visit Date” (Figure 4.13d)



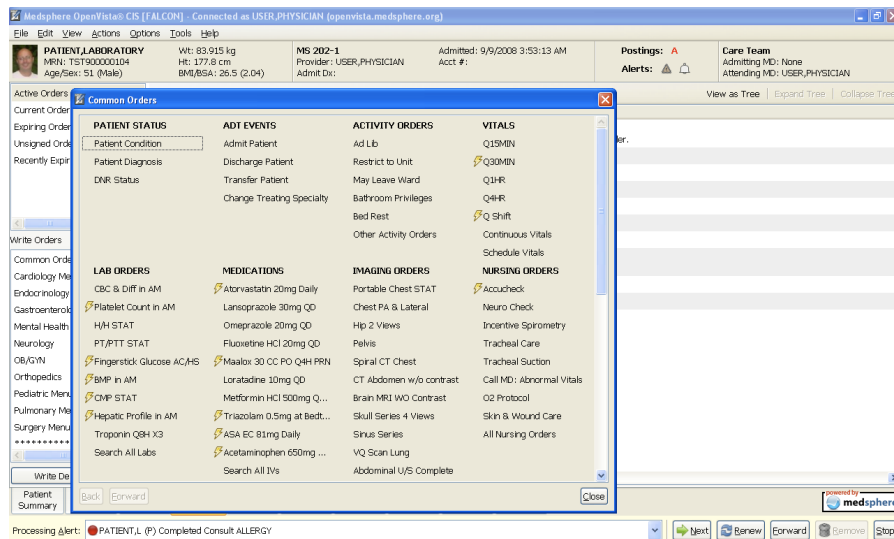
(a) A double click on the list of results opens the tab that has the consultation report in the patient record.

Figure 4.11: Steps (a)-(e) correspond to the different screens and interactions to order a repeat consultation.

- (f) In the popup menu, go to the month from within the calendar (Figure 4.13e)
 - (g) Select the date
 - (h) Select the “Hour”, “Minute”, and “AM/PM” from numeric steppers
 - (i) Click “OK”
 - (j) Click “Select”
 - (k) Click “File” menu (Figure 4.12)
 - (l) Choose “Close Patient”
- **Results:** By comparing the number of steps for ARCs to an EHR system (Table 4.6), the number of steps to execute a simple follow-up action such as ‘confirm’ or ‘inform patient’ can be reduced by half. On the other hand, confirming results that need no follow-up in ARCs did not save any steps. A bigger impact can be accomplished for scheduling a new visit follow-up

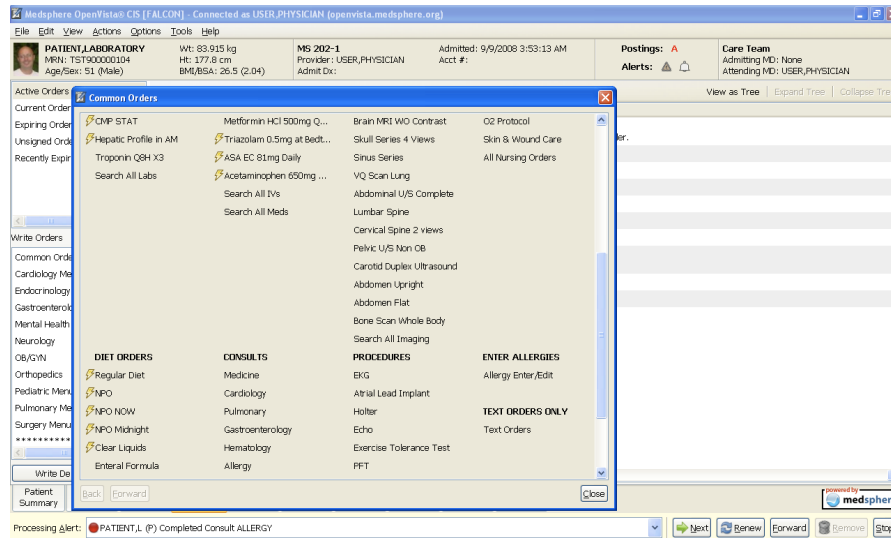


(b) To place an order, users switch to Orders tab.

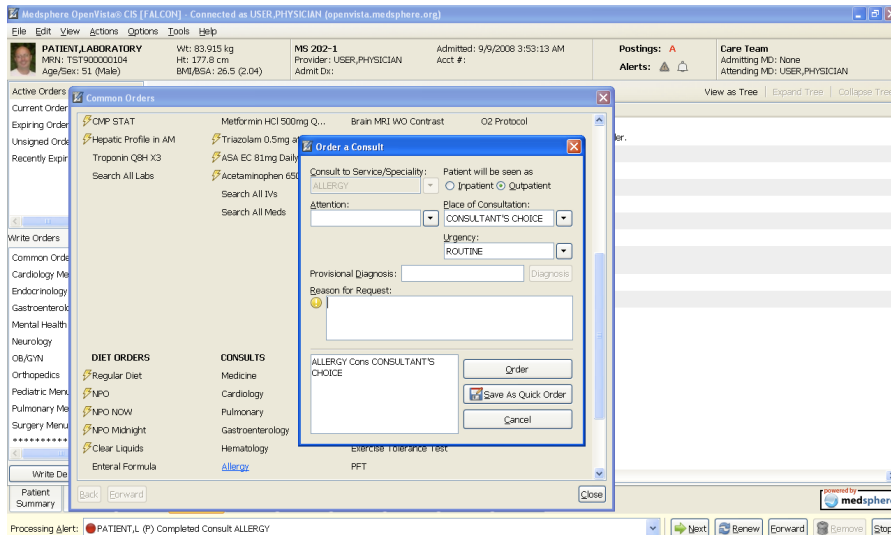


(c) To write an order, the list of common orders are brought up.

Figure 4.11: Steps (a)-(e) correspond to the different screens and interactions to order a repeat consultation.



(d) The desired order type is found in the list by scrolling.



(e) Clicking on the order pops up a new dialog to actually write the details of the order and place it.

Figure 4.11: Steps (a)-(e) correspond to the different screens and interactions to order a repeat consultation.

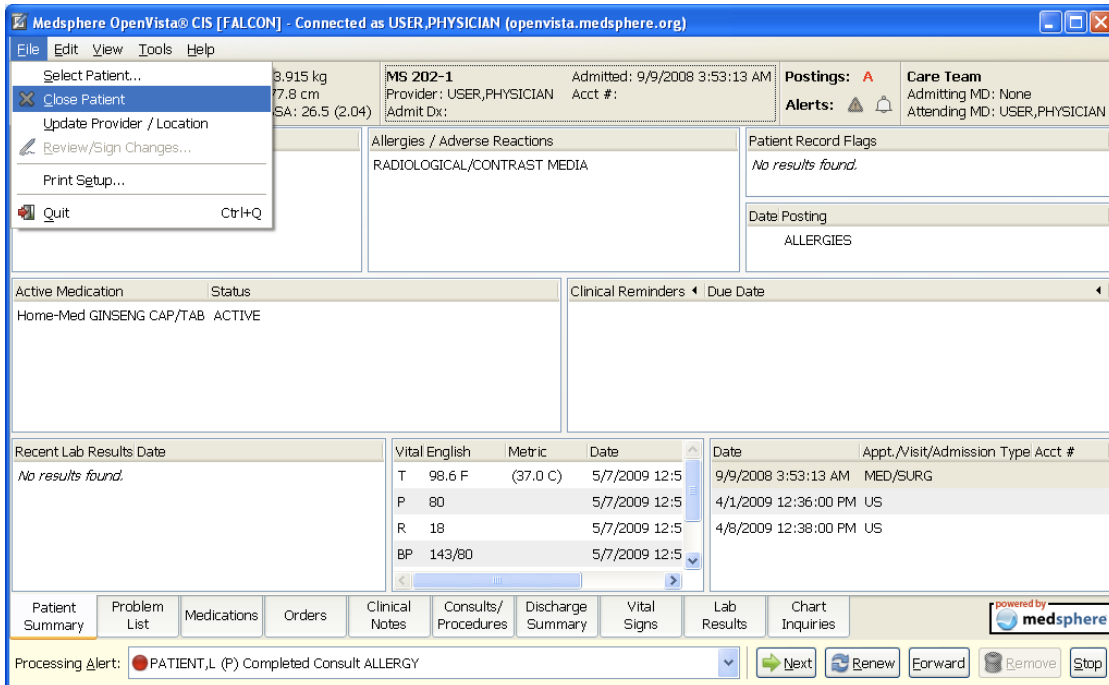
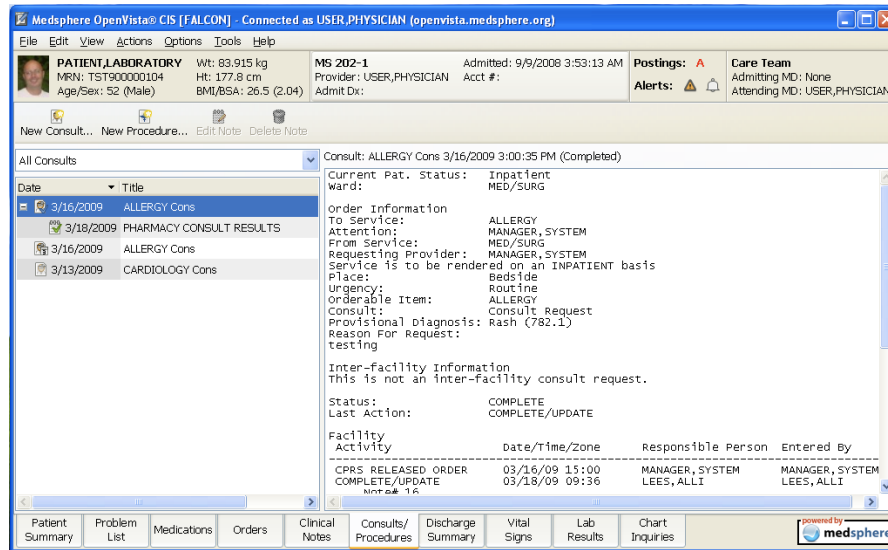


Figure 4.12: Closing the patient record is via the File menu.

action. An EHR may require twelve steps as opposed to the five steps using ARCs. The most dramatic effect was seen for repeating a test: seventeen steps in current EHR systems versus five with ARCs.

Table 4.6: Results of evaluating representative actions using two EHR systems and ARC. The percentages represent the reduction in ARC steps compared to the EHR systems.

Representative Action	OpenVista	Allscripts	ARC
Inform Patient	-	6	3 (50%)
Confirm	-	5	2 (60%)
No Follow-up (Other)	-	4	4 (0%)
Repeat Test (1 month) + Confirm	17	-	5 (71%)
Schedule Visit (1 week) + Confirm	12	-	5 (58%)



(a) Clicking on the physician name in the toolbar initiates the process for scheduling a visit.

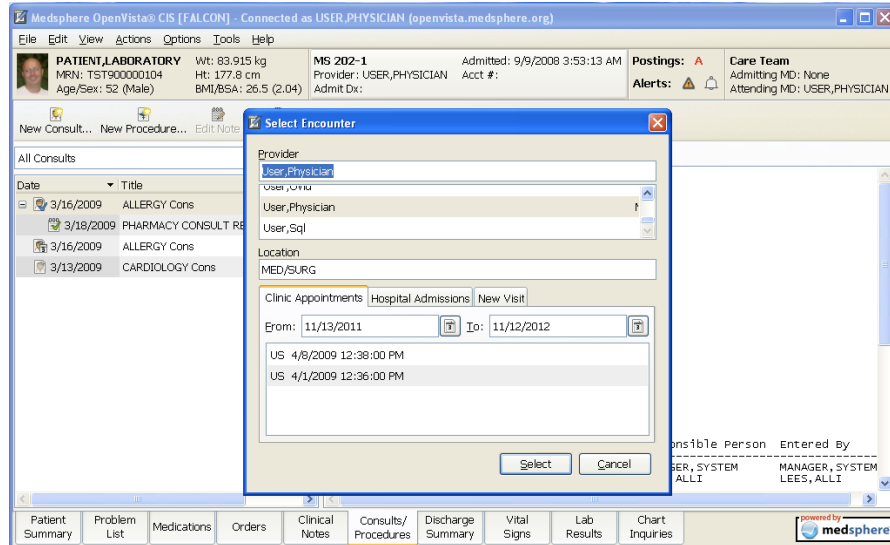
Figure 4.13: Steps (a)-(e) correspond to the different screens and interactions to schedule a new visit.

4.3.3 Discussion

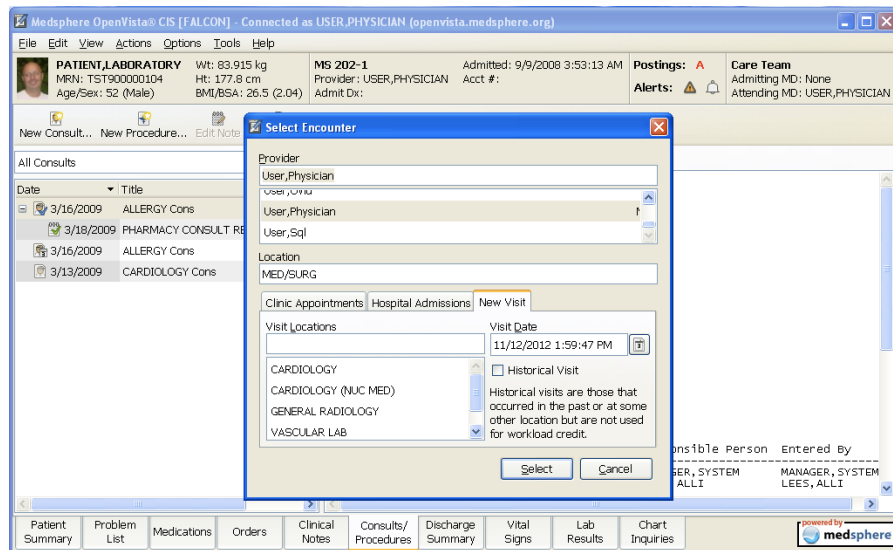
The results of this evaluation imply that with the help of ARCs and its keyboard shortcuts, clinicians can dismiss the simple cases faster and with fewer steps from their orders list. This will give clinicians more time to focus on more complex cases that require analyzing in detail abnormal results, viewing patient histories, and comparing past results.

This evaluation focused on the most common follow-up actions. However, another way to think about ARCs is they provide shortcuts into patient records. In other words, they direct clinicians to the relevant part of patient records without having them switch context or navigate in patient records. Hence, ARCs can be generalizable to all possible actions in a patient record.

Some observations were made during the first evaluation. As the study pro-

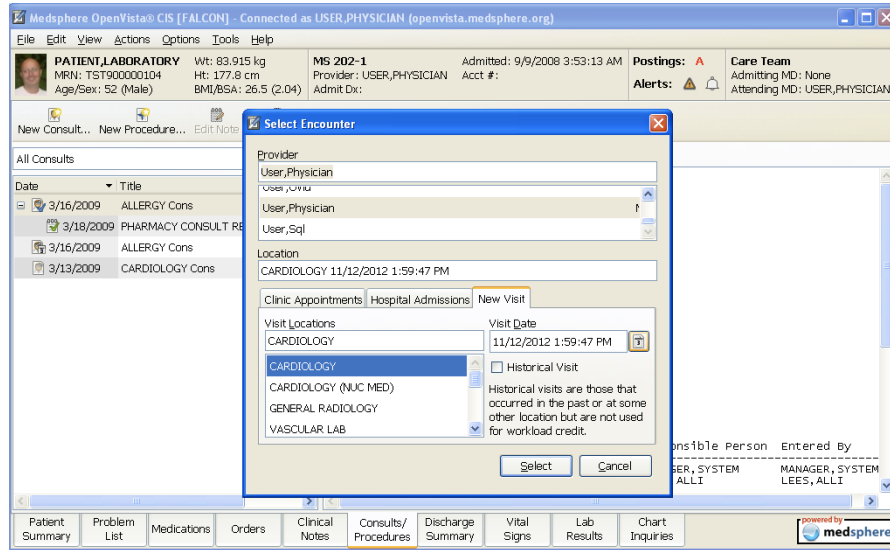


(b) Fields in the 'Select Encounter' popup have to be filled out.

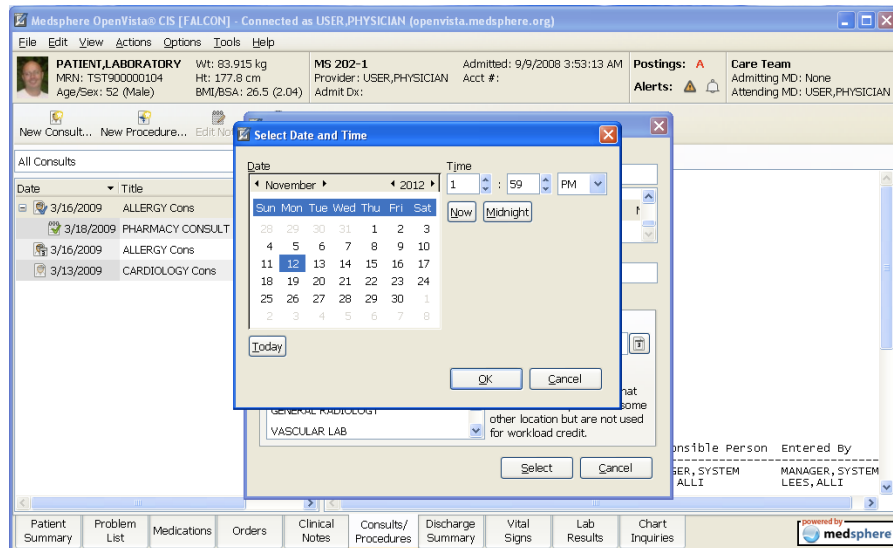


(c) Selecting 'New Visit' tab instructs for information regarding the visit.

Figure 4.13: Steps (a)-(e) correspond to the different screens and interactions to schedule a new visit.



(d) From the list, users choose a visit location, i.e. where patients should go.



(e) A visit date and time must be selected to complete the scheduling.

Figure 4.13: Steps (a)-(e) correspond to the different screens and interactions to schedule a new visit.

gressed, the mouse pointer was moved to the same position to save some time before the state-of-the-art follow-up screen appeared. Unlike ARCs that appear close to the click position, the state-of-the-art follow-up screen always appeared at a fixed location on the screen. Although the mouse travels less with ARCs, this fact does not always save time.

One limitation of this first evaluation was the integration of result review with follow-up actions on the state-of-the-art follow-up screen. As can be seen in the second evaluation, it is common practice to show the follow-up actions and result reviews in different tabs. However, to simplify comparisons regarding two action representations, the state-of-the-art follow-up screens were built to show an example of incorporating result review with follow-up actions for both normal and complex results. However, ARC is still faster than the state-of-the-art screen because ARCs only list the most common options for the normal routine cases. Have result review been separated from follow-up actions, the results of the first study will have been more prominent.

There were some important findings in current EHR systems. There was no way to postpone follow-up at a later time (review later). This can be dangerous because clinicians might accidentally bring up a result and quit out without remembering and the result would be gone. Also, some naming conventions were non-obvious. For example, it is difficult to comprehend the difference between Verify and QVerify by looking at only the labels. Instead QVerify should have been spelled out as Quick Verify, as in the case of ARCs. Finally, mouse travel distances were far because there were too many redundant items, related objects were not placed close together, and opening and closing too many popups were contributing factors.

One drawback was that the results of the second evaluation did not show reduc-

tions in steps for “No Follow-up” action. On the other hand, mouse has to travel a longer distance in current EHR systems than ARCs so the time requirements may still be in favor of ARCs.

One limitation to these two evaluations was that none were run as a controlled experiment, where study variables were tested with participants to show statistical significant results. For the first evaluation, Fitts’ Law [46] already shows that as the target gets farther, the time increases and ARCs are pretty compact compared to their state-of-the-art equivalents. For the second evaluation, the number of steps were compared. Because EHR systems require more steps than ARCs, participants would have more likely made more mistakes and took more steps in total to accomplish the actions.

4.4 Summary

This chapter presented several studies that evaluated various aspects of my systems and guidelines. Table 4.7 summarizes all the evaluations, results, and insights. The results of these evaluations should inspire vendors and medical researchers in designing results management systems.

While these were the important features chosen for evaluations, some were excluded. For example, clarifying responsibility was not formally tested because a formal evaluation would require integration in an operational EHR system. Although prevalent systems support phonebook-style contact lists, they do not employ process management techniques and the person in charge of handling the current step of the process is missing. To compare my prototype system to one without a workflow engine, many simplifications will be necessary and the results of such an evaluation may not be useful or generalizable. Moreover, many guidelines

Table 4.7: Several studies evaluated various aspects of my systems and guidelines.

Evaluation	People	Outcomes	Insights
Iterative Design Reviews with Medical Professionals	~ 40	Experts approved my process model and retrospective analysis. Table and ARC suggestions lead to refinements.	Creation of a system flexible enough to accommodate different needs of clinicians.
Controlled Experiment for Awareness of Order Timeliness	18	Differences in time to answer the questions and the number of corrections given the support for order timeliness awareness are statistically significant ($p < 0.001$, $F(2, 34) = 39.01$).	Showing pending and prioritizing can reduce the time clinicians spend sorting out the issues and increase their chances in providing better care for patients.
Predictive Model for Results Management	N/A	Time to compare names is 0.5s, to look up a date is 2-2.5s, and to compare dates is 3-3.5s.	If interfaces reduce the number of date comparisons, users can be more efficient.
Comparisons of Actions for Rapid Completion (ARC) to popular EHR systems	N/A	ARCs speed the follow-up for normal routine results by 25%. The number of steps for follow-up can be reduced by half with ARCs.	Clinicians can dismiss the simple medical cases faster and with fewer steps, thus having more time to focus on more complex cases.

regarding the table design were not formally tested because most of them have already been proven by various researchers in the field. The results of evaluations for those that have not yet been evaluated are obvious. Although these features were left untested, I believe that they still have validity. I urge future researchers to consider ways to evaluate them in their systems.

Chapter 5

Design Guidelines

“The central problem for human-computer interaction researchers is developing adequate theories and models. Traditional psychological theories must be extended and refined to accommodate the complex human learning, memory, and problem solving required in user interfaces. Useful goals include descriptive taxonomies, explanatory theories, and predictive models. When predictions can be made for learning times, performance speeds, error rates, subjective satisfaction, or human retention over time, designers can more easily choose among competing designs.” – Shneiderman (2010)

This chapter presents a set of Design Guidelines intended for Electronic Health Record (EHR) developers, which may also be valuable to others who deal with presentation of tables of information. Tables are the most common way of managing users’ daily work. Thus, widely accepted principles for tabular displays can improve the design and functionality. A series of design guidelines, integrated with principles for table operations, can be applied to achieve an aesthetically-minimalistic and functionally-effective table approach. These guidelines are devised to foster users’ attention to various states and important items so as to rapidly dismiss items that are deemed non-critical. With justifications and examples, this chapter serves as a reminder to common pitfalls in designing such systems. In addition, the chapter guides the reader to decide which guidelines are relevant, or should be

refined, or which new ones should be added.

Evidence ratings indicate the strength of the evidence for each guideline in the scale of Low, Medium, and High:

- **Low:** No user research findings.
- **Medium:** User research findings are limited and may be conflicting, i.e. there is a mixed agreement between domain experts.
- **High:** User research findings are clear and with a significant number of participants, no known conflicting research-based findings, domain expert opinions agree with the research.

Conformance ratings denote the importance of applying the guideline from Recommended to Mandatory. These are based on Microsoft Common User Interface Guidance [86]. To determine the conformance levels of each guideline, evidence ratings were taken into account at first and finally, the personal opinion of the author – as a designer and implementer of these guidelines – was considered.

- **Recommended:** An implementer is encouraged to follow the guideline.
- **Mandatory:** An implementer must follow the guideline.

These guidelines are imperfect and require a social process to apply them [91, 121]:

- **Education:** Designers and implementers should be motivated with a live or video presentation followed by a discussion of how to apply these guidelines.
- **Enforcement:** An expert manager has to review the interface with a clear process to verify that the guidelines have been applied.
- **Exemption:** To support creative work that was not covered by these guidelines', managers should balance the enforcement process with a simple and

rapid exemption process.

- **Enhancement:** Organizations should produce an annual revision that improves these guidelines and extends them to cover novel topics.

These guidelines build on three groups of published guidelines. The first group of guidelines is for presenting statistical data that is dominated by static numeric values [44, 66, 128, 141]. These tables are used in business presentations. The second group of guidelines is for web-based interactive table widgets [13, 30, 51, 98, 131, 139, 142], which are not well-established and are dependent on the developer toolkit settings. These documents list different web tools that come with one or two of the principles implemented. However, none are steered toward the medical domain so they do not answer the essential question: which guidelines are for users working under busy and time-critical environments? The third group of guidelines describes tables filled with clinical values [34, 86] but they are the same as the first group of guidelines, with clinician preference ratings. They are limited to formatting and tables with numeric values while table functionality is left out.

The design guidelines in this document are split into three categories. Guidelines for medical results management, low-level table design guidelines that apply general HCI principles, and actions for rapid completion guidelines that focus on the interactive features of tables. The contributions are twofold: (i) new guidelines for results management and actions for rapid completion, (ii) extended table guidelines with examples and ratings. The guidelines for results management and actions for rapid completion presented in this chapter are the new contributions of this dissertation work. The table design guidelines in Appendix D are refinements of existing guidelines that emerged from author's own experience of designing and building tables. These guidelines are a substantial advance over existing table de-

sign guidelines with refinements of 25 guidelines and 28 new guidelines. I believe these guidelines will find widespread use in medical informatics and data presentation in many application domains.

The evidence ratings for results management and actions for rapid completion are derived from this thesis work, while the evidence ratings for the table design guidelines come from existing research. Some of the existing table guidelines have never been tested, but they are well accepted because they have sound theoretical foundations or practitioners have substantial experience of trying out variations. Appendix D contributes refinements to table design guidelines, some of which have general applicability, while others are tuned to the medical domain. Therefore, my areas of high contribution are likely to have low evidence in the existing literature. Guidelines which are moderately original (either part of the guideline can be found elsewhere and the other part is new or the guideline has been adapted to the case) have medium evidence. The established guidelines, which did not require refinement, have high levels of evidence in the literature.

5.1 Results Management Design Guidelines

Results management involves the process that starts when an order for a patient is placed by a primary care physician at a clinic and ends when all the follow-up actions for the patient are taken and the physician signs off on the result. While order entry is not part of results management, some follow-up actions may include writing new orders.

These guidelines [133] come from the perceived shortcomings of existing EHR interfaces. In most systems, physicians see a table of results that came back (either for all patients, or per patient), which serves as a reminder to review results.

Pending orders are not visible (unless physicians read the details of the individual patient records or use clumsy reporting tools) so physicians are forced to remember orders they have placed. Systems have no notion of expected latency between order and results. Results are sorted by arrival date, with the newer ones at the bottom necessitating users to scroll. If an expected result is not there, there is no way to know what could be wrong. The only resort is to get on the phone and track the order down. Once results have been seen, there is no mechanism to ensure the follow-up is complete. The results management design guidelines are as follows:

5.1.1 Show pending results (Evidence: High; Conformance: Mandatory)

Whether looking at results of all patients or only one patient, the table should provide access to arrived results, pending orders, and possibly planned orders.

In Figure 5.1, the bottom half of the window shows results that came back. Here, there is little perceptual help for all orders because showing the results only does not remind users about pending orders. Cognitive load¹ is high since the busy and distracted clinicians may forget the details of the orders they have placed. In terms of motor performance, clinicians have to go to patient records to find out the pending orders, which is multiple clicks away (e.g. select the patient, go to the “Orders” tab).

Figure 5.2 shows a better example for all the orders of Dr. Brown. The tables show arrived, pending, and planned orders. One way to implement this is to show all incomplete orders (either resulted or pending) from patient records in an inbox view. A controlled experiment showed the effectiveness of this approach (see

¹Cognitive load refers to the control of working memory, i.e. the system that actively holds multiple pieces of information in the mind to be manipulated.

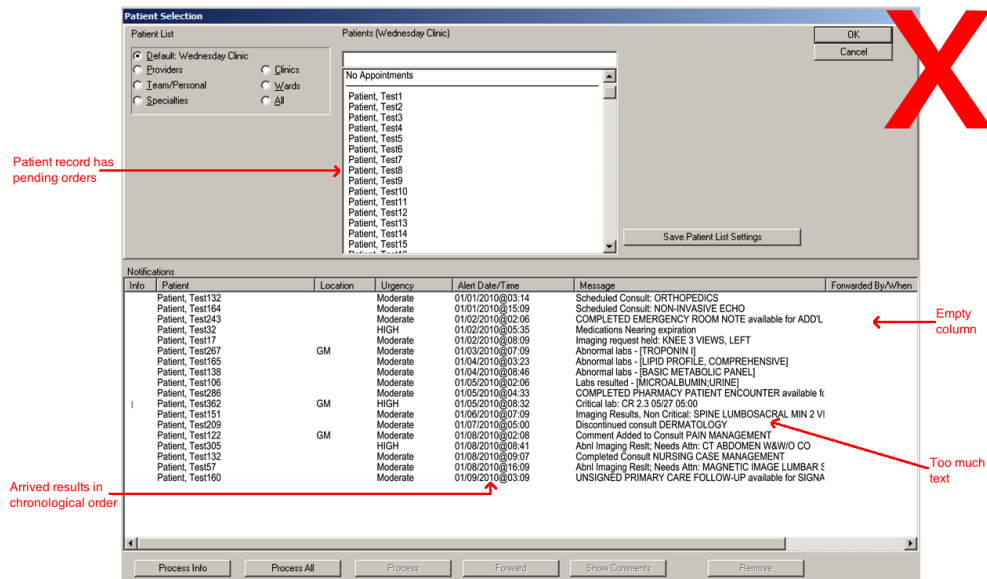


Figure 5.1: Veterans Administration Electronic Health Record View Alert window lists only returned results.

Chapter 4).

5.1.2 Prioritize by late and lost status (Evidence: High; Conformance: Mandatory)

The bottom half of Figure 5.1 has no notion of lateness. Therefore, there is no perceptual help because there is no way to know if an order has been delayed. Cognitive load is high since clinicians need to calculate how much time has passed since the order has been placed to figure out if the expected duration is exceeded. In terms of motor performance, clinicians may bring up a date calculator and enter the information or do the computations in their head (results may be incorrect).

A good example is in Figure 5.2 that sorts all tables by default so as to visually aid users see late and lost results at the top. This example employs an underlying process model, which requires system administrators to specify durations to cal-

The screenshot shows the MSTART application interface. At the top, there's a title bar with 'MSTART' and a menu bar with 'File'. Below that, a status bar indicates 'Physician Joe Brown reviews results from his inbox several times a day' and 'Current Time: Thursday, March 10, 2011 6:33 PM'. The main content area is titled 'Riverside Clinic' and 'Provider: Brown, Joe'. It features a search bar and filter buttons: 'All', 'New', 'Late', and 'Not Completed'. The 'Late' filter is highlighted in red. Below the filters is a table of 'Results to Review (17)'. The table has columns: Patient, Test, Order Date, Review By, and Abnormality. The rows are color-coded: yellow for new results, orange for late results, and red for not completed results. Below this table is a section for 'Pending Test Results (6)' with a similar filter set. A red box highlights the 'Late' filter in this section, and a red arrow points to the 'Pending Test Results' section. At the bottom, there is a section for 'Planned Tests (10)' which is collapsed.

Figure 5.2: Rich tables adhere to the design guidelines. Results that have returned are listed at the top in “Results to Review”, while orders that have not returned to the physician are shown under “Pending Test Results”. Orders that have been placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red.

calculate deadlines at the time of physician order entry and physicians may override this date (see Appendix B). A controlled experiment showed the effectiveness of this approach (see Chapter 4).

For systems, which do not have support for workflow management, it seems difficult to integrate such functionality from scratch. An easier implementation might ask the clinician at order time when they expect the results and when orders should be considered lost [82]. When clinicians enter this information, orders can be prioritized by late and lost status based clinician-provided dates.

5.1.3 Indicate physician acknowledgment and timeliness (Evidence: Medium; Conformance: Mandatory)

When physicians open the results, some systems remove those results from the table and others mark such results reviewed. When results are removed or marked as reviewed, it affects perceptual and motor performance. Clinicians need to figure out the results they have just opened, which is harder if results were opened accidentally. Clinicians may go to other windows to bring results back to their inbox or press extra buttons to change result status from reviewed to unreviewed. In addition, some clinics do not track if physicians are acting timely on their results or not. These have negative effects on perceptual, cognitive, and motor performances due to the same reasons from the ‘Prioritize by Late and Lost Status’ guideline above.

The screenshot shows a table with columns: Patient, Test, Order Date, Review By, and Abnormality. The first row is highlighted in orange and contains: Evans, Anna, MRI / F, Mar 3, 2011 11:11 AM, Mar 10, 2011 11:11 AM. Below the table is a form for the selected patient, Evans, Anna. The form includes a list of names on the left, a 'No Follow-up' checkbox, an 'Ask Nurse to' section with a checked 'Inform Patient (Evans, Anna)' option and a 'Schedule Visit' checkbox, an 'Order' section with a 'Repeat Test' checkbox, and a 'Comments' text area. At the bottom of the form are buttons for '?', 'Review Later', and 'Complete'. To the right of the form is a detailed view of the result, including patient name, study date, comparison, history, and findings.

Patient	Test	Order Date	Review By	Abnormality
Evans, Anna	MRI / F	Mar 3, 2011 11:11 AM	Mar 10, 2011 11:11 AM	

NAME: Evans, Anna
STUDY DATE: 03/09/2011 4:28:29 PM
COMPARISON: None available
HISTORY: Pain
MRI OF THE RIGHT ANKLE (WITHOUT IV CONTRAST)
FINDINGS:
 The bones demonstrate no fracture or gross marrow abnormality. Articular surfaces are within normal limits.
 The spring ligament complex appears to be intact.
 The tibialis posterior, flexor digitorum, and flexor hallucis longus tendons are intact. The peroneus brevis and longus tendons are intact. The Achilles tendon is intact.
 The tarsal sinus is unremarkable. There are no significant fluid collections to suggest ligamentous injury.
 The remaining muscles and soft tissues are unremarkable.

Figure 5.3: “Complete” finalizes follow-up so the result could be removed from “Results to Review” rich table. On the other hand, “Review Later” button means the result was viewed and/or some actions may have been taken but the follow-up is not complete yet. Such a result is still kept in “Results to Review” table for further processing but is marked as “viewed”. In the context of rich tables, Actions for Rapid Completion are close to the mouse-click location and results appear on the right.

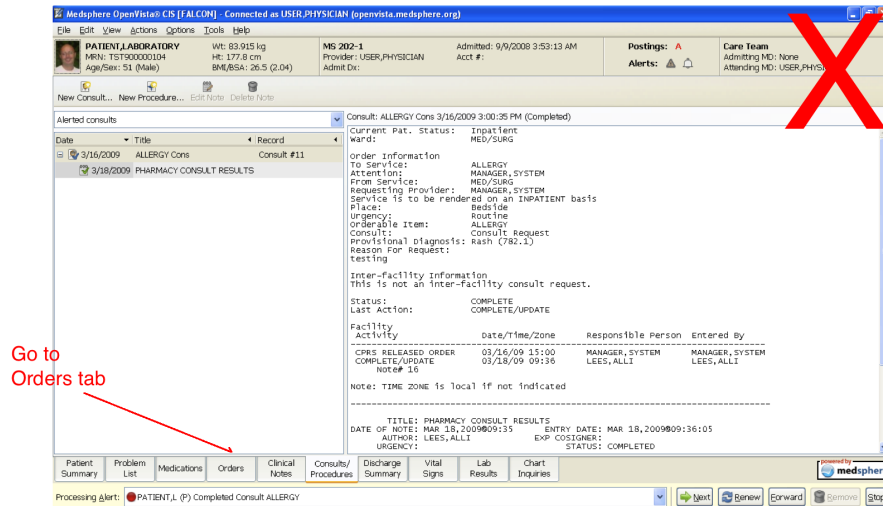
Instead of excessive marking for unread results, the system should prompt

physicians to acknowledge that the results have been reviewed (results with a white background in Figure 5.2 were marked with “Review Later” button in Figure 5.3). The table should keep the results until physicians explicitly indicate completion (“Complete” button in Figure 5.3). If the same principles for order timeliness are applied to physician review and follow-up step, physicians’ work can be marked late or not completed, e.g. Figure 5.2. A study showed the effectiveness of this approach (see Chapter 4).

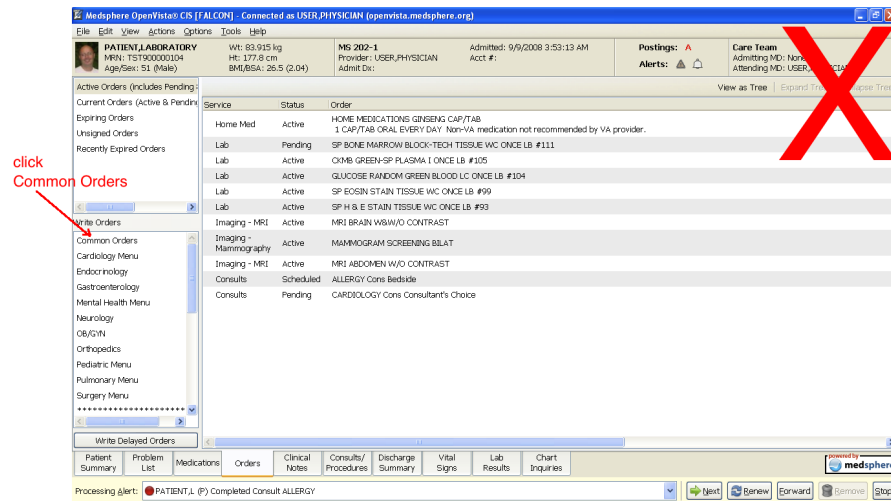
5.1.4 Embed actions when appropriate (Evidence: Medium; Conformance: Recommended)

While there are some better examples, most patient record windows (Figure 5.4a, 5.4b, 5.4c, 5.4d, 5.4e) require users to navigate to different windows to take actions because some results are complex or abnormal and the user must check multiple information resources. The perceptual help is low because clinicians need to look elsewhere to find relevant information. Cognitive load is medium since the clinicians must keep in mind the details of the report. In terms of motor performance, clinicians have to switch context, open multiple dialog boxes to take the actions, which necessitate mouse movements of long distances and are multiple clicks away.

While some results require careful review in separate windows with access to patient histories, there are many situations where actions can be taken rapidly [56], e.g. for normal results of routine orders for well-known patients in a primary care office. These situations can be determined by clinicians’ familiarity with their patients or orders. If physicians know certain patients or orders well, they can quickly decide what to do based on the result. There are two dangers with this approach: (i) Clinicians will not check the details of the patient record or other

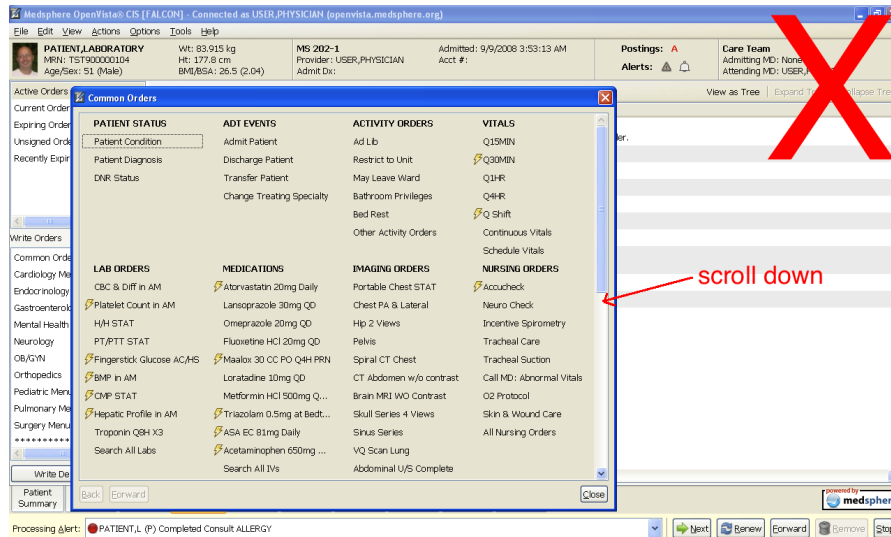


(a) A double click on the results opens the tab that has the consultation report in the patient record.

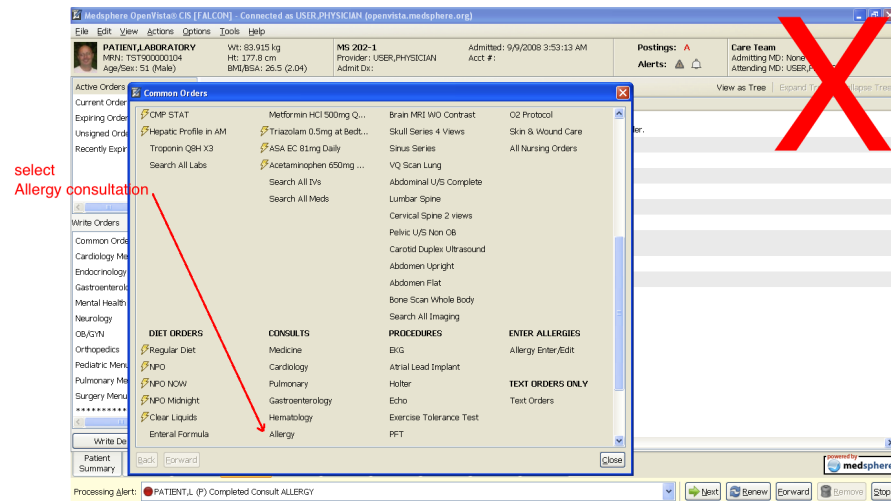


(b) To place an order, users switch to Orders tab.

Figure 5.4: Steps (a)-(e) correspond to the different screens and interactions to order a repeat consultation.

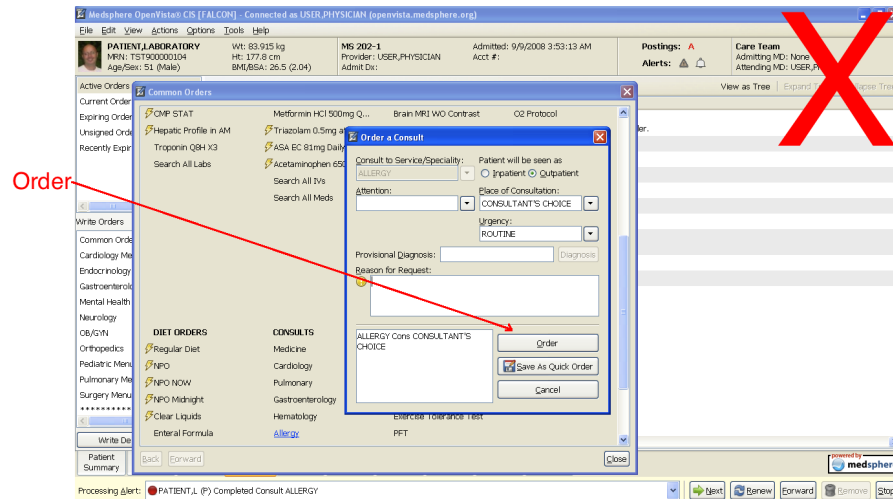


(c) To write an order, Common Orders are brought up.



(d) The desired order type is found by scrolling.

Figure 5.4: Steps (a)-(e) correspond to the different screens and interactions to order a repeat consultation.



(e) Clicking on the order pops up a new dialog to write the details of the order and place it.

Figure 5.4: Steps (a)-(e) correspond to the different screens and interactions in OpenVista to order a repeat consultation as part of a follow-up.

orders, (ii) While clinicians complete simple cases faster, they can leave complex cases for later.

This guideline proposes to allow users to take actions within the results table. Figure 5.3 shows an example where the possible follow-up actions are shown alongside the result report. It still gives quick access to the conventional approach (i.e. patient record) with a double-click on the result. A study showed the effectiveness of this approach (see Chapter 4).

5.1.5 Provide retrospective analysis (Evidence: Medium; Conformance: Recommended)

While most systems log events along with their timestamps, they do not provide retrospective analysis of past data. Hence no aid in perception. If clinics want to improve their results management continuously, managers make use of excel sheets

to write formulas that compute some statistics. Thus, clinic managers have high cognitive load as they periodically make decisions based on these statistics.

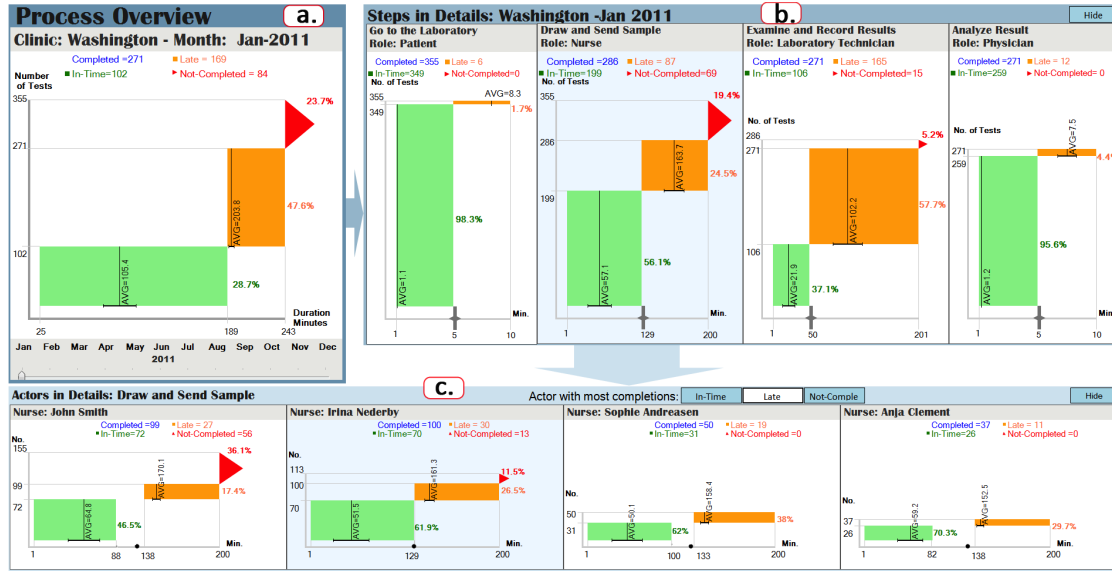


Figure 5.5: This retrospective analysis visualization has three views: (a) *Process Overview* shows the summary of the entire process (blood test), (b) *Steps in Details* shows each step of the process (four steps), and (c) *Actors in Details* shows the performance of the individuals who performed the selected step (“Draw and Send Sample”). Each view uses one or more Process Completion Diagrams (PCD). A PCD consists of two rectangles (separated by the threshold of lateness) and a triangle. Green and orange are in-time and late performances, respectively. Red is not completed orders. Height is the number of orders and width is the min and max completion times.

A retrospective analysis visualization that categorizes completion times for each step into in-time, late, and not completed can be later used to inspect bottlenecks as well as the best and worst performers (Figure 5.5). If the system employs a workflow model, this information can indeed be used to adjust expected durations of each step. This results in fewer or more late or not-completed orders. A usability test showed the effectiveness of this approach (see Retrospective Analysis of Medical Orders: MSProVis).

5.1.6 Distinguish preliminary and final results (Evidence: Low; Conformance: Mandatory)

Sometimes the outside facility generates an initial report for physicians to review although the order has not been finished processing [92]. In current systems, this will be documented in a textual form (see the “Message” column of Figure 5.1) that could easily be bypassed by busy and distracted users. Therefore, the system lacks perceptual help.

Physicians’ responsibility is to review results but the outside facility is still responsible for finalizing reports. Thus, such results should appear in both pending and results tables but be clearly marked preliminary or final in the results table so that physicians know their status during review. For example, Figure 5.2 has ‘P’ and ‘F’ in the test column for preliminary and final results, respectively. Preliminary results also show up in pending results (e.g. Megan Reed’s preliminary mammogram result).

5.1.7 Support views for different clinician roles (Evidence: Low; Conformance: Recommended)

Various users look at and act on the same orders differently. Current systems either do not support this, make customization too complicated, or do not link separate views (i.e. an action completed in one view is not reflected in another view). This guideline is for enhancing perception and motor performance as users will be able to see orders easily and results management will take less time and effort.



Figure 5.6: Care provider, Joe Brown, at the Riverside Clinic is currently signed in.

Figure 5.6 indicates the logged in user. Physicians or residents review and follow-up on results, nurses regularly check if pending orders of patients coming in today have arrived, managers overview the clinic and forward results to alternative clinicians if needed (e.g. in case of physician illness). The table contents and possible actions depend on the role of this user. Lateness information should be available to all users of the system.

5.1.8 Clarify responsibility (Evidence: Low; Conformance: Recommended)

None of the results management windows (Figure 5.1) have any information about who is performing the current step of the process or guidance to contact the responsible parties. There exists no perceptual help as clinicians spend hours on the phone tracking the results down. Cognitive load is high since the clinicians need to estimate who should take care of the current step of the process, and guess where to contact them. In terms of motor performance, clinicians may use a phone/e-mail directory search function within the system to find out the relevant information. Then, they might either write a message or call multiple places to understand what went wrong.

The current window should indicate the status (order status column for pending orders in Figure 5.2), which can then be expanded (the popup menu in Figure 5.7 that appears with a click on the row) to enumerate who did what and when, as well as the deadlines. This guideline depends on workflow management capabilities. If the actions of responsible parties are unrecorded, and the responsibilities are not transferred between various systems, the only feedback to provide is ‘Test ordered’. Many systems capture most of these data (with or without workflow management)

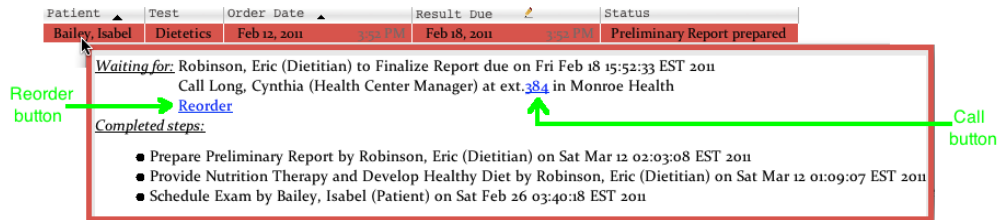


Figure 5.7: Popup menu for pending orders identifies the responsible person (dietitian, Eric Robinson) who is handling the current step of the order (finalizing the report) along with their deadline (February 18th, 2011) and the manager’s contact information (Cynthia Long’s phone extension is 384). Below, it has a reorder button for the lost consultation and illustrates the progress of the order step by step.

but what is missing is different systems do not communicate on the backend.

5.2 Actions for Rapid Completion (ARC) Design Guidelines

This section uses various terms for user interface components that can hold multiple widgets (labels, buttons, text boxes, drop-downs, tables). Screen refers to a computer monitor which may show one or more windows. Users can open, move, resize, minimize, maximize, or close windows. Windows are visible until users explicitly close them. A dialog or dialog box is a special window that is limited to communicating a message with users (e.g. present feedback to the user, prompt the user for a response). Users have to acknowledge or explicitly dismiss the message to close dialogs. Dialogs disable interaction with other windows. Popups or popup menus appear upon user interaction and may offer information and/or a limited set of choices that are available in the current state or context. Popups disappear when users select one choice or click away from the popup. Panels are generally part of windows but they do not always have to be. Unlike popups, they

allow multiple selections. They disappear when users confirm their selections or click/move outside of the panel. Tooltips appear when mouse hovers on a widget without clicking it and disappear automatically after some delay.

5.2.1 Allow In-place Editing for Cells (Evidence: High; Conformance: Mandatory)

If a table contains data users can edit, it is best to let users edit individual cells [3,121] without having to open a secondary window to speed up motor performance (Figure 5.8).

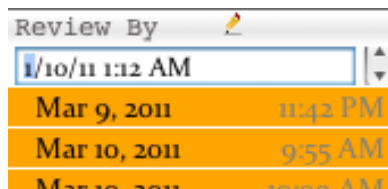


Figure 5.8: In-place edit for a table cell value.

5.2.2 Reveal a panel on hover or click of a row for more complex actions (Evidence: Medium; Conformance: Mandatory)

If there is more than one action to perform on a row, a panel can be revealed on hover or click [87] (Figure 5.3). Other options may be opening a dialog box or putting the actions on a separate window (Figure 5.9). A panel is more lightweight than these options in the sense that panels do not obscure the main window (context-switch), disallow access to the main window, or have to be dismissed with a “close” or “back” action. Note that a popup is not possible in this case because multiple choices need to be selected. Therefore, panels offer better perception, cognition, and motor performance compared to their counterparts. An affordance

to popup the ARC may be color or sound indication during hover. With frequent users a typical solution would be to have training to let them know.

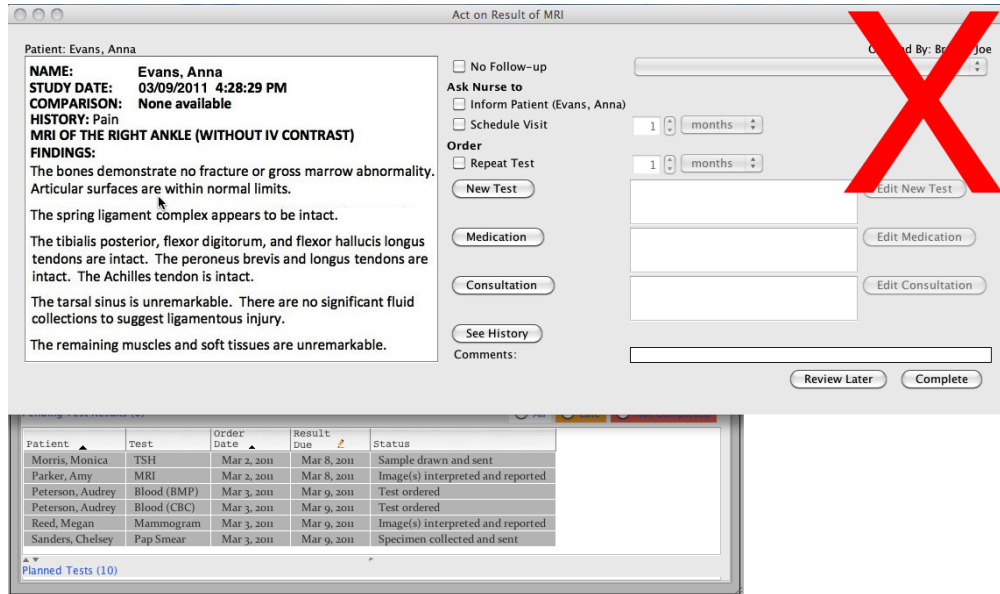


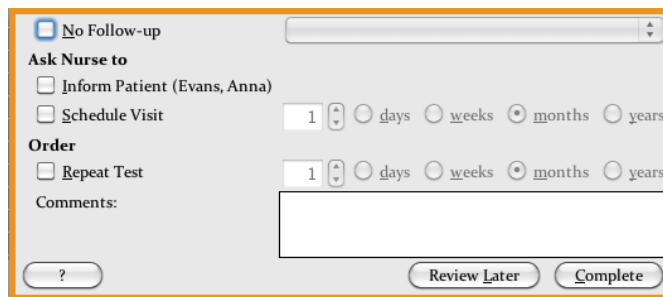
Figure 5.9: A dialog box enumerates actions that can be performed on a row. The main drawbacks are it obscures the table and the row, and needs to be closed to return to the previous window.

5.2.3 Make the entire row clickable if there is one action set on a row (Evidence: Medium; Conformance: Mandatory)

If specific columns have actions associated with them and each column has a disjoint set of actions, only those cells should be interactive [94]. For a single action that concerns the entire row, the entire row should be clickable (single-click) for faster motor performance. This is based on Fitts' Law [46] that implies wider targets are easier to click.

5.2.4 Support key accelerators to speed up action taking (Evidence: Medium; Conformance: Mandatory)

When speed is vital, a keyboard shortcut is an excellent strategy, which is especially appreciated by power users [121]. This significantly improves motor performance. Not only there should be a help button for key commands, key commands can also be highlighted on the panel to further increase perception. As a reminder, when users hold the activating key (such as Alt, Command, Control, and Shift), the key to select an action is underlined on the panel (Figure 5.10). Note that quick undo/redo operations are crucial.



The image shows a software interface panel with several sections. At the top, there is a checkbox labeled 'No Follow-up' with the letter 'N' underlined. Below this is the section 'Ask Nurse to' containing two checkboxes: 'Inform Patient (Evans, Anna)' with 'I' underlined, and 'Schedule Visit' with 'S' underlined. To the right of these checkboxes are input fields for a number (set to '1') and radio buttons for 'days', 'weeks', 'months', and 'years'. The 'months' radio button is selected. Below this is the 'Order' section with a checkbox 'Repeat Test' where 'R' is underlined, and similar input fields for a number and time units. At the bottom left is a button with a question mark '?'. At the bottom right are two buttons: 'Review Later' and 'Complete', where the 'C' in 'Complete' is underlined. A large text area for 'Comments' is located between the 'Order' section and the bottom buttons.

Figure 5.10: An activating key reveals all letters underlined to press for shortcuts.

5.2.5 Spell everything out while keeping the content compact for visual scanning (Evidence: Medium; Conformance: Mandatory)

The goal of a panel is to improve efficiency and reduce memory load so that the most frequent actions are one-click away. Designers need to keep the content compact for visual scanning and simplify the language to achieve better perception [121]. For cognition, minimal text with no abbreviations is desired.

5.2.6 Support undo/redo of actions (Evidence: Medium; Conformance: Mandatory)

Any action – including saved ones – should be reversible [121]. An undo/redo option must be supported not only shortly after the action has been taken but also from the table where completed items are kept.

5.2.7 Avoid drop-downs that necessitate multiple clicks to access and select (Evidence: Medium; Conformance: Recommended)

Selections should be made with a single click on radio buttons, check boxes, etc. Drop-downs (a.k.a. combo boxes) require selection to see a list of items, move and scroll, and click on the desired item. Drop-downs introduce more errors, impair perception of the items and motor performance [91]. However, with multiple widgets laid out on the panel, mouse may travel longer distances than with drop-downs. Figure 5.11 has only one drop-down, which lists long words that would not fit on such a small panel.

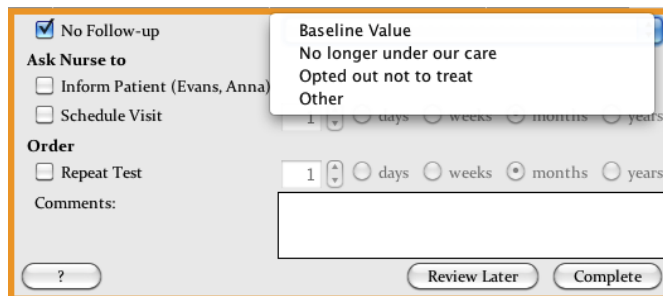
The image shows a screenshot of a web-based form. On the left side, there are several sections: 'No Follow-up' with a checked checkbox, 'Ask Nurse to' with two unchecked checkboxes labeled 'Inform Patient (Evans, Anna)' and 'Schedule Visit', 'Order' with one unchecked checkbox labeled 'Repeat Test', and a 'Comments:' label above a text input field. On the right side, there is a 'Baseline Value' drop-down menu that is open, showing a list of options: 'No longer under our care', 'Opted out not to treat', and 'Other'. Below the drop-down menu, there is a numeric input field with the value '1' and radio buttons for 'days', 'weeks', 'months', and 'years'. At the bottom of the form, there are three buttons: a question mark icon, 'Review Later', and 'Complete'.

Figure 5.11: Drop-downs require a click to see the list of items, cursor move, and a click on to the desired item.

5.2.8 Provide sounds for feedback (Evidence: Medium; Conformance: Recommended)

To provide extra feedback and support accessibility, a sound may give confirmation of the action status [121] by extending the perception and cognition. When “Complete” is clicked (Figure 5.3), a sound gives confirmation. This may not be of help to users with hearing disabilities.

5.2.9 Keep the panels and popups on the screen until the user intentionally dismisses them (Evidence: Low; Conformance: Mandatory)

Users should have as much time as they need to process all the information in the popups and select their actions on the panels. This is suggested for improvements in perception and cognition. Because panels or popups do not have a close or a cancel button, they are dismissed on click/hover somewhere else on the screen than the panel/popup itself. This will close both the panel and popup (if any) simultaneously without affecting anything in cases when users unintentionally clicked on a row by mistake or want actions to remain unsaved. This has benefits for motor performance. For example, the mouse cursor is on the panel in Figure 5.12 but as soon as the mouse leaves the panel and popup, both of them disappear.

5.2.10 Keep the row visible when the panels are shown (Evidence: Low; Conformance: Mandatory)

No matter where the panels or popups open, the row provides context and acts as an information resource for users to choose their decisions on the panels. This is useful for perception.

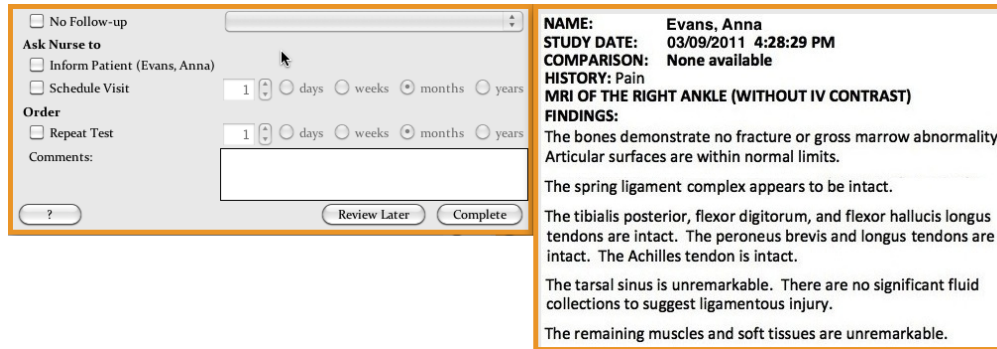


Figure 5.12: The panel (left) and popup (right) stay visible until the mouse moves outside of their borders.

5.2.11 Highlight the selected row while unhighlighting the rest of the data in all tables (Evidence: Low; Conformance: Mandatory)

The selected row should be highlighted while the rest of the data in all tables gets grayed out (Figure 5.3). With a sheer amount of data on a tabular display, this dissipates the distracting information and brings the focus onto the item which is being handled. Perception is under consideration for this guideline.

5.2.12 Rows that appear in more than one table should be linked (Evidence: Low; Conformance: Mandatory)

If the same row appears in more than one table, any interaction with one of the rows should also bring up the other one. For example, Amy Parker's preliminary late MRI result is shown in both tables and selecting one highlights the other one (Figure 5.13). This is important for perceptual performance.

The screenshot shows the MSTART Simulator interface. At the top, it displays 'MSTART File' and 'MSTART Simulator'. Below that, it shows 'Physician Joe Brown reviews results from his inbox several times a day' and 'Current Time: Thursday, March 10, 2011 6:33 PM'. The interface is for 'Riverside Clinic' and 'Provider: Brown, Joe'. There are buttons for 'Pause', 'Continue', and 'Sign out'. The main area is divided into 'Order', 'Track', and 'Complete' sections. The 'Results to Review (17)' section contains a table with columns: Patient, Test, Order Date, Review By, and Abnormality. The 'Pending Test Results (6)' section contains a table with columns: Patient, Test, Order Date, Result Due, and Status. A row in the pending table is highlighted, and a popup window displays details for that row.

Patient	Test	Order Date	Review By	Abnormality
Evans, Anna	MRI / F	Mar 3, 2011 1:12 AM	Mar 10, 2011 1:12 AM	
Parker, Amy	MRI / P	Mar 2, 2011 11:42 PM	Mar 9, 2011 11:42 PM	
Reed, Megan	Mammogram / P	Mar 3, 2011 9:55 AM	Mar 10, 2011 9:55 AM	
Stewart, Kim	Mammogram / F	Mar 3, 2011 10:09 AM	Mar 10, 2011 10:09 AM	
Cooper, Molly	X-Ray / F	Mar 8, 2011 10:22 AM	Mar 15, 2011 11:22 AM	⚠
Howard, Amanda	Blood (BMP) / P	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	⚠
Howard, Amanda	Blood (CBC) / P	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	⚠
Phillips, Sarah	TSH / F	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Phillips, Sarah	Pap Smear / P	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Richardson, Allison	Blood (CBC) / P	Mar 7, 2011 11:07 PM	Mar 15, 2011 12:07 AM	⚠
Bennett, Claire	Pap Smear / F	Mar 7, 2011 8:01 PM	Mar 14, 2011 9:01 PM	
Cox, Pamela	Blood (BMP) / F	Mar 8, 2011 9:34 PM	Mar 15, 2011 10:34 PM	

Patient	Test	Order Date	Result Due	Status
Morris, Monica	TSH	February 15, 2011 2:22 PM	February 21, 2011 2:22 PM	Sample drawn and sent
Parker, Amy	MRI	March 2, 2011 11:42 PM	March 8, 2011 11:42 PM	Image(s) interpreted and reported
Peterson, Audrey	Blood			
Peterson, Audrey	Blood			
Reed, Megan	Mam			
Sanders, Chelsey	Pap S			

Waiting for: Jackson, Paul (Radiologist) to Finalize Report due on Tue Mar 08 23:42:11 EST 2011
Call Powell, Amy (Imaging Center Manager) at ext. 329 in Accu Imaging

Completed steps:

- Interpret and Report Image(s) by Jackson, Paul (Radiologist) on Thu Mar 03 11:14:11 EST 2011
- Take Image(s) by Wilson, Robert (Radiology Technician) on Thu Mar 03 05:45:11 EST 2011
- Schedule Exam by Parker, Amy (Patient) on Thu Mar 03 04:31:21 EST 2011

Figure 5.13: Selecting one item in the pending table highlights the same item in the results table.

5.2.13 Open panels as close to the mouse click location as possible (Evidence: Low; Conformance: Mandatory)

With a click on a row, the panels and popups should appear either just a little below or above for perception and cognition motives. Putting them far away would have a negative effect on motor performance because the eye and mouse will need to jump a greater distance between the rows and the panels. Figure 5.3 presents popups along with the panels but to the right so as to give quick access to the actions. On the other hand, opening the panel at a specific location might prove useful for consistency.

5.2.14 Make sure the panels and popups are visible at all times (Evidence: Low; Conformance: Mandatory)

This means that popups should not overlap with panels. In addition, neither of them must go off the screen (Figure 5.14). The system should automatically compute the optimum location for the panel within the screen space without obscuring the row or the popup. This guideline concerns with perception mostly.

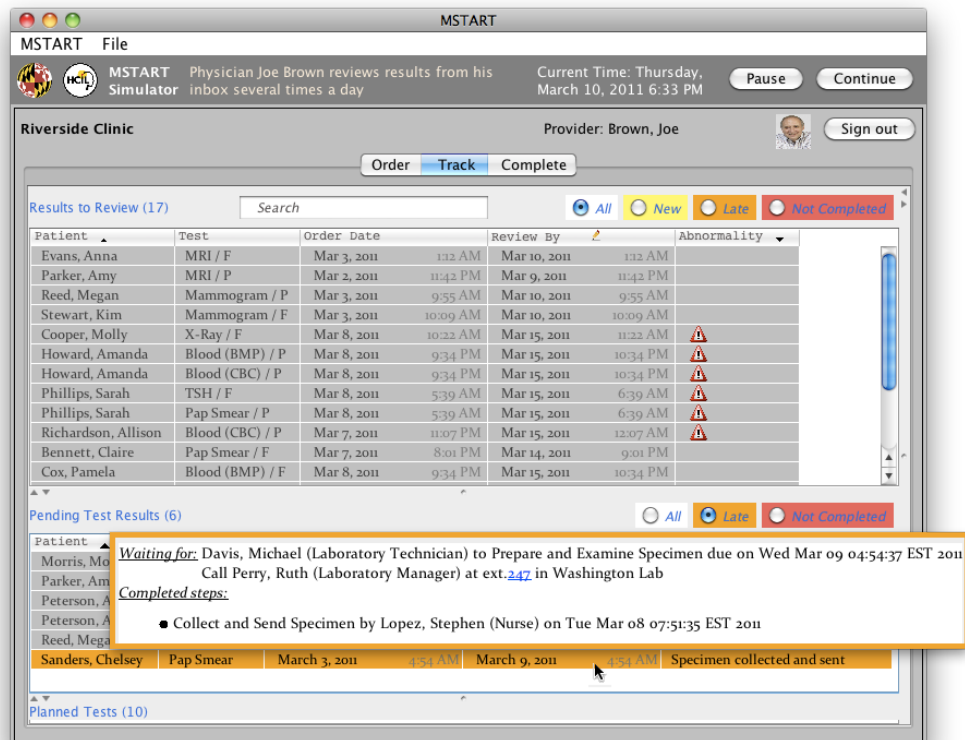


Figure 5.14: Assuming that there is no screen space available below this application window, the location is computed so that the popup does not go off the screen and the row is still visible. Normally, the popups display right beneath the point of click location, however, the mouse cursor is very close to the edge.

5.2.15 Save and allow access to completed actions (Evidence: Low; Conformance: Mandatory)

Items with completed action status should be saved automatically. This helps with perception of completed items. Completed items can be kept in the same table but may increase the table size significantly over time as more items are completed, impairing motor performance. One option is to provide a filter to show/hide completed items. Another option is to move them to a separate table where users can access when desired (see “Complete” tab at the top of the tables in Figure 5.2). In the latter case, animation can be used to illustrate the item is moving to another table.

5.2.16 Present a popup of row information along with the actions (Evidence: Low; Conformance: Recommended)

Sometimes there is external data about the row that does not fit in a column (e.g. an image) or is intentionally hidden. Users who take actions on such rows should know all the data to make the best decisions. In these cases, a popup menu showing this information along with the actions is helpful for perception and cognition of the extra information (see the MRI report in Figure 5.3). Another way to implement this guideline is by interleaving the actions with the information in the popup menu (Figure 5.7). Then, accessing the actions may require more mouse travel. This is only optimal if there are not too many actions and these actions can be put close in proximity.

5.2.17 Color the panel and popup the same as the color coding of the row (Evidence: Low; Conformance: Recommended)

Row colors could be missed when new panels or popups appear as panels and popups may partially hide the table. It does not hurt to remind the color coding. For color coded tables, colors should be repeated in the panels and popups. To avoid extensive color usage, a good choice is border coloring (Figure 5.3). Perception is reinforced here.

5.3 Summary

These guidelines were developed by studying existing systems, interviewing medical informatics experts, demonstrations to expert users, and reviews by HCI professionals. These guidelines were sent to researchers for review and seven of them replied with detailed feedback. There were three negative and four positive comments; the rest were suggestions for improvement within the document. Here are four positive followed by three negative comments from researchers:

The guideline content looks great! There is a lot of useful information in them that we could use for mocking things up. If we think of these guidelines in terms of the National Institute of Standards and Technology use cases, there is great potential for implementation. Incorporating these guidelines to our current mockups and design would further improve readability and interaction.

You have compiled an impressive list of guidelines and supportive evidence that strike me as valuable on face value as well, as a family physician.

I think this a great beginning. There has been a lot of work put into these guidelines and there are a lot of really great points.

Overall, the guidelines are well-done and helpful for human factors professionals. Some would be helpful to non-human factors professionals. Most guidelines are well-written and would be useful for the human factors professional working on electronic health/medical records.

My main suggestions are aimed at making it easier to comprehend for software engineers and managers who lack a lot of human factors training.

I see now the document is very specific to the Veterans Affairs Electronic Health Record system. If the goal is to make this more generalizable I would focus less on the “bad” examples and the errant problems and provide more and better examples of the various list information found in tables from medication lists, to problem lists, to lists of orderable items in Computerized Physician Order Entry.

I don’t think many folks from the vendor side who are making these decisions will take to the time to read the document. I wish it was different but, I think that is the way it is.

Tables 5.1, 5.2, 5.3 summarize all 90 guidelines (8 results management guidelines, 65 table design guidelines, 17 actions for rapid completion guidelines) along with their evidence and conformance ratings. For each guideline the perceptual, cognitive and motor skills of users were considered, with the goal of reducing the load on users so as to speed performance while reducing errors. Each guideline suggests a direction for research to validate and refine it, especially to refine it for

new platforms such as small or large displays, use of voice controls, use by medical professionals with disabilities, and other special situations.

Table 5.1: Summary of Results Management design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).

Results Management design guideline	Evidence	Conformance
1. Show pending results	H	M
2. Prioritize by late and lost status	H	M
3. Indicate physician acknowledgment and timeliness	M	M
4. Embed actions when appropriate	M	R
5. Provide retrospective analysis	M	R
6. Distinguish preliminary and final results	L	M
7. Support views for different clinician roles	L	R
8. Clarify responsibility	L	R

Table 5.2: Summary of table design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).

Table design guideline	Evidence	Conformance
Data Arrangement		
<i>Columns / Rows:</i>		
1. Sort the table according to one or more column(s) by default, arranged vertically down	H	M
2. Permit re-sorting of tables with a click on the column header	H	M
3. Avoid horizontal scrolling in the default view	H	R
4. Focus on the data itself	M	M
5. Use sort icons in column headers to communicate that the table is sortable; conventionally upward/downward arrow for ascending/descending values while the arrow size indicates sort priority	M	M

6. Perform computations for users; value, derived from data, should be readily available in the cells	L	M
7. Reduce the number of columns whenever possible	L	R
8. Remove a column that always has the same value to save space	L	R
9. Use endless scrolling when all results do not comfortably fit within one page	L	R
10. Combine columns when appropriate	L	R
<i>Row Sequence:</i>		
1. Put the most severe row at the top of the table while ensuring that the most important rows are still visible	M	M
2. Group related rows together so they are close in proximity for comparisons	L	R
<i>Column Sequence:</i>		
1. Offer rearranging of columns	H	R
2. Place sets of categorical values to the left of the quantitative values associated with them	M	M
3. Place columns containing data that should be compared close together	M	M
4. Organize the most important columns on the left to permit reading in the conventional left-to-right order	M	R
<i>Related Tables:</i>		
1. Use just enough space between tables to make them noticeable	M	R
2. Size the tables according to their frequency of usage	M	R

3. Filter a table that is not used often to show only the important data or stretch the table with support for full-view or expansion on-demand	L	M
4. Order the tables according to their frequency of usage	L	R
5. Keep table structure consistent from table to table	L	R
Labeling		
1. Style headers differently but ensure consistency	H	M
2. Give the table a descriptive title with a total row count	H	M
3. Keep the headers visible in the window at all times	H	M
4. Align column headers with their associated data	H	M
5. Avoid a header that is significantly wider than the data it is indicating by spreading such headers into two or more lines	M	M
6. Indicate editable columns	L	M
7. Show a tooltip for the title that describes the table's function	L	R
8. Do not truncate column headers; break long headers by full words whenever possible, otherwise split in the middle with a hyphen	L	R
Settings & Help		
1. Provide custom filtering on-demand	H	M
2. Allow settings to be saved and loaded	M	R
3. Do not let tooltips go outside of screen space	L	M
4. Provide help in a separate window	L	M
5. Derive filter values from current table entries, not all database entries	L	R
6. Group filter values by range if possible	L	R

7. Support search for large tables	L	R
8. Show a description for columns in a tooltip	L	R
Delineation		
1. Feature light white space between the rows, no heavy gridlines	H	M
2. Allow sufficient space between columns to clearly separate them but no more	H	M
3. Calculate initial column widths from data but offer resizing of columns	M	R
4. Define a min and max width for each column depending on the data it presents; these widths should be user changeable	M	R
5. Do not stretch tables to fill available space; only the last column may be stretched, if not right-aligned, to align with other tables	M	R
6. Add some padding to columns if there is room; allow user-defined values	M	R
7. Add some padding to rows for easy visual scanning; allow this to be set by the user	M	R
Formatting		
<i>Orientation:</i>		
1. Avoid text orientations other than horizontal, left to right	M	M
<i>Alignment:</i>		
1. Ensure consistency in the alignment of similar data	M	M
2. Align the time and date markers for all numerals in a column	M	R
3. Bottom-align column headers	L	M
4. Align numeric (time) values to the right while keeping all other values in the table left-aligned	L	R

5. Left-align columns with icons for horizontal continuity	L	R
6. Double-align the day and time in a date column	L	R
<i>Number and Date Precision:</i>		
1. Round values displayed in table cells where it is not misleading to do so	M	M
2. Use abbreviation for those values that have the same substring	M	R
3. Truncate values if abbreviation does not apply, but use tooltips for showing details of values that do not fit in the cell	M	R
4. Gradually show more precision if space permits or users seek	L	R
5. Keep the precision consistent from column to column	L	R
6. Split cells into two lines when a value is too long after abbreviation and truncation	L	R
<i>Font:</i>		
1. Select a font family and size that is legible	H	M
2. Ensure consistency in the typeface	M	M
3. Make table values prominent with a font style that is easily distinguishable from others used throughout the interface	M	R
4. Do not drop to a smaller font to fit a table on the window	L	R
<i>Color and Icons:</i>		
1. Change color saturation to make important information stand out	M	R
2. Use icons whenever possible	M	R
3. Provide a color legend to show and filter based on the coding	L	M
4. Use color coding to show grouping	L	R

Table 5.3: Summary of Actions for Rapid Completion (ARC) design guidelines and ratings for evidence (High, Medium, Low) and conformance (Mandatory, Recommended).

Actions for Rapid Completion (ARC) design guideline	Evid- ence	Confor- mance
1. Allow in-place editing for cells	H	M
2. Reveal a panel on hover or click of a row for more complex actions	M	M
3. Make the entire row clickable if there is one action set on a row	M	M
4. Support key accelerators to speed up action taking	M	M
5. Spell everything out while keeping the content compact for visual scanning	M	M
6. Support undo/redo of actions	M	M
7. Avoid drop-downs that necessitate multiple clicks to access and select	M	R
8. Provide sounds for feedback	M	R
9. Keep the panels and popups on the screen until the user intentionally dismisses them	L	M
10. Keep the row visible when the panels are shown	L	M
11. Highlight the selected row while unhighlighting the rest of the data in all tables	L	M
12. Rows that appear in more than one table should be linked	L	M
13. Open panels as close to the mouse click location as possible	L	M
14. Make sure the panels and popups are visible at all times	L	M
15. Save and allow access to completed actions	L	M

16. Present a popup of row information along with the actions	L	R
17. Color the panel and popup the same as the color coding of the row	L	R

Chapter 6

Retrospective Analysis of Medical Orders:

MSProVis

*“The visual-information-seeking mantra:
Overview first, zoom and filter, then details on demand.” –
Shneiderman (1996)*

While MStart (Multi-Step Task Analyzing, Reporting, and Tracking) helps users track and act on their ongoing processes, there remains the question of retrospectively analyzing all archived orders in the past months. Clinic managers are responsible for continuous improvement of their processes. They pose questions, such as “Are orders processed more timely this month than the last month?”, “Who are completing their orders on time?”, “Which laboratories are dropping orders?” to make decisions and take corresponding actions. With MSProVis (Multi-Step Process Visualization) interactive visualization¹ (Figure 6.1) [97,96], managers can study an overview of the process, durations of each step to identify the bottlenecks (in-time, late or not-completed orders), performances of individuals to compare the best and worst performers, changes in process and actor performances when the value of the threshold of lateness increases or decreases, and differences over time. Pilot studies with two physicians and five novices showed that users were

¹This work was done in partnership with Kostas Pantazos during his 6-month visit to Human-Computer Interaction Lab.

able to efficiently use MSProVis to answer questions about a medical process. Two surveys explored different design choices.

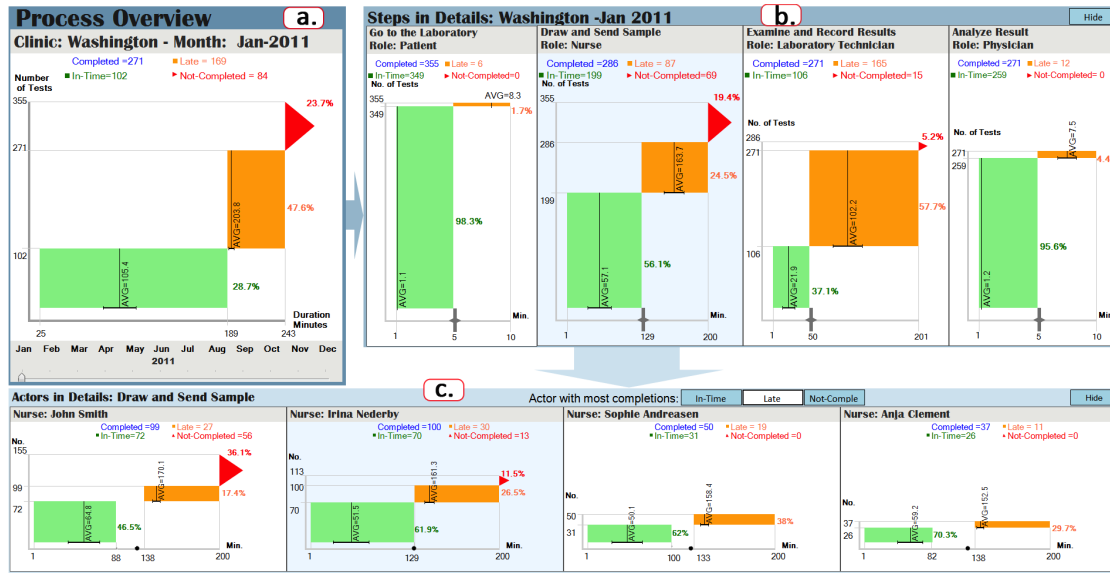


Figure 6.1: MSProVis has three views: (a) *Process Overview* shows the summary of the entire process (blood test), (b) *Steps in Details* shows each step of the process (four steps), and (c) *Actors in Details* shows the performance of the individuals who performed the selected step (“Draw and Send Sample”). Each view uses one or more Process Completion Diagrams (PCD). A PCD consists of two rectangles (separated by the threshold of lateness) and a triangle. Green and orange are in-time and late performances, red is not completed orders. Height is the number of orders and width is the min and max completion times.

6.1 Introduction

Advancing technology has accelerated the growth of stored data, but extracting meaningful insights remains a substantial challenge. The challenge is amplified in dynamic environments, where multi-step processes are present. Examples include the medical orders in healthcare, the paper-review in academic journals, the return merchandise authorization in the electronics industry, the package shipping

process in the logistics industry, personal financial planning, the internal supplier agreement in IT companies, sales within a huge corporation, the bug fix process in software maintenance, etc.

A multi-step process is composed of a number of steps, where each step is executed by one or more actors from an organization. The performance of an actor can be *in-time* (when completion's duration is less than or equal to an expected duration time), *late* (when completion's duration is greater than the expected duration time), and *not-completed* (when completion's duration is undefined). Retrospective analysis can assist in comparing these processes, locating the critical steps, understanding the root-cause and facilitating managers in making better decisions. However, investigating completions (the raw data of event-logs) is a time-consuming and tedious process.

In multi-step processes, managers who monitor the process are overwhelmed by information, and often limited in the time to investigate data from different viewpoints. According to the "Bounded Awareness" theory, the limited cognition of humans and large volume of available information can harm the decision-making process [24]. Information visualization enhances human cognition by visually presenting abstract data [19] and revealing patterns, trends and outliers [121]. Visual analytics is the science of combining interactive visualizations with analytical reasoning techniques to enable users to understand their data, reflect more effectively, and make better decisions [65].

The bounded awareness challenge in multi-step processes can be addressed by introducing visual analytics. Exploring the data using a visual analytics approach can support managers in obtaining an overview, identifying the bottlenecks, and highlighting the root causes and contributing factors. In addition, it can deal with

fundamental questions such as: what is the average time for the whole process and each step; how many steps were performed in-time and late; are there consistent patterns for not-completed cases; which actor is causing the delay; how are processes and steps performed over different points in time, etc. Therefore, visual analytics has the potential to assist managers in making better decisions by exposing the timeliness of multi-step processes.

6.2 The Process Completion Diagram (PCD)

In a process such as lab tests, completions can be executed in-time, late or not-completed. To encode effectively the information regarding the in-time, late, and not-completed cases, the PCD uses shapes and colors, information encoding techniques suggested by Bertin [12]. The PCD is composed of two rectangles, one for the in-time (green rectangle) and one for the late (orange rectangle) completions, and one triangle for the not-completed (red triangle). The PCD uses a different shape to distinguish easier the completed from not-completed. These three shapes are placed in a time series plot, where the X-axis shows the completion duration and the Y-axis shows the number of completions. This is inspired by cumulative plots of fraction or percentage over time (or other parameter) that physicians are quite familiar with. Examples of these plots include Cox survival [26], Kaplan-Meier [64], cumulative hazard [37], and receiver operating characteristic (ROC) [161] curves.

Figure 6.2 presents a blood test example from the medical domain and illustrates how 454 blood tests are summarized and presented using the PCD. The X-axis shows that there are 224 in-time, 185 late, and 45 not-completed blood tests. On the Y-axis, in-time completion durations range from 1 to 50 minutes and late completion durations range from 50 to 200 minutes where 50 is the threshold of

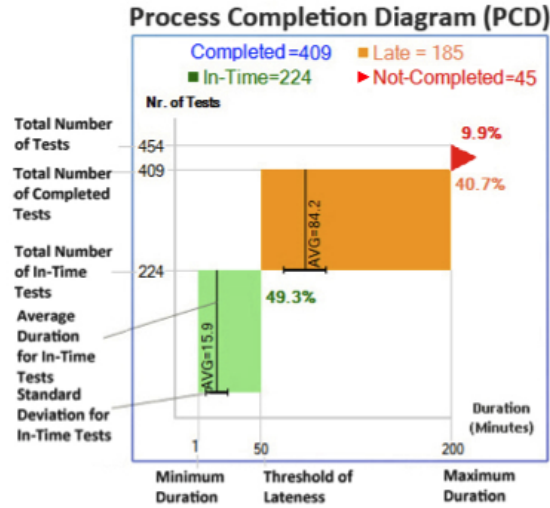


Figure 6.2: The Process Completion Diagram (PCD) visualizes data using shapes and colors in a time series plot, where the X-axis presents the duration and the Y-axis the number of tests. It also shows the average completion time, the standard deviation, minimum and maximum duration time, and the number of in-time, late and not-completed completions.

lateness. Thresholds are usually defined in the process specification. However, they may correspond to statistically computed values, such as the average duration.

The final version of the PCD is a result of an iterative design, where different representation aspects were considered. For instance, initially the axes were inverted, showing the number of completions on the X-axis, and duration on the Y-axis. Taking into consideration that time is usually presented horizontally (see examples in [1]), directed us to change the design.

6.2.1 Layout

The PCD design was guided by the desire to support managers to rapidly recognize tall-narrow green rectangles that indicate good performance or large orange rectangles and big red triangles that point out the problems. In this design, we

attempt to visualize data using simple visual objects that are familiar to any user. The PCD presents hierarchically the grouped completions. The orange rectangle and red triangle are placed at the top-right corner of the green rectangle (unless there is a gap) and the orange rectangle, respectively.

In the PCD, the size of shapes is calculated by aggregating all the event-log data from the database, and grouping them to in-time, late and not-completed cases. The height of shapes is bound to the number of event-logs in the database, and the width of the rectangles represents the duration. As the not-completed do not have a duration time, the triangle's width is not bound to the X-axis, and can expand up to predefined pixels number (based on available space). Furthermore, the X-axis is stretched till the right side of the orange rectangle (Figure 6.2). This aims to help managers understand easier that not-completed category does not have a duration.

Classifying completions into in-time and late is realized using the threshold of lateness. Figure 6.2 illustrates that the threshold of lateness can correspond to the maximum duration for the in-time completions, and the minimum duration for the late completions. However, this might not be always the case. The maximum duration of the in-time completions can be less, or the minimum duration of the late completions can be greater. As a result, there may be a gap between the green and the orange rectangle. The presence of the gap provides more detailed information, especially when managers compare individual actors' performances. An example can be seen in Figure 6.1.c.

6.2.2 Color Encoding

Ware [153] suggested a 12-color palette for presenting data, and we decided to use green, orange and red. These colors are consistently used in the shape and the label coloring. For example, the late rectangle and the related labels are colored in orange. Initially, we considered using yellow, rather than the orange, but on a white background yellow is not very readable. Green denotes good performance and red represents critical situations.

6.2.3 Detailed Information

The PCD is enriched with detailed information. At the top of the PCD, there are labels enlisting the total completed, which is made up of the in-time and late completions, and the number of not-completed. Small colored shapes are placed close to each label, to link the text with the shapes in the plot. The colored percentage labels for each performance type are positioned on the mid-right side of the shape, aiming to confirm assumptions derived from the shape's size.

The PCD provides managers with a visual representation of the average duration and the standard deviation for in-time and late completions. It uses a vertical line inside the rectangle to show the average duration, and a horizontal line, placed at the bottom of each rectangle, for standard deviation. The numeric value of the average is aligned properly (vertically or horizontally based on available space) and positioned close to the average line. Showing the average and the standard deviation indicates to managers the distribution of completions, i.e. a change in the threshold may have a high or low impact on the in-time or late completions. For example, let's assume the threshold of lateness is set to 79 minutes, the average duration of late completions is 137, and the standard deviation is 20. Adjusting

the threshold from 79 to 90 will not dramatically affect the size of orange rectangle (i.e. the number of completions considered late in the process). While moving the threshold bar close to the average value (i.e. 137), will significantly change the rectangles' sizes.

6.3 Multi-Step Process Visualization (MSProVis)

MSProVis (Figure 6.1) is an interactive visualization composed of three main views: *Process Overview*, *Steps in Details*, and *Actors in Details*. MSProVis allows managers to review and compare a series of PCDs at different levels of detail, allowing comparisons between steps or between actors executing those steps. MSProVis obtains default thresholds that define lateness, and allows managers to adjust those thresholds interactively. This visual approach aims to facilitate retrospective analysis, to identify problematic steps in the multi-step process, and to expose the best and the worst performers.

The design of MSProVis was driven by the need to efficiently present an overview of the multi-step process, and details-on-demand for steps and actors. It attempts to provide a space-efficient presentation by arranging the views as shown in Figure 6.1. The height, the header's font and color, and the position of each view represent the task analysis hierarchy: obtain an overview of the multi-step process, investigate the steps, and analyze actors' performances.

The Process Overview and Steps in Details are the most interactive parts. In the Process Overview, the managers can interact with the slider and view historical data for different months. Automatically, Steps and Actors in Details views show the steps and actors' performances for the selected month. MSProVis receives initial predefined thresholds of lateness from the process specification, and allows

managers to adjust those thresholds interactively in Steps in Details. Threshold changes are also reflected in the Process Overview and Actors in Details view. When the changes are saved back to the process specification, MStart is updated to list more or less late processes to users.

6.3.1 Process Overview

The Process Overview uses only one PCD to summarize the overall process performance. The process duration is computed using all steps' durations. The threshold of lateness for the process is calculated using the threshold of lateness defined for each step in the process.

Figure 6.1.a gives an overview of a blood test process using data from January 2011. Managers click on a shape to view details-on-demand (e.g. number of tests, minimum duration) in a tooltip. They interact with the slider at the bottom of the area to see the data of another month. Consequently, they gain insights on a specific month, but also are able to compare the overall process performance for different months. Next, they double-click on the *Process Overview* and the *Steps in Details* view (Figure 6.1.b) appears.

6.3.2 Steps in Details

A process can have one or more steps. Steps in Details uses one PCD for the performance of each step. In the example, the blood test process consists of four steps: “Go to Laboratory”, “Draw and Send Sample”, “Examine Specimen and Report Results”, and “Analyze Report”. Therefore, *Steps in Details* is composed of four PCDs (Figure 6.1.b).

The in-time and late completions do not affect the performance of the pro-

ceeding steps, although they clearly impact the performance of the overall process. Not-completed in one step has an impact on the next step, as the total number of performed tests in the next step is decreased. For example, 355 were performed in step two, while in step three only 286 tests, as 69 tests were not-completed from step two (Figure 6.1.b). Positioning the PCDs in order, using the same normalized scale for the number of completions, and aligning them on the Y-axes of the *Process Overview* allows managers to quickly view how the number of completions changes along the steps, compare each performance with the overall process performance, and identify bottlenecks and good performances.

Different steps may have different lengths of execution time (10 minutes for step one, 210 minutes for step two in Figure 6.1.b). This means that timescales used for the X-axis may differ from one step to another. Time-oriented visualizations face the challenge of presenting data in different timescales and using different granularity of time [1,65]. In *Step in Details*, the PCDs use consistently the same granularity of time (e.g. minutes). However, as space is limited and timespans may differ significantly, we decided to divide the space based on the number of steps and slightly adjust them based on the timespans. This attempts to foster managers to rapidly realize the different timescales.

Managers can interact with thresholds of lateness in the PCDs of *Steps in Details*. This means that the manager can reconsider what should be treated as in-time and late completions for a step. When the visualization is opened for the first time, the threshold of lateness in each step has a default value from the workflow specification. The default values usually correspond to estimations made from process management. In the example (Figure 6.1.b), in the second step (*Draw and Send Sample*) the default threshold is 129. As a result, the PCD shows 199

in-time (where duration time is lower or equal to 129 minutes), 87 late (where duration time is higher than 129) and 69 not-completed. The managers adjust the thresholds interactively by moving the threshold line, shown as a thicker gray line with two small arrows over the X-axis. When the threshold value changes, *Process Overview* and *Actors in Details* are updated. This feature lets managers investigate several scenarios and extract insights from the data. The thresholds of lateness are also used to adjust the triggering of alarms in MStart. Next, the managers double-click inside the second step (*Draw and Send Sample*) in *Steps in Details*, and *Actors in Details* is revealed (Figure 6.1.c).

6.3.3 Actors in Details

A step can be executed by more than one actor. *Actors in Details* presents the performances of each actor involved in a step, using a PCD per actor. It applies a descending ordering of the PCDs using the total number of completions, and can contain an unlimited number of PCDs. However, as space is limited, users have to use a scroll bar when there are more than seven PCDs. In the example, there are four PCDs, one for each nurse involved in step two (Figure 6.1.c). Managers use the simple menu with three buttons to highlight in light blue the actor with most in-time, late, and not-completed cases.

The PCDs use the same scale for the X-axis (duration) and the Y-axis (number of completions). More precisely, the timescales on the X-axis correspond to the timescale of the step (200 minutes). The maximum value on the Y-axis corresponds to the maximum number of completions performed by one of the actors (155 by John Smith).

A gap between the green and orange rectangle is possible, whenever the actors

maximum in-time performance and minimum late performance do not match the step threshold of lateness. For example, the actor *Sophie Andreassen* executed in-time completions in 100 minutes, and her minimum time for late completions corresponds to 133 minutes. While, the actor *Irina Nederby* drew and sent samples equal to the threshold of lateness. As a result, managers can rapidly identify who has caused the lengthy step.

The numbers of in-time, late, and not-completed results for each actor are computed by selecting those completions performed by the actor in the selected step. For example, the actor, *John Smith*, in Figure 6.1.c has 72 in-time, 27 late, and 56 not-completed completions. From this dataset, the PCDs show the minimum and the maximum durations based on the threshold of the selected step. The minimum duration (1 minute) is equal to the minimum duration performed by this actor. The maximum duration (88 minutes) for the in-time completions can be smaller than or equal to the threshold of lateness (129 minutes) of the step. The same principles apply for the minimum and maximum durations of the late completions. The minimum duration (138 minutes) can be much higher than the step's threshold of lateness, and the maximum duration corresponds to the highest duration performed by this actor (200 minutes).

6.4 MSProVis for Software Development Cycle

In Figure 6.3, MSProVis visualizes an Internal Service Agreement process in an IT company. In this process, completions are executed in-time, late, or very late. All completions have a duration, unlike the previous scenario from the medical domain, the triangle's width represents duration. Therefore, the PCD design is easily adaptable to different performance types. Similar to the medical domain,

the reason for using a different shape is to emphasize the critical situation. In comparison to the previous example, this process is composed of eight steps and managers interactively adjust two thresholds: the first divides in-time from the late completions, and the second separates the late from the very late completions.

6.5 Implementation

MSProVis is built in uVis [75], a formula based visualization tool for relational databases developed at the IT University of Copenhagen. uVis also supports user interface development, and allows users to extend their applications with visualizations.

uVis has an integrated development environment, where designers can create visualizations by dragging and dropping controls (e.g. triangle, rectangle, etc.) in the design panel, specifying spreadsheet-like formulas in the property grid, viewing immediate feedback, and trying end-user interactions without switching context. By means of formulas, controls can be bound to data, refer to other controls, define the appearance and behavior of each control. The formula language and the integrated development environment enabled us to rapidly create iterative visualization prototypes and led us to MSProVis.

6.6 Usability Studies

There were two studies conducted. The studies were performed with two medical doctors and five novice students. The studies were performed using a de-identified medical dataset of 1606 records, extracted from a four step medical process in a hospital. However, as this dataset did not include all the required information

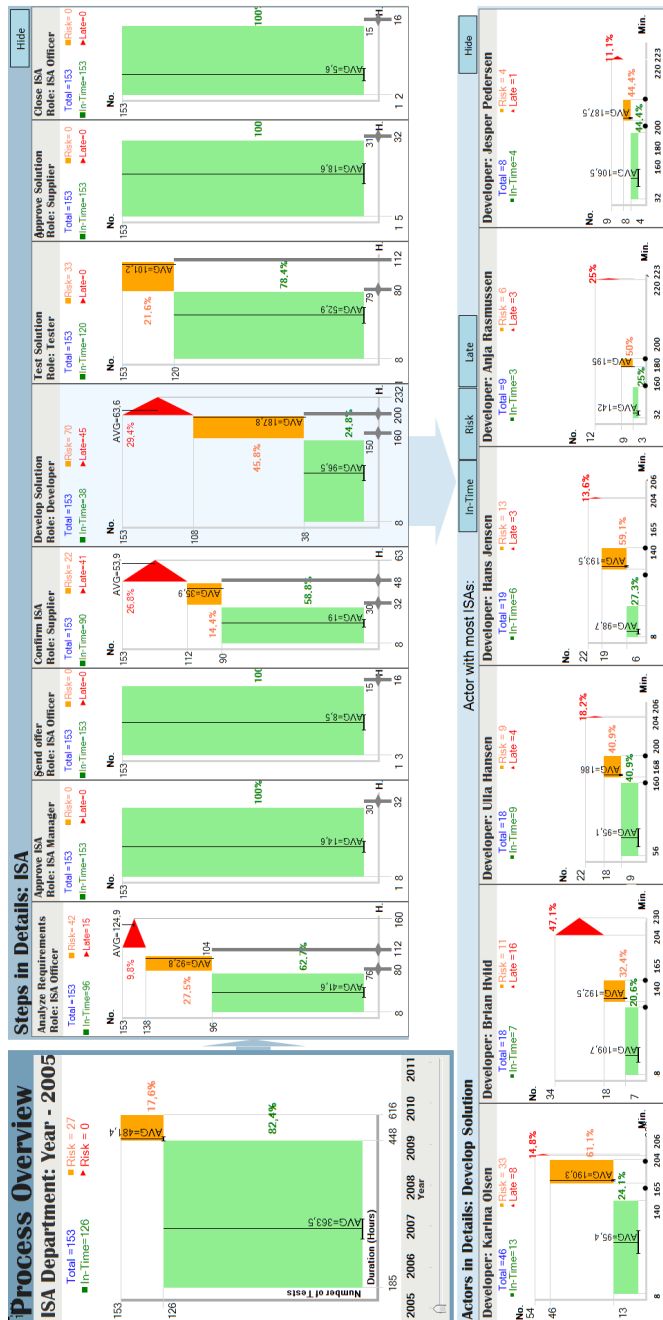


Figure 6.3: Internal Service Agreement process in an IT company from 2005 visualized in MSProVis indicates that the fifth step of the process, i.e. develop solution, has too many very late completions and the developer, Brian, is responsible for the lateness.

(e.g. not-completed tests, actors involved in each step), we processed this dataset and added synthetic data. The purpose of the studies was to investigate if users were able to understand the information presented in the PCD and to efficiently use MSProVis. In addition, these studies focused on identifying potential usability issues, and on providing suggestions for further improvement.

6.6.1 Study with Physicians

The PCD and MSProVis were evaluated with two medical doctors from the Copenhagen region. Currently, they were not clinic managers, but from their experience (25-30 years) they were very knowledgeable of how lab test results were handled or mishandled.

- *Understanding the PCD:* Physicians were able to distinguish completed from not-completed by means of shapes, and the color encoding helped in distinguishing the good from the bad performances. One of them reported, “The colors are well-chosen, and help me more than the shapes to distinguish performances.” He could easily read the information in the Y-axes and X-axes, but the triangles without durations was not immediately understood. He appreciated the way of presenting the average, and it helped him to correctly guess the annotation for the standard deviation. He added “I think clinicians might prefer placing the line in the middle of the average, rather than at the bottom.” The labels assisted him in deriving the correct assumptions, and he found their color-coding appropriate, “It is good to color the labels to the corresponding shapes.”
- *Using MSProVis:* Once physicians opened MSProVis, the Process Overview was shown. It showed an overview of all tests for January 2011. They used

the slider to view data of another month, and one said “I don’t remember the figures, but I can tell that this month is better than the previous one.” When they double clicked on the Process Overview area, Steps in Details was shown. They could easily compare steps and identify the critical one. They found the interactive thresholds very useful. They adjusted the threshold of the most problematic step to decrease the number of bad performances. After exploring the steps, and creating different scenarios on how to improve each step, they started using Actors in Details to investigate actors’ performances. While viewing the lab technician, one of them said “I can see without reading the figures which lab technician is performing faster.” The marker (black dot on the Y-axis) in Actor in Details was helpful, as they easily related it with the threshold, but it took some time to understand the gap between the rectangles. Once, they found out, they could rapidly compare the minimum and maximum durations for the actors. Comparing actors’ performances using the heights of shapes was easier than referring to the percentages. Finally, one of them asked for an additional feature that indicated the reason for a not-completed test (e.g. lost sample, not a good sample, etc.).

- *Results:* The physicians were able to understand the PCD and MSProVis rapidly. They could quickly obtain an overview of the process, locate problem steps in the process, and identify responsible actors. The physicians appreciated that MSProVis presented data in different levels using the PCDs. Enriching the visualization with textual information allowed one of them to quickly conclude: “I compared the relative heights, next I looked at the figures, and after viewing the figures I concluded.” Further, they found the interactive thresholds very useful. Adjusting the threshold of the most prob-

lematic step showed how it would decrease the number of overall late performances. One of them added that in Copenhagen region several hospitals use a tool, known as the Controlled Chart, that monitors events and triggers alarms but it is not meant for retrospective analysis of a large number of events as MSProVis does. Therefore, it would be very useful to integrate MSProVis to these environments.

6.6.2 Studies with Novice Users

We conducted a user study with five graduate students (3 female and 2 male) from the University of Maryland. Participants were good at computer usage in general, but they were not familiar with process analysis and did not have a medical background. They were introduced for the first time to the PCD and MSProVis during this usability test.

- **Procedure and Tasks:** Each usability test consisted of four parts and was conducted in 30 minutes. First, participants were shown one PCD using the Process Overview screen and asked to interpret this PCD without any training. Whenever there was a misinterpretation, the experimenters assisted them before they moved on to the next step. The second part took 10 minutes and consisted of an exploration of MSProVis. After getting an understanding of the PCD and MSProVis, participants used 10 minutes to carry out these tasks in a think-aloud manner:
 1. How many blood tests were in-time, late, and not-completed in April 2011?
 2. What is the minimum, maximum and average duration for performing late blood tests in April 2011?

3. Which step has the highest percentage of in-time, late, and not-completed in April 2011?
4. If you increase the threshold of lateness in step 3, “Examine Specimen and Record Results”, how many tests are late in April 2011?
5. Why do you think the number of late tests changed that way?
6. Which actor in Step 2, “Draw and Send Sample”, has the highest percentage of not-completed tests in April 2011?
7. If you decrease the threshold of lateness, how does it affect the performance of this actor?
8. If you compare March, April, and May, which month has the highest percentage of late and not-completed tests for Step 2, “Draw and Send Sample”?
9. Which actor for each month is responsible for the not-completed tests?

These tasks tested if participants were able to quickly obtain an overview of the process, view details for steps and actors, adjust thresholds interactively, and compare performances for different months. In the last 5 minutes, participants were asked if they had any suggestions for improvements, and to respond to these subjective evaluation questions using a 10-point scale:

1. Give your evaluation of how comprehensible the PCD is.
2. Give your evaluation of how comprehensible MSProVis is.
3. Give your evaluation of the interactive features in MSProVis.

- **Results:**

- *Understanding Data in the PCD:* The evaluation showed that all users were able to read most of the information presented in the PCD. They were readily able to identify the number of in-time, late, and not-

completed tests on the Y-axis and associate them with colors. Information presented on the X-axis was more difficult to understand at a glance. Positioning the triangle next to the orange rectangle made two of the participants assume that not-completed ones have a duration time higher than the maximum time for late completions. One of them suggested using a circle, instead of a triangle, may have been less confusing. Another participant suggested the triangle should have a proportional width and height, and the steps should not be distributed evenly.

Two participants found the average confusing because the text was aligned vertically and the line was not extended until the X-axis. Standard deviation notation was not guessed by two participants, while others related the horizontal bar with the average and recognized it. Positioning the in-time performance percentage below the orange rectangle caused a misunderstanding to one of the users, who associated the percentage with the late completions. The color-encoding did not help in this case. Others found the consistent color encoding helpful and easy to associate with the shapes. Participants accessed tooltips on-demand. Two participants preferred to obtain details on mouse hover instead of mouse click.

- *Exploring Data using MSProVis*: After the first step of introducing participants to the PCD, the next step included exploring MSProVis. Participants started exploring the data presented in the Process Overview by interacting with the slider. They were able to compare the performances between months. Participants were told that by making a double-click on the Process Overview, Steps in Details would open. One

of the participants clicked over an orange rectangle and expected to see only information related to the late completions. All participants found Steps in Details view intuitive and easy to understand. They could recognize the steps in the process and compare the number of tests that were completed and not-completed in a step. However, the different timescales in the steps caused some confusion when participants compared the duration.

Participants appreciated the fact that they could adjust the threshold of lateness. All participants were able to understand what happened when the threshold changed. However, threshold of lateness bars were not very noticeable and selectable as interactive affordances. Participants were able to compare different steps identifying which step caused the lengthening of the process. Finally, the participants explored Actors in Details view by double-clicking in one of the steps. They were able to rapidly gain insight on actors' performances through the PCDs. Participants were able to more easily compare actors' performances than steps' performances, due to the consistent scales used in the X- and Y-axes between actors. Initially, participants found it strange to see the gap between the green and orange rectangle. However, they were able to figure out what it meant on their own. One of the participants pointed out that a marker for the threshold of lateness might have assisted them more in understanding the gap.

- *Executing Tasks*: Participants used the data from April 2011 for the first seven tasks. In task eight and nine, participants were asked to compare performances for March, April and May 2011. All participants

answered all tasks correctly. The first seven tasks were performed easily and quickly. Average execution time for all seven tasks was 5 minutes. The last two tasks were more complex, and took approximately 2 minutes for both. Participants had to switch from one month to another, and memorize the late and not-performed percentages for 3 months. Although the comparison was successful, the results indicated the need for a better approach in comparing processes, actors, and steps' performances at different points in time.

6.6.3 Discussion

The usability test showed that users unfamiliar with the PCD and MSProVis were able to rapidly perform representative tasks, gain an overview of the process, view details-on-demand for the steps and actors, interact and create scenarios to extract insights, and compare performances across different levels. Overall, participants could interpret the PCD and use MSProVis. On average, they rated them 8.5 out of 10 for being comprehensible. Interactive features in the visualization were rated 7 out of 10. Showing tooltips on mouse hover, and having a thicker adjustable threshold, should improve interactivity. Accessing Steps and Actors in Details was easy and intuitive once the participants knew how to interact with MSProVis.

This evaluation showed that the PCD allows users to quickly identify good and bad performances by means of color, shape, and size. Comparing steps' and actors' PCDs was easier than comparing the PCDs in different months, as MSProVis is limited in showing data for a single month. During the study, we observed that users initially were seeking information using the size and color of shapes, and once they identified large colored shapes they referred to the labels to ensure their

assumptions were correct (list of details at the top of each PCD, percentages, etc.).

Several easy improvements were identified and incorporated into MSProVis. For example, the future versions of the PCD changed the percentage label placement. The top position was misleading so labels are instead placed near the center-right of each rectangle. A small circle is introduced on the X-axis of Actor in Details for the threshold defined in the selected step. The direction of the triangle is changed to explicitly distinguish from the duration on the X-axis. Steps in Details do not always have the same width to indicate that the timescales are different.

Three buttons in Actors in Details are added to highlight the person with the highest number of in-time, late, and not-completed cases. Although Steps in Details view is normalized to accommodate all the steps, scalability is still an issue for the Actors in Details view. The button menu in Actors in Details could actually sort by worst or best performers so that the user can see the critical ones at the top.

The major limitation to the usability studies is none of the participants were clinic managers. In the observations (Appendix A), some clinic managers were experienced physicians, who computed on paper percentages of delayed physician reviews. However, not all clinic managers are experienced physicians. Clinic managers should have been observed or interviewed as they work to better understand their tasks and the current state-of-the-art, such as the Controlled Chart.

6.7 Surveys

There were two survey-based evaluations with 21 and 28 participants, respectively. The surveys were distributed on paper and took participants less than 10 minutes to answer. The goal of these surveys was to learn more about user preferences for

the PCD. The first survey compared different design options (bar, pie, and stacked bars with a triangle) for the PCD while the second survey compared the PCD design to two different bar charts (binned and not binned).

6.7.1 Survey 1

Before a presentation of MSProVis to a group of researchers from Scandinavia, the first survey was distributed. The responses were also collected before the presentation. There were 21 people in total. They were researchers, Ph.D. students, nurses, and professors. However, none of them worked in Human-Computer Interaction or Information Visualization.

- **Setup:** The survey instructed participants to assume being a manager in a clinic or hospital, who wanted to see how the results of lab tests are processed. They were given three custom charts (Figure 6.4) that aggregate and visualize the number of lab tests (e.g. blood tests) and their duration.

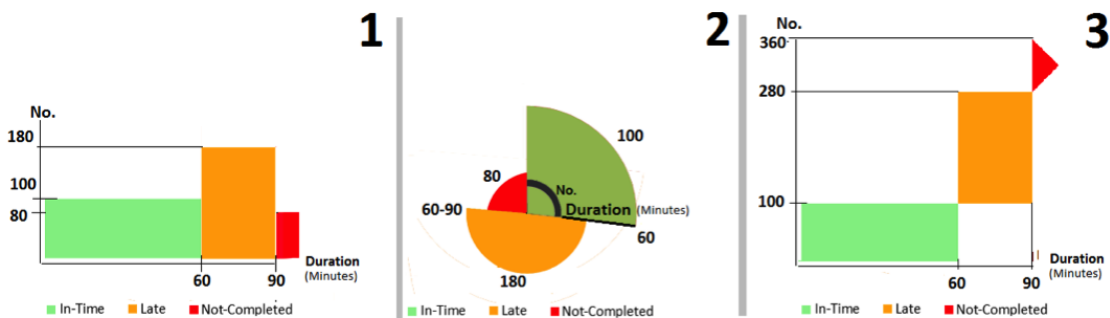


Figure 6.4: Three variations of charts to represent in-time, late, and not-completed cases: (1) Bar chart-style, (2) Pie chart-style, (3) stacked bars with a triangle.

Since not all participants had a medical background, they were told about the categories of expected durations for the lab tests:

1. In-time: lab tests performed within the expected time (e.g. duration ≤ 60 minutes)
2. Late: lab tests performed after the expected time (e.g. duration > 60 minutes)
3. Not-completed: lab tests never completed (e.g. no duration as the samples were lost)

The following questions were directed to obtain subjective preferences and the participants were asked to denote one chart of their choice (i.e. 1, 2, or 3) for each question:

1. Which chart can you understand better?
2. In which chart can you better understand the duration?
3. In which chart can you better understand the number of tests in each case?
4. Which chart provides you more detailed information?
5. In which chart can you better distinguish the not-completed?
6. In which chart can you better distinguish the total number of tests?
7. In which chart can you better distinguish in-time from late?
8. Which chart do you like the most?

- **Results:** One of the survey participants did not reply. This person preferred to give his interpretation. Two of the survey participants skipped some questions. Table 6.1 summarizes the answer counts for the different designs. The numeric results suggest that the pie design was almost never preferred while the bar design was ranked the best except for one case, i.e. the total number. The PCD can compete with the bar design on providing detailed information and distinguishing the not-completed.

Table 6.1: Results of each question are displayed in the total number of participant responses for the different charts.

Question	Number of Replies	1	2	3
1	20	16	0	4
2	20	15	1	4
3	20	15	0	5
4	19	10	2	7
5	19	10	2	7
6	19	4	0	15
7	18	11	2	5
8	20	13	1	6

6.7.2 Survey 2

At the end of a graduate level lecture, the second survey was distributed and the responses were collected. There were 28 people in total. The participants were graduate students in their first two years.

- **Setup:** Participants were given a scenario, “Assume you are a clinic manager and want to investigate the performance of two nurses. The charts in Figures 6.5, 6.6, and 6.7 visualize the same completed medical orders (blood tests) by Amy (left) and Beth (right). These charts allow you to compare Amy and Beth’s performances side by side in terms of the number of completed orders and the actual completion times.”

Since the participants did not have medical background, they were given the following time requirements for in-time, late, and not-completed cases:

1. In-time: medical orders completed within expected time (completion time \leq 60 minutes)
2. Late: medical orders completed after the expected time (completion time $>$ 60 minutes)

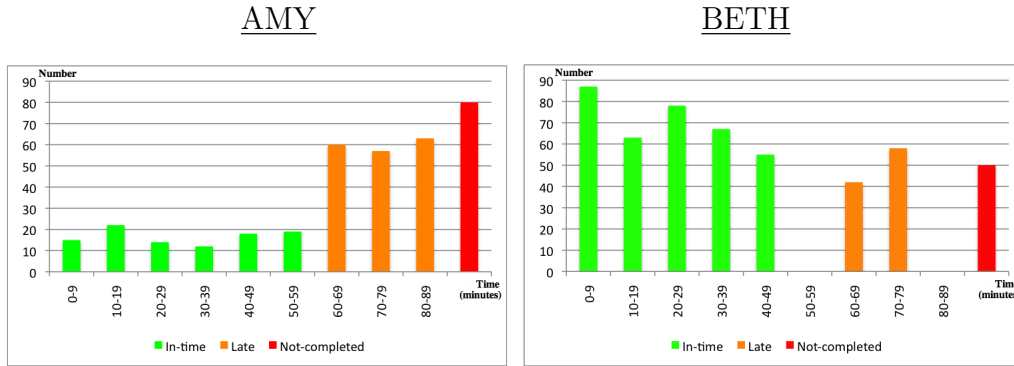


Figure 6.5: Chart 1

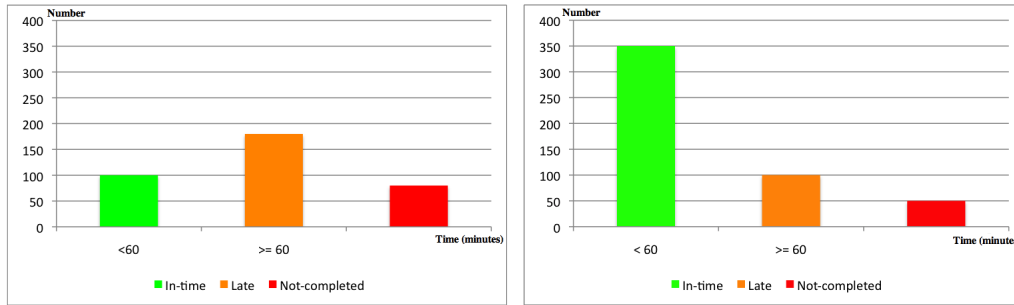


Figure 6.6: Chart 2

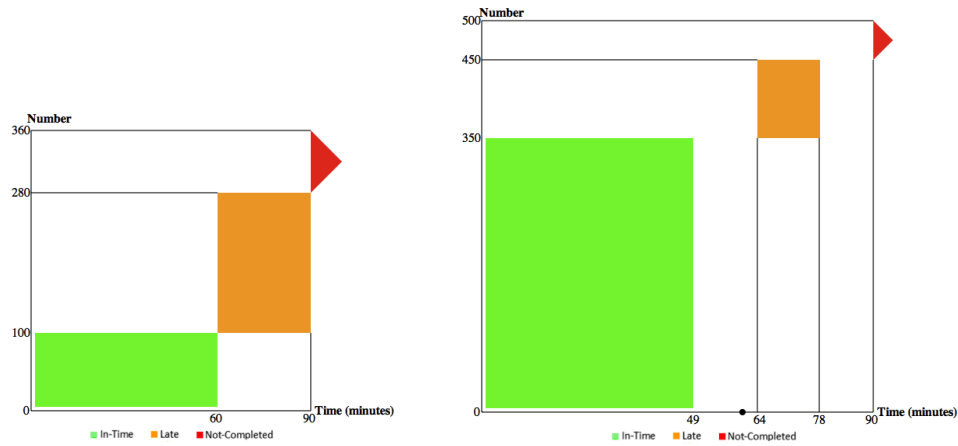


Figure 6.7: Chart 3

3. Not-completed: medical orders never completed (no completion time as the data was lost)

The participants were asked to circle a rating for each of the charts (1, 2, and 3) according to the chart's usefulness in answering the following questions as accurately as possible. The ratings were in the range of 0-9, where 0 corresponded to 'Not Helpful' and 9 was 'Very Helpful'.

1. How many orders did Amy have for each of in-time, late, and not-completed?
 2. How long did it take (i.e. the range of time values) for Amy to complete in-time orders? How about late orders?
 3. Did Amy have not-completed orders?
 4. Did Amy have more orders in total than Beth?
 5. Did Amy complete more in-time orders than Beth? What about late orders? What about not-completed orders?
 6. Did Amy take longer time (i.e. range of time values) to complete in-time orders than Beth? What about late orders?
 7. How much longer did it take (i.e. range of time values) for Amy to complete in-time orders than Beth? How about late orders?
 8. Overall impression.
 9. Comments and suggestions.
- **Results:** Some questions were skipped by some participants. Table 6.2 summarizes the ratings for the designs by averages, one-way repeated measures ANOVA (three treatment levels), and pairwise comparisons using the Holm adjustment method. The participants preferred Chart 2 (the initial design for the retrospective analysis) and 3, in general. However, Chart 2 did very

poorly on duration-oriented questions (2, 6, and 7). All differences are statistically significant, except for the third question about showing not-completed orders.

Table 6.2: Results for each question are computed using the mean ratings, one-way ANOVA (within subjects), and post-hoc paired t-tests.

Q	Chart 1	Chart 2	Chart 3	F value	Pr(>F)	Pairwise
1	3.500	8.148	6.444	F(2, 54) = 52.970	< 0.001	< 0.001 for 1 & 2, 1 & 3, 2 & 3
2	6.679	3.929	6.000	F(2, 54) = 7.346	< 0.01	< 0.01 for 1 & 2, 2 & 3
3	7.821	8.107	8.000	F(2, 54) = 0.312	0.733	
4	3.926	6.296	7.815	F(2, 52) = 21.070	< 0.001	< 0.001 for 1 & 2, 1 & 3; < 0.05 for 2 & 3
5	4.536	8.107	7.571	F(2, 54) = 30.660	< 0.001	< 0.001 for 1 & 2, 1 & 3
6	5.821	3.214	6.357	F(2, 54) = 9.113	< 0.001	< 0.01 for 1 & 2, 2 & 3
7	5.179	2.964	6.250	F(2, 54) = 9.720	< 0.001	< 0.05 for 1 & 2; < 0.001 for 2 & 3
8	5.643	6.143	6.893	F(2, 54) = 4.123	< 0.05	< 0.05 for 1 & 3

After the survey, there was one question by the participants about what the triangle meant. The confusion about the triangle is also apparent in the participant comments for the last question:

- “Show some distribution in the box of Chart 3?”
- “The triangle is redundant. What does it mean?”
- “Chart 3 is a good representation, but might have issues with large sets. I think except for ingrained details, Chart 2 gives a quick information presentation.”

- “If comparing, it would be more useful if both sets of data were displayed on the same axes.”
- “The last one takes more time to understand but is very helpful after knowing the structure.”
- “Confused that the red (not completed) was a triangle. Also didn’t immediately realize that the X-axis of Chart 3 represented the range information”

6.7.3 Discussion

The implication of survey 1 results is that participants preferred the bar design and that pie designs are much harder to interpret. One plausible explanation is the bar design resembles the standard bar-chart, which is familiar to participants. This is also the reason behind the PCDs having simple rectangles rather than more complex shapes (such as circles) that are harder to interpret. However, it seems novice users may have a hard time understanding the triangle and the stacked up design. The PCD can be updated accordingly to have only rectangles that are aligned near the X-axis. The reasons for these design choices were: (i) the triangle can not be associated with a duration as it does not lie parallel to the X-axis, (ii) the total number of completions are important especially when comparing individuals because actors may be rewarded for more completions even if they had late completions.

The first implication of survey 2 is that the triangle did not make a difference because the differences were not statistically significant for question 3, which was related to not-completed orders. However, most participants commenting and asking about the triangle suggests the triangle is indeed not easy to understand

initially. Another implication of survey 2 was that the standard bar chart and the PCD were preferred (question 8). The argument is that they are much simpler than the more detailed binned chart so they are easier to understand and compare (see the results for questions 1, 4, and 5). The next implication is that although the standard chart is preferred, it performed very poorly on duration-oriented questions (2, 6, and 7) due to the lack of time-related data (one participant even wrote, “impossible to tell”). The final implication is the binned bar chart with details performed a little better (not statistically significant) than the PCD in question 2 only. Question 2 was about looking at one chart and participants preferred to see more details.

There was no winning chart and this result is not surprising. The PCD was designed to help answer particular types of questions (especially comparisons) and these surveys provided some representative ones. The standard bar chart was simple and good enough, however, it did not help with duration-related questions. Many experts, who saw the first prototype that employed a standard bar chart, asked questions about durations. How late completions are as important as the number of completions as 1 minute lateness might not be a big deal while 2 days could affect the process significantly. It is important to set the thresholds accordingly. For all other types of questions (comparisons, percentages, totals), participants could not utilize all the details provided by a binned bar chart easily.

There were some limitations in these surveys. The surveys compared only a limited number of charts. Other types of charts might have been preferred by the participants. One alternative is the bar chart that stacks in one bar in-time, late, and not-completed on top of each other. The surveys merely asked preferences of first time users. After using the different designs for a longer period of time, users

may see some benefits they have not noticed before. The results will also be more reliable if there was a longitudinal study that asked preferences periodically for over a period of time. Another limitation is that these are just preference ratings. Had the questions directly asked and timed users on how fast they performed in each question, the results could have been dramatically different.

First recommendation to designers is to use the simplest design for the types of questions users deal with. If time is not important, the simple bar chart works perfectly. However, if time is of concern, the PCD offers an alternative. Another recommendation is to use designs that are more familiar and easy-to-understand. The triangle and the stacked up design were found to be confusing to users who have never seen them before. Final recommendation to designers is to have more details (e.g. the binned bar chart) show on hover of the standard bar chart or the PCD, as one of the participants commented. Because of the rich data, users can gain more insight when they are focused on one chart.

6.8 Summary

This chapter presented a visual approach to expose delays in multi-step processes by retrospective analysis. A novel visualization called the PCD was designed to summarize complex event-log data and present them in a space-efficient and comprehensible way. MSProVis combines several PCDs, to visualize process, steps and actors' performances. MSProVis is a multi-view presentation that supports managers to analyze and compare several scenarios interactively using thresholds of lateness and different points in time.

The approach was evaluated using examples from the medical and software development domains. Users were able to understand most of the data visualized

in the PCD without any prior training, and effectively use MSProVis after 10 minutes of exploring the tool. In addition, two surveys were conducted to suggest new design choices for the PCD. The results show that there is no winning design and instead a combination of approaches have to be considered depending on the tasks.

The PCD and MSProVis are generalizable to other domains, where multi-step processes are present. Lessons learned can be applied to the retrospective analysis of other processes where various responsible parties collaboratively handle different steps, being separated by time and distance. Examples include but are not limited to the paper-review process in academic journals, and business processes such as online shopping, personal financial planning, the return merchandise authorization process in the electronics industry, and sales within a huge corporation.

Chapter 7

Conclusion and Future Directions

Real, important, and complex processes are generally not handled by one person, nor even one organization, and as the work gets passed from one to another, failures cascade. Busy and dynamic environments make it more difficult to track progress and ensure timely completion of processes. Delayed and incomplete processes occur frequently [149], physically harm people [53], and often cause malpractice litigation [104]. There are many issues medical professionals face that can benefit from carefully designed, implemented, and tested systems. In clinical settings, processes are big and tracking the progress of these processes is routine. In addition, clinicians are becoming increasingly equipped with mobile technologies, which increase the amount of interactions and data generated. Improvements can save human lives and reduce the financial and legal consequences of failures.

7.1 Conclusion

This dissertation started with Jane's true story. To avoid the disastrous consequences, the system proposed in this dissertation alerts clinicians when the results have not been returned after a certain amount of time has passed. Even if there is no communication with the outside lab system and no workflow defined, the results get marked late. Clinicians can immediately take actions, such as calling the patients or asking the patients to contact the lab about their results. This

dissertation contributes to ensuring timely management of results, summarized as follows:

1. **Clinician observations and interviews in clinics and hospitals, with discussions of findings that can improve results management:** My observations at seven hospitals and clinics and an extensive literature review led to the discovery that pending orders and delays are important to capture. My approach is to provide a visual display that incorporates taking actions within the same display because clinicians need a display that aids monitoring orders from start to action completion.

For this purpose, I developed MStart (Multi-Step Task Analyzing, Reporting, and Tracking) that generates an interactive display showing the progress of orders in tabular format, with visual enhancements and rich controls that adhere to a set of design guidelines. Orders move between different tables depending on their status. MStart employs a workflow specification that defines multi-step processes on top of a database of actors and organizations. More specifically, an XML (eXtensible Markup Language) [17] workflow model defines a high-level description of order processes. This model captures default normal and maximum durations for each step. These durations allow calculating an expected duration for the entire order so that at order placement, MStart can accurately predict an estimated due date. After the normal and maximum duration, orders are late and not completed, respectively.

More importantly, since rich tables are designed to assist in decision making, clinicians can act on the progress and make decisions based on the results. To reduce response time, each step in the workflow model also lists possible user actions. This way, rich tables lead clinicians to particular task-specific

actions, i.e. common appropriate actions that users generally carry out. Rich tables allow operations such as sorting to prioritize, and color coding and icons to draw users' attention, filter and search to find ones that are of interest, and double-clicks on results open follow-up actions. By accessing a particular order, clinicians see what actions they can take or alert others about. An interaction technique called ARC (Actions for Rapid Completion) is proposed. MStart employs a strategy to automatically generate actions that offer appropriate choices at each step to help users make decisions. MStart embeds these actions in the display to allow rapid response.

- 2. Comparative evaluations of the displays and action techniques for use in time-critical applications where distractions undermine users' ability to track and react to orders:** I continuously engaged with my target users and evaluated my system by different approaches. Rich tables were refined by conducting dozens of iterative design reviews with clinicians, designers, and researchers. This was intended to gather early user feedback to learn what is necessary for a complete system. Refinements were done in terms of the workflow, actions, terminology, functionality, and data presented to build a prototype practical for use in a daily work environment.

Design guidelines in the categories of medical and action-based were compared to the common interfaces in-use today for reviewing results. A controlled experiment with 18 participants under time pressure and distractions indicated that rich tables statistically significantly reduce the time (from 9.5 minutes to 1 minute) to correctly identify late orders compared to the traditional chronologically-ordered serial lists. A predictive model was built. It suggests that date comparisons should be reduced on these interfaces. An-

other study demonstrated that ARCs can speed the performance up by 25% compared to current systems and the number of interactions necessary to execute follow-up actions can be reduced by a third.

3. **Design guidelines for an interactive visual display to overview the dynamics of ongoing orders and to act on the results rapidly:** Based on my observations and evaluations, 90 design guidelines were devised to improve the timely management of orders: (i) results management (8), (ii) table (65), and (iii) ARC (17). The first aims to improve the current medical systems for results management while the last two are more general and allow users to better view tables and rapidly take actions, respectively. These guidelines build on Stephen Few's table design recommendations [44] and Microsoft's Common User Interface design guidance [86] for medical systems.
4. **A novel retrospective analysis visualization and usability studies for this tool:** While MStart helps users track and act on their ongoing processes, there remains the question of retrospectively analyzing all processes completed in the past months. Clinic managers are responsible for continuous improvement of their processes. Due to information overload, data analytics alone are not enough, and instead visualization techniques are needed to leverage humans' innate ability to complete otherwise intractable tasks. With an interactive visualization called MSProVis (Multi-Step Process Visualization), clinic managers can identify common problems. MSProVis uses the same underlying results management workflow description to display in detail due to which step(s) delays occurred and who is/are responsible for the delay. By employing a Process Completion Diagram (PCD), which shows the number and duration of in-time, late, and not-completed orders, MSProVis

gives the decision maker an idea to investigate further and make decisions.

A usability study with two clinicians and five novices showed that participants were able to understand MSProVis and efficiently perform representative tasks. Two subjective preference surveys were distributed to evaluate design choices for the PCDs. The first survey revealed that because pie designs are hard to interpret, bar charts are preferred. A second survey found out that the PCD is comparable to the standard bar chart design but standard bar chart does not give enough information to make informed decisions.

7.2 Future Directions

A goal of this dissertation is to open up new directions for other researchers. Below lists four possible future directions that may expand upon this dissertation.

7.2.1 Workflow Management Systems

While I have designed and built my own process language, there are many workflow management systems available. These systems have advanced features that are useful to results management such as verification capabilities that formally check that the model is complete and sound. Results management can also advance the research in workflow processes. It has unique features that may provide researchers with new frontiers to discover. Examples include the introduction of actions and the generation of advanced user interface interaction capabilities.

Although my examples are tied to the medical domain, the approach is generalizable to other processes where various responsible parties collaboratively handle different steps, being separated by time and distance. I believe that this research

can be applied to a variety of other application domains such as the software development cycle, paper-review in academic journals, and business processes (e.g. online shopping, personal financial planning, the return merchandise authorization in the electronics industry, the internal supplier agreement in IT companies, sales within a huge corporation). These domains may indeed include more complex workflows. My tools can be adapted to specific user needs via observing domain experts interacting with these systems, and iteratively improving the designs after obtaining feedback. Conducting user studies in different domains will solidify the guidelines and improve the systems further.

The lessons from this dissertation were derived from focused work on medical results management, but these lessons can easily be applied to a wide range of applications. Complex workflows are increasingly common, requiring more careful design of the types of monitoring tools studied in this dissertation. The methods, designs, and results from this dissertation provide a basic for future work in many application domains and a conceptual foundation for an emerging theory of results management. Principles such as showing pending results, prioritizing lateness, clarifying responsibility, providing retrospective analyses, and supporting multiple roles could invigorate design research and suggest fresh thinking.

7.2.2 Design Guidelines

My design guidelines are limited to visual displays on desktop computers. Clinicians are becoming increasingly equipped with mobile technologies, which increase the amount of interactions and the data generated. Improvements to these guidelines in terms of touch-based and web interfaces that offer access from any machine can save human lives and reduce the financial and legal consequences of failures.

My guidelines primarily focus on tables and table-based interactions. However, different layouts or interactions (such as animation) may support results management better. Formal comparisons of various techniques for each results management step could prove to be useful for developers to determine which method to employ and when to utilize that method.

While this research focused on result and order lists, there are many other lists used by clinicians on a daily basis (e.g. drug lists). Moreover, my work is on results management but other lists may support different tasks. The next step could be designing, implementing, and testing new interfaces and visualizations as solutions to problems other than tracking processes. There are many complex tasks domain experts carry out frequently that could benefit from similar design guidelines. My guidelines can be extended or changed for different lists and tasks. By working closely with the experts, researchers can investigate the daily tasks of users, brainstorm about the findings, and propose novel interfaces and interactions that meet users' needs. Testing my guidelines for various lists and tasks can show the benefit and generalizability of my research approach.

My research does not address handling complex results, i.e. ARC guidelines are suggested for normal and routine results. The same principles for managing simple cases could be applied to complicated cases. The interactions proposed in this dissertation can be considered as shortcuts with care given to sensitive situations. Rapid actions should eliminate the disadvantages of current interaction techniques and be readily reversible. The feasibility of this approach is left for future researchers to investigate.

7.2.3 Longitudinal Studies

One of the biggest challenges in my research was implementing my prototypes in real systems. While this is essential for technology transfer, it proved to be extremely difficult without influencing the vendors. Vendors develop EHR (Electronic Health Record) systems, which are in use today by hospitals and clinics. These systems have thousands of users and impact the lives of millions of people in the country. My approach to evaluation was to run an experiment and five user studies in a controlled setting rather than testing my research with real users in hopes that the positive results of these studies can spur more interest.

If vendors fully adopted these ideas into their systems, clinicians can start using them in their real environments. This way, researchers can run longitudinal studies or case studies of how the new systems affect clinicians' daily work. Clinicians can provide more direct feedback as they continuously use the system and experience the advantages and disadvantages. Before and after the introduction of these research ideas, data could be collected to compare user performances and care quality. Comparative evaluations over time at multiple hospitals, half using the new system proposed in this dissertation and the control conditions using their baseline systems, can prove the impact on mortality rates. Success and failure stories can bring more researchers together to improve the systems further.

7.2.4 Visualization

MSProVis was tested using a de-identified dataset with senior physicians, who are familiar with clinic processes. In the future, it would be wise to conduct usability studies with clinic managers who regularly perform retrospective analysis of their processes, using real data. It is also important to apply the same retrospective

analyses in other domains.

This approach is limited in dealing only with linear processes, but the principles could be applied to parallel processes as well. Further, Actors in Details can contain a large number of PCDs while the current implementation shows only seven without interacting with the scroll bar.

The feedback from the studies can be used to produce a better version of a retrospective analysis tool. Implementations include aligning the rectangles near the X-axis, making the triangle a rectangle, implementing mouse hover for more information, designing more visible threshold bars, etc. Challenging improvements such as comparing more than one month in a single screen, space limitations for large numbers of steps and actors, and using parallel processes should be addressed by future researchers.

7.3 Summary

This chapter summarizes all results and contributions of this dissertation. Each contribution was supported with evaluation results that demonstrate its benefits. I believe that this dissertation has introduced new ways to look at results management and shows the impact of human-computer interaction and information visualization for saving human lives. This chapter also combined lessons and feedback from my users and colleagues into a list of interesting future directions, which could lead to challenging research projects. There are still many promising opportunities for designing EHR guidelines, waiting for the research communities to explore.

Appendix A

Requirements Analysis

Electronic Health Record (EHR) vendors typically do not make screenshots or videos available and do not publish guidelines. To alleviate this problem, the best way to understand EHR systems is to directly ask clinicians who use them. This appendix will describe after appropriate IRB and HIPAA training, how the requirements were gathered during visits to 7 hospitals and clinics throughout the United States. More specifically, this appendix will report the observations during each visit. Then, a discussion of findings will elaborate on how these observations lead to the prototype application.

A.1 Meeting with a Primary Care Physician

This meeting was arranged with the primary care physician of one of the colleagues in our lab on September 21st, 2010. The system used at the clinic was called MicroMD, which had two components, Practice Manager (PM) and Electronic Medical Record (EMR). At login, the system shows the schedule on an hour-based daily calendar and color is used to indicate different types of work: conference, emergency, and routine (new types could be defined). The EMR tool has three tabs: Appointment Book, Desktop, and Charts. Billing information can be entered from the Practice Manager.

A.1.1 Observations

This system has some general issues. For example, after log on, there is an error message which the physicians do not understand and ignore. Medications screen have the most error messages. Physicians cannot edit only the dose of a drug, they have to go back to the list that contains hundreds of items to just change the dose. User interface does not automatically obtain the previously entered information so physicians have to go through the medication list once more (i.e. same as the last time they entered drugs). When the delivery fails, no reason is shown and physicians have to check the failures one by one and do it from scratch (i.e. no re-send option). The pharmacy list is not sorted by location and the physicians have to vertically scroll all the way across because the dialog box is not big enough. When authorization is needed in the last page of the dialog, physicians have to click each one and cannot authorize everything at once. For drug interactions, a warning is issued but the physicians do not even read the information because they are aware of the warning they see several times a day but do not have a way to suppress it. Physicians can prescribe drugs for a dead person. When the drug is finally added for the patient in the system, that patient is lost in the patient list (i.e. the most recent patient is not highlighted) so physicians have to locate the patient again. The same drug can be added twice and no warnings are displayed.

For the medical results, there are also many problems. While the old paper reports grouped blood test results in a meaningful way, the computer orders results alphabetically which is not helpful to physicians. Results are sometimes not grouped by the patient which should have been the default. Physicians prefer to see results changing overtime (in a line graph) and a single result is not interesting. Tables that present values are not interactive and there is no way to reorder

columns, rows, or even sort them (i.e. no customization). As results come during the day at anytime, physicians make use of a checklist on paper to keep track of medical results. When a physician is away, only critical results are taken care of and the rest wait until the physician comes back. To do this, the surrogate physicians sign in with the attending physicians' accounts to take care of their patients. There is no forwarding function or any integration with hospitals, labs, imaging centers etc. Paid college students scan imaging results, and files can be opened from the Document Manager. If the folder contains documents, it is marked yellow to differentiate from folders that have nothing inside (in blue). The folder listing does not show the number of sub-items and it is impossible to tell if there is something new to look at.

The physicians are afraid the computer introduces more errors in their processes and do not trust it. The physician complains about the number of clicks and the number of items in the pull-downs. Every minute physicians work on the computer, they lose communication with patients. Thus, the physician prefers to type details in progress notes faster than inputting through combo boxes or radio buttons.

A.1.2 Discussion of Findings

There was an important discovery in this meeting. Tables or lists are the main way to manage results. As results can be considered as to-do items, it confirms the research on to-do lists [10]. If table issues are resolved, the problems in results management can be reduced significantly. The following lists other findings and some proposed fixes for tables:

1. *Tables are common.* Not only for medical results but there are lists for medications, patients, pharmacies, data values, etc. An enhanced table design

can have an effect on the management of the overall electronic system.

2. *Tables may contain many items.* Each table contains hundreds of items and the important ones may get missed by the busy clinicians. Therefore, the table design has to minimize the information overload or enhance users' abilities to inspect large amounts of data.
3. *Edits are laborious.* With current systems, making a simple change is an onerous process with redundant steps. The Eight Golden Rules of interface design state that actions should be easily reversible [121]. Thus, edits should happen in-place and require a single step.
4. *Actions are not supported.* An entire user task can be repeated instead of a single action, e.g. a resend. Quick actions for failures or common cases can mitigate the time spent on figuring out and solving issues [56].
5. *Too much scrolling is required.* Because the tables are not organized in a compact way, scrolling is inevitable. New designs should reduce necessary scrolling as much as possible [56].
6. *Dialog boxes are hard to interact with.* Dialogs are either small, change the context, require more interactions, or open multiple other dialogs. That is, they are not designed to yield closure according to The Eight Golden Rules of interface design [121]. A more lightweight solution is needed.
7. *Transitions are not seamless.* When users switch context, they are not guided with enough feedback. To alleviate the drawbacks, highlights, animations, or even sounds can be used.
8. *Table sorting and grouping are not coherent.* Tables are sorted or grouped by features that are not helpful to users and there is no way to re-sort or move items up and down. This also adds to the scrolling issue, described above.

The default setting should sort the most important information vertically down while grouping related items together.

9. *Tables have limited interactivity.* Tables are conventionally a place to store data [44] but interactivity can enrich tables significantly. Interactive tables allow controls to make changes to data (e.g. see more information, take actions) or to presentation (e.g. sort, filter).
10. *Systems require too many interactions.* In these systems, the number of clicks are high, the mouse traversals are long, and context switches happen frequently, which all negatively affect motor performance. When designing these systems, perception, cognition, and motor performance should be taken into account [56].
11. *Surrogate logins are dangerous.* When a surrogate signs in and manipulates data for another person's account, the account holder is unaware of the modifications although the actions are recorded for the account. This raises many privacy concerns in a clinical setting. One way to mitigate these concerns is to support views into someone else's account. Each action can be recorded for that account and everyone can see what was done by whom.
12. *Systems are not fully-integrated.* While the basic information is available in different systems, each require a separate login and they do not communicate in the backend. While this is not a research question, if this was resolved many improvements can be made to current system designs.

A.2 University Health Center Visit

This meeting was arranged with the director of our university's health clinic on September 24th, 2010. The system demonstrated was called ApplicationManager

v10.1 by a vendor named Point-and-Click (PNC) Solutions. This tool was used for students only, faculty/staff/assistants had a separate system, Employee System's Program.

A.2.1 Observations

Physicians have access to patients, visits, documentation, portals (to communicate results and forms with patients), alerts and sub-alerts (created by physicians), immunizations, reminders, and follow-ups from the taskbar. This view has a task summary (type, description, patient, due date, and status all in one place), reminders for appointments, and messages. In the patient record, a bar on the left allows access to patient's history: medical summary (problem list, active medications), reminders (that have categories), appointments and immunizations, health/disease tracker (health maintenance and disease management), visits and notes (of the past), all results, lab/crosstab (shows in a graph but not utilized by providers), lab specimens (can show status as "to be done"), vitals (could be charted by weight or blood pressure), diagnoses, procedures, referrals, orders (number provided on the bar and could be filtered by many checkboxes), flow sheets, compliance forms, surveys (sent to the patient), scanned documents (like radiology reports), outside care, messages (sent to the patient), letters (sent to the patient), encounter note.

Some issues are raised by physicians. Some of the numbers given on the side bar do not match the actual counts. For lab results, there is no way to differentiate abnormal ones. Everything is in one table list that can be sorted/looked up by date and the pending reports can be excluded/included from the list. Entries contain follow-up flags, description, acknowledgment, ordered by, category, and date (most

recent is relevant) but column customization is not supported. Physicians need to get out of patient history to add information to a current visit, which would then get inserted into the patient history. There are custom templates with sections for the visit (problems), medications, allergies, and past medical information. However, partially modifying something is not possible and physicians have to delete and create again. There are mandatory fields that do not allow moving onto the next step without being filled. Problems are usually procedural, e.g. physicians need to talk to the nurses besides entering in the system. Unless the nurse remembers to hit a second button, old stuff remains in lists. The system does not bill correctly. The learning curve of the system is steep and midsize (small and busy) practices need to make time for this. This health center made use of live webinars in 2008 and in general, they learn from each other. Due to HIPAA violation, no VPN access is allowed except for radiology during midnight hours.

Overall, the health center is happy with their system and there are not many alerts to make the system overwhelming. The system only crashed once in 2 years so it is pretty stable. Their user group meets 1-2 times a year to influence the vendor to update. This is organized by the vendor.

A.2.2 Discussion of Findings

The most important finding during this visit is one table is not enough. Tables overloaded with a variety of and too much data increase the chance to miss the truly important information. Placing everything in one table is a bad design choice if the data has different types or even different features. Splitting the data into multiple tables that separate their use offers more advantages. Below discusses some other findings:

1. *Counts are not updated automatically.* If the screen is out of sync with the data, users are possibly going to miss important information. It becomes a serious threat to users working in distracted and data-intense domains as such users highly depend on the clues on the display.
2. *Abnormality is missing.* Some information has severity ratings that are crucial to know so as to take actions accordingly. If this information is hidden from users, they may not immediately act on the important items and delays may cause even more damage [56]. Another drawback maybe busy users completely missing this information because it is not put upfront for their attention.
3. *Columns are not customizable.* Although the designer makes the best effort to show everything in the most comprehensible way, there are special cases where users prefer customizations. The developer should take into account different user preferences and implement the interfaces with flexibility.
4. *Actions require context switches.* Completing an action should never be more than one or two steps away [56]. When users get interrupted, they may forget what they were doing and never complete it.
5. *Incomplete state is not supported.* Between not started and successfully complete, there is an incomplete state. Systems that do not automatically save actions, frustrate users. Also, in cases of system failures, users may assume a complete state for unsaved actions or not even remember which ones were incomplete.
6. *Items are not forwarded through the system.* When an item is completed by one and sent to another person, it should appear in the work list of the forwarded person without having to remind the person. External communi-

cations to pass the items through introduce both delays and errors [108].

7. *The learning curve is steep.* Not all primary care clinics have the resources and motives to learn complicated systems. Systems should support different levels of users and as the user progresses from novice to power, the system functionalities should evolve accordingly [121].
8. *Systems built with user feedback have happy users.* Systems should take into account user comments and be updated based on needs. When users see the changes, they will be more willing to use the system and provide more ideas on how to improve it even further.

A.3 Second & Third University Health Center Visit

At our university's health center, two additional meetings were arranged with the clinic lab and urgent care units on January 20th, 2012 to learn more about the result processing. Afterwards, a follow-up short visit took place on November 5th, 2012 to obtain answers to some of the remaining questions.

A.3.1 Observations

The workflow is defined as follows. While the physicians use Point-N-Click (PNC) system to order labs (except Women's Health that can draw their own specimens), the laboratory uses OrchardHarvest's Lab Information System (LIS). New orders sent by the physician are accessioned in the lab, i.e. drawn, processed (centrifuge, etc.) and sent. CBC, pregnancy, HIV, urine, culture, strep tests are performed in-house. Otherwise, the specimen along with a print out from the orders system are zipped up in a bag, which are stored together until the (one) pick up time in the evening. Specimens are sent to three different labs: State Department Lab,

LabCorp, and ACM Medical Laboratory. These labs can forward results a couple of times to other facilities. If names do not match on the specimen and the order sheet, the laboratory does not run the test. If there is an error in the order, the physician is contacted by the lab tech. If the patient does not show up during the day of the order, physicians are notified (via messages) to call the patient. Physicians determine when to call patients; if it is not urgent, they may wait until the next visit. After a week, the orders are cancelled in the system by a supervisor who has password and the physician will call the patient, if necessary. It is the physician's decision to check this. Supervisors also have privileges to correct orders and mark not performed. Moreover, lab techs inspect the list of sent orders everyday in the mornings and if test-dependent wait time has passed but no results are received yet, they contact the lab. If the lab has never received the specimen, the physician is notified to call the patient so as to ask to come again (to obtain another sample). However, if the lab is slow in processing, they are asked to speed-up for that particular order. Physicians call the lab when they notice a delay and spend most of their efforts in tracking. Laboratories return results through either computer or fax, which are scanned in and entered to the computer. Approved results are released for physicians to review. If results come back abnormal (panic or critical values), physicians are contacted by the lab techs directly. Physicians check abnormal results right away but other results remain in the taskbar, which is cleared at the end of each day. There is a peer-review every semester to inspect if physicians follow-up with abnormal results. The results are computed as numeric values and printed on paper.

In the lab system, orders are kept in a scrollable long list, which is called the pending list and sorted by the lab. Physicians have access to all results and all

orders (including past). Physicians can also enter future orders in the PNC. The number of entries depends on the date/time and during the winter break on a friday afternoon, there were 46 orders in the list. The columns are: Destination Lab, Ordering Physician, Order Date, Order, Collection Date, Accessioned By, To Be Done (for future orders), Patient, Provider, Organize, References, Web Links. Tests can have the following status: Sent, Final, Ordered, Future Order, In-Process (for in-house tests).

A.3.2 Discussion of Findings

The most notable finding is that delays are reported as a loophole. When delays occur, the system does not notify clinicians. Late orders are not recorded electronically and clinicians spend hours to track these orders down. Here are the other findings:

1. *The process is not fully electronic.* Most steps are processed outside the electronic environment. As mentioned in the literature [38], results management heavily involves paper printouts, specimens, mail, and so on. These interactions are not captured in the electronic environment.
2. *Clinicians know and determine the time.* While clinicians make their judgments to respond to results, the systems are unaware of this knowledge. In other words, clinicians have no way of telling the system that they want to review a result at a certain time.
3. *Phone communication is crucial.* There is some electronic communication with result reports and messages. For errors or critical situations, clinical personnel depend on calls as reported by other researchers [54].

4. *The pending list is checked once a day.* Technicians do their primary work during the day and go through the list of pending orders once. The list may contain hundreds of items that can be forgotten in the interim.
5. *Peer-review ensures timely follow-up.* A retrospective analysis of the past performance for physicians helps identify root causes of issues [56]. This idea could indeed be extended to include all steps in results handling to guide clinics in making decisions about the overall process performance with a visualization.

A.4 Emergency Department Visit

A night observation shift was arranged with the emergency department of a local hospital on January 20th, 2012. The system was Azyxxi, which showed the last 48 hours.

A.4.1 Observations

There are three teams during shifts: red, blue, and green. In the red team, there were 16 total entries (fitted in one screen) sorted by room number at the time of the visit. Every row is a patient in a room. Cells show the last processing time for Labs, Meds, Radiology, X-Ray, CT scan, Dic(tation), Log. Tabs in the opened pop-up dialog for labs are all, pending, cancelled, unclassified, special, urine, chem, body fluid, hemo but the physicians always looked at the all tab. There are options to trend the results and values are shown in a list with L/H indicator. When physicians are leaving, they talk to the next physician about the patients and they visit each patient together (a.k.a. handover process). Discharge instructions are printed from paper. Physicians can assign a medical student/resident to see the

patient and after the resident's visit, they debrief.

Some findings are as follows. Physicians have to convert in their head the time passed. The physicians keep a list on an extra paper to note who are seen, written a comment about, or discharged (circled and put the timestamp on paper). Every 30 mins to an hour, physicians check the list of patients on the screen. The physicians also have to do mental calculation of expected delivery based on their experience. Physicians enter comments in the system to remind themselves that they are waiting for something. However, these comments are not connected to anything and the lab does not see them. These comments for the future look like a code (e.g. c, ua, upreg, ivf > 1230h). Visit notes are written on a paper and the cleric puts them into the computer system. Physicians have to hit refresh each time the console is opened and finding out what is pending is 3-4 clicks away. Entire screen background is red and the new info is not apparent, i.e. the refresh button does not draw attention to the newly changed data by highlights or animation. Not only there are no alerts when new results become available, the system rewrites the old timestamp so there is no way to know if there is an earlier result. Physicians have to remember this information but it is not easy for novices or when there are more patients. After 60 mins, physicians ask nurses to start calling the lab for a late test, e.g. during the shift, a physician had to call the X-Ray room to ask why a test was not completed yet.

Some issues were observed during the shift. When the triage physician left or a team of physicians were working together in the emergency department, the new physician was not aware of what had been ordered previously or by other physicians. The physician ordered a test that was not put into the system but when the results came back he realized it was not in the system so he asked the

nurse to put it into the system again.

A.4.2 Discussion of Findings

An important discovery was that clinicians had to do many calculations in their head. In situation awareness theory, researchers indicate that busy and time-pressured users can make mistakes [41]. The system should show the desired value, e.g. time passed, expected delivery.

1. *Last processing time alone is not enough.* Although timing is important, a screen with a bunch of timestamps is nothing but useful, especially if clinicians are going to convert these time values to another value that is more useful to them.
2. *Clinicians make reminders to themselves.* Whether by keeping a paper or writing comments on the screen, clinicians use reminders extensively because the systems obscure critical data for decision-making [107]. A good system would not necessitate extensive use of reminders.
3. *Second-hand data entry introduces errors.* Because the computer notes are not communicated through the system and visit notes are on the paper, clerics type in the data. As observed during the shift, clerics also make mistakes. To reduce such errors to a minimum, physicians should easily be able to enter the information themselves.
4. *The number of clicks is too many.* It is a bad design choice to have users press an extra button to refresh or to put the most useful information a couple of clicks away. This takes the time away for physicians to make decisions or see their patients.

5. *Bad color choices and lack of animation is concerning.* Systems need to alert users with correct colors [56] and animation techniques to draw attention to the newest information. In a dynamic and collaborative environment, such information may be missed by different clinicians.
6. *Systems are not designed for novices.* Complex systems that are not suitable for novices lead users to errors, frustrate and scare users.

A.5 Meeting with an Internist

A primary care clinic in Houston, TX was visited to meet with an internist on January 25th, 2012. This was a busy clinic and the physician reported seeing 20-30 patients daily. The physician was using one of AllScripts EHR systems.

A.5.1 Observations

Results management process was observed. The first thing physicians see when they sign in is the tasklist. From the tasklist, physicians can sign, cosign, and review a patient's note. There are computers in examination rooms and physicians write their notes before the patient leaves since it is required by law to give an explanation to the patient. Sometimes, physicians can tell patients which laboratory to go. Paper orders are scanned in by the staff (nurses) to keep a record. If there is an emergency, the laboratory calls back. If the patient does not show up, the laboratory calls the clinic. The results go directly to the patient charts and the tasklist of physicians. Usually, there are 100 things to-do in physicians' tasklists; 10-15 arrive a day. Physicians check the tasklist one or two times a day and clean the list everyday. Although X-Rays and blood work get populated automatically, paper results are scanned and put in the system by nurses. The electronic system

has a flowsheet for a specific patient that shows values changing over time. If physicians are looking for a result, they ask the staff to go and find results (i.e. call the laboratory).

There are many problems. Since there are too many steps in the order entry and it generates so many printouts, physicians do not use it. If the orders are submitted from the system one-by-one, each one has to be printed and handed to the patient separately. The physician has to remember to write free-text notes to self in progress notes about the patient orders. However, there was one instance when physician forgot she ordered a test that came back and did not write what she ordered in her notes either. Because patients assume they are fine if they do not hear back, physicians ask patients to call if they have not heard in a week. While the system interfaces with some of the outside laboratories, not all of them do. The physicians report that 20% of X-Rays (from one lab) do not come back. In these cases, physicians call the laboratory to obtain a hardcopy which is scanned into their system. Offline results get lost more often than digital ones. Although some laboratories do a bad job with some tests, they are better at others (e.g. the lab that did not return X-Rays was consistently good at returning mammograms). X-Rays do not alert physicians. There were situations in the past when the physician saw the result but forgot to follow-up. Once, the results were sent to an unknown physician due to a system error and there was no one checking this. If physicians write a letter, they also need to assign someone to mail it, otherwise it disappears from their tasklist. Forwarding the work is not automatic, physicians have to communicate with the person they want to handle either verbally or through the system (for example, asking a nurse to print it out). When the physician is on vacation, re-assigning the work does not work.

Other issues were observed during the meeting. While there is an icon in red, “due :-” (hovering over shows “4.13h late”) the physicians do not know what that means and have not noticed it before. The labeling choices are poorly chosen, e.g. there is a button called “QVerify” which does not notify while “Verify” meant the physician will do something about the result (follow-up). Once the order is sent to the printer, the system does not print it and there are no error messages so physicians have to cancel the order and re-order. There are options like “overdue in 10 days” or “overdue if important” for the lab to contact the physician but there is actually no computation performed in the system, it is like an email/communication system. Search is difficult because although the terms are long and confusing, it does not work until the terms are spelled correctly. There are no search suggestions, “did you mean?”, or autocorrect options.

Physicians also commented on other systems they used and provided suggestions. For example, they prefer Epic over AllScripts. Epic has templates but the physicians find it too generic and they do not like the checkbox or radio button style although they generate billing easily. Care4 is an inpatient system and using for outpatient is scary since the physician is never notified but instead the results directly go to the patient chart. If physicians knew there is a delay or when the result will come, that would be awesome. It would be nice to have an option to say “let me know” or “I need to know if this does not happen”. If reminded, the physicians would more likely to enter in the system.

A.5.2 Discussion of Findings

One important discovery here was about the follow-up. Current systems did not seem to have the mechanisms to ensuring that the follow-up was complete [124].

1. *The tasklist is checked 1-2 times and cleaned daily.* As was also mentioned in one of the previous visits, the list is not viewed regularly during the day but it is cleaned at the end of the day. This confirms that the 100 items in the list may be forgotten by users until next check.
2. *Tasks with too many steps are unused.* If the tasks take longer or produce extra outputs on the computer than on paper, users refuse to use the computer. For users to prefer computers, tasks should be at least as efficient if not more efficient [56].
3. *Progress notes and patients serve as reminders of orders.* In systems that are error-prone and do not guarantee timely completion, clinicians have to rely on their notes and patients [150]. With carefully-designed systems, the need for these yet other error-prone approaches can be reduced.
4. *One fifth of orders do not come back.* The clinician confirms that a huge percentage of their orders do not come back with results. Other researchers stated similar percentages of missed results [36]. This presents a serious threat to patient safety.
5. *Clinic systems can be connected to lab systems.* Different systems used at various facilities can communicate to some degree. In this clinic, the results are shared from some labs. Given this fact, more information (e.g. the person or the step) can be forwarded to clinics that could be useful to physicians so that they can make decisions faster and take actions accordingly.
6. *Non-electronic results are lost more often.* It is not unexpected that results, which are not submitted through the computer, are lost frequently [40]. This confirmation means moving the results to an electronic environment can mitigate the bad consequences of lost or missed results.

7. *Not all results alert physicians.* While too many alerts interfere with physicians' daily workflow, alerts still play an important role [56]. Alerts can be implemented in a non-intrusive way that could even be turned off with markers for critical results [90].
8. *Follow-up is not ensured.* Current systems fail to enforce that the follow-up is complete. Either results disappear after (accidentally or intentionally) opening them or get automatically dropped from the list by a system setting after staying there for a long time. Managers should be able to check if physicians followed up on time.
9. *System may not send results to the correct physician.* While making life-critical systems foolproof is of paramount concern, a system administrator should be monitoring error cases to avoid further failures. This means there should be different views of the same data for a variety of users.
10. *Tasks disappear without completion.* Orders fall through the cracks in systems that necessitate communication outside the digital environment [39]. Incomplete tasks should stay in the list until they are confirmed of their completion.
11. *Tasks are not forwarded automatically.* The prior person may forget to inform the next one [56]. Once the assigned person is finished with their task, the next person should automatically be enabled for the completion of the task.
12. *Surrogate assignment is not supported.* Physicians go on vacation or may not be able to come in on one day, someone has to take care of their patients. A surrogate should be able to view and take actions on another physician's patients [54]. This could be best supported with views.

13. *Legends are missing.* Abbreviations, colors, or icons facilitate fast reading, however, there should be a legend somewhere on the same page they are used as a reminder to their intention.
14. *Labels are misleading.* Widgets whose meanings are not obvious introduce errors [56]. Either they are never used or are used accidentally causing more issues. Widgets should have a brief but meaningful label.
15. *Errors are not communicated well.* When there is an error, users should know what went wrong with a lucid message [121]. Rather than requiring multiple steps to fix what went wrong, errors should be fixable in one-step.
16. *Systems do not compute.* Computers are powerful computational machines that can be used to compute deadlines and alert the physicians at the time [56]. However, current systems do not seem to make use of this feature of computers.
17. *Search is difficult.* Because clinical terms are long and come from Latin roots, search suggestions and autocorrect becomes handy [56]. Search functions in systems that do not integrate these features stay unused.
18. *Some systems are more preferred.* By using various systems at different clinics, clinicians have developed preferences over others. Because they compare their systems against others, their feedback is more fruitful.
19. *Clinicians prefer text over widgets.* Due to mouse interaction, widgets take time to select. Free-text is easy to generate either by typing or dictation. To enhance the experience for power users, keyboard shortcuts may help.
20. *In- and out-patient systems are different.* Inpatient systems cannot be used for outpatient without changes. There are unique features to outpatient systems that are not default in inpatient systems.

21. *Clinicians want to know the delays and deadlines.* Clinicians show excitement in knowing delays and having deadlines. This may have some practicality to their workflow.
22. *Clinicians like system reminders.* Rather than setting offline reminders to themselves, which may be less reliable, clinicians are interested in electronic reminders [70]. These options are not difficult to add to current systems.

A.6 Visit to Outpatient Clinic

An outpatient clinic in a hospital was visited in Houston, TX on January 26th, 2012. This clinic used Epic.

A.6.1 Observations

The physician and residents explained their workflow. The physician checks results in the afternoon and handling results takes 5-6 hours. The physician reads the notes before meeting with the patient and writes notes in the evening. Results come to in-basket and in the morning on the day of the visit there were 13 results. New results are shown in bold. Status of a result can be final, in-process, cancelled, or pending future. Orders should be in the notes. Referrals are written by the physician and faxed, printed by the staff. Working with staff is critical. There are 15 residents, 11 computers (no paper) and simultaneously 6 residents are working. Residents are responsible for putting in the orders depending on the note. In the patient examination rooms, there are computers to enter the orders. Most labs are done in house but not all of them. When there are acute labs, the lab calls. There are barcodes on prescription orders.

There were some major pitfalls reported. Clinicians need to make time to

figure out what is missing and if there are any delays (an onerous process). The only way for the physician to find in the system if there is a missing result is to go to the chart and to read the notes or to talk to the patient. Patient orders are not reviewed if patient does not show up for the next visit. Patient record is spread out into different tabs. For example, Procedures tab contains EKG results, Other Orders lists inpatient orders. Media tab includes legal documents for outside orders (so no one checks it). Results are shown by the date they were ordered and unless the user clicks each, it is not easy to understand which ones resulted. Maybe there are ways to customize but the physician does not know how. The system works slowly because the in-basket turned into an “interest list”. More specifically, physicians keep every patient they have seen in their in-basket lists (as there may be interesting events during previous encounters). There were 214 in-basket results (sorted by visit day) that were unread by the resident. The reason is that the resident prefers to review results from the patient chart but the system does not link in-basket to the patient chart view. Thus, residents have to double-document in-basket communication in patient charts. Human communication problems exist, especially when the notes are badly written. If a resident places an order, the result does not come back to the physician. Some results such as Papsmear do not get marked as abnormal so users have to open each result. Once a result is opened, it is marked read. A red dot is used for attention but the system does not give feedback as to what it indicates. There is no column header, no tooltip but a hardcopy paper on the side of the computer shows the meaning. The clinicians like Epic results review but comment that AllScripts has many problems.

A.6.2 Discussion of Findings

Observing the workflow around different staff led to devising use case scenarios for this clinic.

1. *Handling results takes 5-6 hours daily.* As was confirmed in other visits, handling the results takes a large portion of a physician's work day [126]. Any improvements to reduce this time can have impacts on patient care.
2. *Working with staff is critical.* To handle the results, there is a workflow that involves other staff members [38]. Current systems miss to capture this essential point. A better designed system should center around working with staff.
3. *It is onerous to find missing or delayed results.* Because current systems fail to point out what is missing or delayed, physicians repeatedly report that it takes too much time and effort to figure out such results [39]. This information should be upfront.
4. *Review depends on visits.* There is a lag between when the orders are placed and the patient visit. However, the patient visit determines if the orders will be checked for completion or review [110]. This is due to textual documentation of orders in progress notes and physicians not reading the notes on a regular basis. If orders were listed separately from the notes where physicians can regularly check, the lag could be minimized.
5. *Pending versus results are not differentiable.* In the patient record, orders that have and have not resulted are not separated into two different lists. This requires users to click on each to see if there are results available.
6. *Results are sorted by order or visit date.* There is one default sorting criteria which is either order or visit date. This makes it impossible to see which are

late or to group them by patient.

7. *Customization is non-obvious.* Customization of default settings are rarely used, if at all because it is not obvious to clinicians. Easy customizations can be done within the list while other preferences can be put on a separate menu [56].
8. *To-do list becomes an interest list.* Because physicians do not want to miss important things they devise a workaround [56]. They turn a to-do list into an interest list by adding items of interest. However, searching and handling a longer list is more difficult. If the todo list is implemented in a smarter way, interesting results can be easily reachable.
9. *In-basket and patient records are not linked.* Although handling already takes too much time, double documenting will take twice more time. Whatever is done in one list should be reflected on the other.
10. *Notes are crucial in communication.* Because notes are the only way to track orders, they are currently crucial in communicating with other staff although they are hard to understand if written poorly [108]. Standardizing this text form into a better structure like a list might help with communication issues [56].
11. *Results should go to the ordering clinician.* Each patient has a primary care physician who is responsible for their care. Sometimes residents can see patients during their visits and write orders. Residents are not regular employees of the clinic so they change from time to time. Whatever they do are monitored by the primary care physician. Therefore, the results of orders placed by residents should also be forwarded to physicians.

12. *Not all results are marked abnormal.* Abnormal results have to be checked and followed up with patients as soon as possible. Knowing which results are abnormal can expedite this process so this information should be known via the results list [56].
13. *Opened results are marked read.* Results can be opened accidentally or to skim through. However, this does not mean that they are completely read. To assure that this does not happen a confirmation from clinicians that they have read the document can be required.
14. *An icon's legend is on a paper.* While icons can result in faster reading, not all icons are universal. Thus, the legend should be visible on screen from wherever the icon can be seen.

A.7 Medical Center Visit

A local medical center was visited on April 6th, 2012 to meet with two nurse practitioners (NP) and an epidemiologist about their X-Ray results management. This hospital had both in- and out-patients.

A.7.1 Observations

General information about the hospital was explained by one of the nurses. There are 7 clinics in the hospital, each of which have 5 orders so weekly 35-40 studies are managed. Orders can be placed 3 months in advance. For inpatient, TRACKS (Cerner) is used because inpatients do not usually stay longer than a day. Outpatient presents problems because there is paper trail so more human-errors. Due to errors in the past, there was a case when a child lost a leg.

The following is the procedure for ordering/scheduling radiology exams, written

by a Nurse Practitioner (NP):

1. Written requisition completed (one copy given to the patient/one copy kept by NP for filing)
 - Once the copy is handed to the patient, they are asked to go downstairs to schedule with the radiology
 - The nurse also faxes the requisition to the radiology receptionists (there are 4 of them) with a patient phone number
 - Patients are not expected to call because the voicemail in the radiology department is usually full
 - NP gives a copy to another department if authorization is required for a. CT b. MRI c. Bone scans d. Some Ultrasounds (dependent on insurance)
2. NP places requisition into “ordered” file folder in the NP office
 - There is an entry made in the CERNER system as well
 - There are options as “Priority”, “4h”, “STAT”, “Today (with time + order comments)”, “Next Available” but the nurse cannot make it push through
3. Weekly, NP checks to see if “ordered” tests become “scheduled” in Synapse system
 - Radiology uses a completely different system (Synapse) that is accessible and readable to nurses but nurses cannot modify the data in that system
 - Radiology department enters/updates Synapse entries (the orders are transcribed from CERNER)
 - Synapse system does not show ordered studies, it only shows if it has

been scheduled (hence the reason for file folders)

4. If an order is noted as “scheduled” in Synapse, NP does the following:
 - (a) Add requisition to the “scheduled” file folder in the NP office
 - (b) Adds study date/time to shared outlook calendar as an appointment & invites attending as well as “Surgery NP” contact (in shared address/contacts on outlook)
5. If an order is not scheduled, NP evaluates the emergent nature of the order & facilitates scheduling by contacting radiology CORs & patient’s family
 - An emergent order could be placed on Friday, scheduled for Tuesday but non-emergent ones could be a month from now
 - During the call, nurse asks the reason, i.e. why it was “cancelled”
 - If the orders are not-completed after scheduling a couple of times, she sends a note to the family in the mail with delivery/signature confirmation

Procedure for obtaining Radiology Exam Results (once scheduled):

1. Weekly, NP checks the shared outlook calendar and follows up on all scheduled radiology exams
 - There are 5 nurses who check the pile of orders, each spends 2.5 hours per week to sort the orders for the entire week
 - Note that this is the test schedule date, NOT the test completion date thus nurses usually go back a day or two to keep track of completed results
 - The turnaround time for radiology is 24h

2. Response to the shared appointment is sent to all parties, including the attending with a copy of the exam results (radiologist's interpretation)
 - The way to determine if something was done is to pull up films put into the Synapse & CERNER systems
 - CERNER's patient chart only shows orders for the last 72 hours and when the date criteria is changed, it freezes so the nurse uses Synapse anyway
 - Once the results arrive, nurse puts away the "order sheet" and deletes the entry in her calendar because physician will take care of it (no higher authority)
3. If patient was a no-show/cancelled, NP does the following:
 - (a) Sends "cancel" update to all parties on invite
 - (b) Documents no-show/cancellation in CERNER system as "other note" in the patient chart
 - CERNER system is chart-based and comes with no inbox support
 - Synapse and CERNER do not talk to each other
 - (c) Contacts family to attempt to reschedule – documents communication in CERNER note
 - (d) Places requisition back into the "ordered" file folder in NP office to ensure continued follow-up

A.7.2 Discussion of Findings

The most important finding was that the consequences of failures are heartbreaking. It is confirmed by this hospital that when results are not managed properly, patients are harmed [54].

1. *Outpatient is more problematic than inpatient.* It is not surprising that outpatient results are more difficult to handle than inpatient results because it takes longer time, the patients are not under observation, and not all results are handled in the same institution [107]. The problems in outpatient settings are more challenging and have more potential for improvement.
2. *There is paper trail.* Because not all systems support outpatient results management, clinicians come up with other solutions which require paper-based checking mechanisms [89]. This leads to more errors.
3. *Patients cannot contact radiology.* To shift all the responsibility to clinical staff can be dangerous in cases where patients are actively involved in the workflow.
4. *Computer does not enforce faster results.* Electronic requests to push things faster do not necessarily result in faster results. However, showing that the requested time has passed can urge clinicians to check up on expected results more often.
5. *Different systems do not communicate.* Because the workflow includes various institutions, which have systems to capture a part of the process, communication of this data is crucial for the completion of the process.
6. *Users can view but not write data.* Some information is writeable by only those who produce it and access to this information is available via views. Outside users should be able to annotate or mark information when they view the data.
7. *Not all orders are shown.* This system only shows scheduled but not all pending orders. This is the reason for having paper trail which keeps a record of pending orders.

8. *Shared outlook calendar holds the appointments.* Because the system does not display the order progress, a different system holds this information. It complicates the process because besides the EHR tools in use and the paper list, there is an additional system to check the progress.
9. *Lists do not have an indication of completion.* Whether or not an order is processed is possible to see only by opening up the details. This is more inefficient than showing the order's status.
10. *Physician is the highest authority.* Although physicians are humans who can make errors, current systems do not enforce a higher authority who checks if the follow-up was done efficiently. A system for managers, who can retrospectively analyze performance can help identify the best and worst personnel.

Appendix B

Process Model

This appendix introduces my process modeling language via the example of a medical result management workflow. A knowledgeable system administrator can write the workflow specifications in an eXtensible Markup Language (XML) file that is read by the running application, MStart (Multi-Step Task Analyzing, Reporting, and Tracking).

B.1 Processes

In the medical result management workflow [134, 135], every order is a *process*. Every process may have a number of *actors* that indicate the responsible agents (patient, nurse, physician, etc.) doing their assigned work in concert with other actors. A process requires a unique *id* and a *name*. Each actor can sequentially perform one or more *tasks* of the process. Actors have a *role* attribute denoting their role in the system. As certain tasks can be completed by different roles, the role attribute may represent a list. For a task, a duration range is set, and used to compute the expected lifetime of the entire process. If the duration is unspecified, this means that task is optional (i.e. it may be done or skipped for the process completion). A task is defined with a required unique *id* and a *name*. Mandatory tasks also have *start*, *end*, and *unit*. The *start* and *end* attributes take numeric duration values, while *unit* can be any of “mins”, “hours”, “days”,

“weeks”, “months”, or “years”.

Each task can have zero or more *actions* associated with it. Actions are lists of appropriate choices during task execution. An action has a unique *id* (required), a *type*, an *actor*, a *process*, and an *object*. A *type* can be assigned to one of “order”, “consult”, “refer” values; *actor* captures a role; *process* means the id of another process from the same file; *object* can be a workflow artifact used in the system, e.g. a medication. There is a notion of *groups* and *options*. Groups are used to categorize actions that have common properties together. Options on the other hand, enumerate all possible values of an action or a task.

B.1.1 Simple Linear Process

This is a simple scenario where the process does not depend on other processes. The execution is linear, meaning that to proceed to the next task, the previous tasks, if not optional, have to be successfully fulfilled. Figure B.1 shows an example XML specification for part of a process called “Test”. In this simple situation, the physician hands the patient a lab test order and expects a result to arrive in his/her office. The actor “Patient” has a task named “Get Test Done” which has to be completed normally in 6 days and in 24 days at most (see *start*, *end*, and *unit* attributes in Figure B.1). When the results come back to the “Provider” or “Resident”, that person must perform the “Analyze Result” task within 1 day to 4 days.

The durations of the two tasks in Figure B.1 allow the calculation of a normal and a maximum duration for the overall process, “Test”, which are equal to 7 and 28 days, respectively. This information is later used to predict an estimated result time range when the order is placed.

```

<process id="100" name="Test">
  <actor role="Patient">
    <task id="1000" name="Get Test Done" start="6" end="24" unit="d" />
  </actor>

  <actor role="Provider|Resident">
    <task id="1001" name="Analyze Result" start="1" end="4" unit="d" />
  </actor>
</process>

```

Figure B.1: Test linear process consists of two tasks completed by two different actors. First, the patient gets the test done in 6 to 24 days. Then, either a physician or a resident analyzes the results that come back in 1 to 4 days.

B.1.2 Extended Linear Process

An *extended process* inherits from another process, referred to as *super-process*, by using this super-process' specification. Besides being able to modify any task or action of its super-process, an extended process may extend its super-process by (i) adding more tasks or actions in addition to the ones already defined in its super-process, and (ii) breaking any *container* task of its super-process down into multiple sub-tasks. Each sub-task might be performed by different actors than the actor designated in the super-process specification. The execution is still linear, since tasks and sub-tasks still happen one after the other.

The simple linear process model (Figure B.1) can be extended to more concretely defined tests, such as imaging studies or laboratory tests. For instance, Figure B.2 partially shows laboratory test tasks. First of all, the reader may notice from the first line in Figure B.2 that a laboratory test is-a simple linear test (*isa* attribute refers to process id="100" defined in Figure B.1). Then, it elaborates on the patient's task of getting the test done (task id="1000") by splitting this task into four tasks completed by two different actors. Has-a (*hasa*) attributes assign new ids to these tasks so that these tasks in turn can be referenced else-


```

<process id="102" name="Laboratory" isa="100">
  <actor role="Nurse">
    <task id="1000" hasa="1006" name="Draw and Send Sample" start="2"
      end="8" unit="d" />
  </actor>
  <actor role="Laboratory Technician">
    <task id="1000" hasa="1007" name="Process and Examine Sample"
      start="2" end="8" unit="d" />
    <task id="1000" hasa="1008" name="Record Preliminary Results" />
    <task id="1000" hasa="1009" name="Finalize Results" start="2"
      end="8" unit="d" />
  </actor>
</process>

```

Figure B.2: Laboratory test process definition extends from the simple test. It consists of four tasks that are subtasks of the first task of simple test. There are two different actors that complete the tasks in given times. Note that the third task, “record preliminary results”, is optional because there is no time requirements.

where (Figure B.2). First, the nurse draws and sends the sample between 2 and 8 days. Next, the actor switches to a laboratory technician role that could be one individual or a team made of laboratory technicians, and the same person does not have to complete each of the tasks. First, the sample is processed and examined within 2 to 8 days of arrival. After processing the sample, the optional task is to record preliminary results (note that start, end, and unit are undefined for this task in Figure B.2). Finally, the laboratory technician finalizes results within 2 to 8 days. The last task of the provider analyzing results are inherited from Figure B.1 without any changes.

Expected duration for the laboratory test is computed as follows. All task durations in Figure B.2 are summed together to compute 6 and 24 days (equal to patient getting the test done in Figure B.1). Because it inherits from simple linear test (process id=“100”), the last task of provider analyzing the results from the simple linear test automatically appends at the end of laboratory test, which takes from 1 to 4 days. Therefore, the total duration for a laboratory test is expected

to be 7-28 days.

B.1.3 Parallel Process

A parallel process that consists of multiple other processes, named *sub-processes*, could not be captured with the aforementioned models. Unlike the previous two processes explained, sub-processes take place autonomously, so some tasks can be executed synchronously. For example, when evaluating a patient’s sore throat, a clinician may order a rapid strep throat test and a throat culture at the same time to look for harmful bacteria. The rapid strep test takes approximately an hour, but is less accurate than the throat culture, which may take three or more days for final results (Figure B.3). This structure allows different actors in parallel to carry out tests independently of each other, but initiated from the same previous action.

```
<process id="114" name="Strep Throat">
  <subprocess id="40000" name="Rapid Strep" isa="113" />
  <subprocess id="50000" name="Throat Culture" isa="103" />
</process>
```

Figure B.3: A strep throat test is a parallel process that consists of two sub-processes, rapid strep and throat culture, executed simultaneously.

The XML model shown in Figure B.3 also captures this. Sub-processes “Rapid Strep”, which is-an (*isa*) office test (defined elsewhere in the XML file but shown here with process id=“113”) and “Throat Culture”, which is-a (*isa*) laboratory test (id= “103”), constitute a strep throat test. For such parallel tests, expected duration calculations are completed for each test separately.

B.2 Actor Actions

This process model helps physicians review and take follow-up actions on results [135]. During result review, a physician could be guided to follow-up actions. This is accomplished by listing custom actions for tasks in the model.

More extensive model allows specifying feasible actions for tasks of every actor, not just of the last actor of the process [134]. If a necessary action is not given, a “write comment” action exists by default for each task. Since processes can extend other processes, it is also possible for tasks to inherit actions from super-processes, change some of those actions as appropriate, or add more process-specific actions. This is explained in the subsequent subsections.

B.2.1 Actions during Process Result Review

At the end of a process while reviewing the process results, the actor can be guided to particular process-specific actions, i.e. common appropriate actions that actor would generally carry out. For instance, Figure B.4 partly illustrates leading the physician through the analysis of test results, building from the simple linear process example (Figure B.1). The first line shows that providers can “Access” the “Report” *object*, which is the report of the test result that comes back to the office for care providers to review.

To improve interface usability, support for meaningful grouping of actions, options of actions/tasks, and default values of options were added. In Figure B.4, *group* specifies a set of actions that belong to the same category. The first group in this example is to “Ask Assistant” that encapsulates the following *actions*: (i) “Inform” the “Patient” actor, (ii) “Schedule” a “Visit” object within the following *options*: 1-30 days, 1-52 weeks, 1-30 months, or 1-5 years, with the default value

```

<actor role="Provider/Resident">
  <task id="1001" name="Analyze Result" start="1" end="4" unit="d">
    <action id="10" name="Access" object="Report" />
    <action id="11" name="No Follow-up">
      <option name="Baseline Value" />
      <option name="No longer under our care" />
      <option name="Opted out not to treat" />
      <option name="Other" />
    </action>
    <group type="Ask Nurse to">
      <action id="12" name="Inform" actor="Patient" />
      <action id="13" name="Schedule" object="Visit">
        <option min="1" max="30" name="d" />
        <option min="1" max="52" name="w" />
        <option value="1" min="1" max="30" name="mo" />
        <option min="1" max="5" name="y" />
      </action>
    </group>
    <group type="Order">
      <action id="14" name="Repeat" process="100">
        <option min="1" max="30" name="d" />
        <option min="1" max="52" name="w" />
        <option value="1" min="1" max="30" name="mo" />
        <option min="1" max="5" name="y" />
      </action>
    </group>
    <option name="Review Later" />
    <option name="Complete" />
  </task>
</actor>

```

Figure B.4: A physician or resident are assigned some common follow-up action choices during review. These include accessing the result, recording no follow-up is necessary, asking the nurse to inform the patient or schedule a new visit, or ordering a repeat test.

being “1 month” (note the *value* attribute). Other actions presented to a care provider in Figure B.4 are grouped under “Order”, which encloses the action “Repeat” current test (shown in the figure with its process id=“100”). Tasks can also take options. “Analyze Result” task can be completed with either “Review Later” or “Complete” options.

Similar to tasks that are extended from super-processes, actions can be extended from parent tasks. This is accomplished by adjusting the actions of a super-process to the current process. For instance, Figure B.5 indicates how an

```
<actor role="Provider/Resident">
  <task id="1001">
    <group type="Order">
      <action id="27" name="Consult" actor="Orthopedic Surgeon" />
      <action id="28" name="Consult" actor="Physical Therapy" />
    </group>
  </task>
</actor>
```

Figure B.5: Follow-up actions can be extended from super-processes to modify or add process-specific actions. Here, order follow-up action of a physician or resident during an imaging study is extended to include consultations to an orthopedic surgeon and physical therapy.

imaging study adds two new actions to the last task of simple linear test (since an imaging study inherits from a simple linear test). A possible common action for the provider analyzing imaging study results is to order a consultation to an orthopedic surgeon, or to physical therapy. This action does not apply to other types of tests, so it is part of neither the simple linear nor the laboratory test process.

B.2.2 Other Actor Actions

The model definition tolerates actions for not only the actor reviewing results as a final task of the process, but also different types of actors involved in the process.

An example is given in Figure B.6 for the laboratory technician recording preliminary results, i.e. the third task of a laboratory test (Figure B.2). While doing this, possible actions could be accessing the exam, indicating no results could be entered due to some reason (e.g. error), repeating the test, or entering the result. One could likewise enumerate other clinical personnel's frequent actions for other tasks.

```

<task id="1000" hasa="1008" name="Record Preliminary Results">
  <action id="29" name="Access" object="Examination" />
  <action id="30" name="No Results">
    <option name="No processing" />
    <option name="Processing error" />
    <option name="Other" />
  </action>
  <action id="31" name="Repeat" process="102" />
  <action id="32" name="Enter" object="Result" />
</task>

```

Figure B.6: The most common actions for a laboratory technician recording preliminary results are accessing the exam, marking no results can be entered at the time, repeating the lab test, or entering in the results.

B.3 Summary

This appendix described the process management and action specifications in MStart. The processes file used throughout this document consists of 270 lines for nine concrete processes. Once the generic processes (such as test, imaging, laboratory, pathology, consultation) are defined, the rest of the processes usually entail one line, with possibly a couple lines of modifications (see Figure B.7).

```

<process id="109" name="TSH" isa="102" />

```

Figure B.7: A TSH test inherits directly from a laboratory process without any changes.

For this project, an existing workflow engine [22, 143] could have been used. However, the main purpose of the project was to create high-quality interfaces from the process definitions. None of the commercial or research tools provided such functionality. Moreover, supporting actions would have been really difficult because it necessitated changing the specification language itself. Due to these reasons, it was decided to create a custom process language.

Appendix C

Implementation

This appendix describes my implementation of the techniques and algorithms for modeling workflow processes, generating actions for rapid completion (ARC), and predicting process times in MStart (Multi-Step Task Analyzing, Reporting, and Tracking). MStart is an inspirational prototype developed as a Java application, which consists of approximately 70 classes (each between 100 and 1000 lines) to illustrate my ideas to domain experts. There are six packages: process hierarchy, interface generator, database objects, date and time, reader/writer, and interface models. Each of these will be discussed briefly below, except the interface models that contain renderers, listeners, editors, and so on.

C.1 Workflow Element Relationships

MStart extracts the relationships between workflow processes, actors, tasks, actions, groups, and options shown in Figure C.1 class hierarchy. This hierarchy is independent of the domain of the specification and is merely based on the eXtensible Markup Language (XML) specification elements (defined in Appendix B).

As seen in Figure C.1, `WorkflowElement` is a superclass of all the other classes and it stores an `id` and `name`, along with a `description`. `WorkflowProcess` descends from this class and has a `getExpectedDuration` method that returns the expected duration of a process. This class also saves its super-process as well as its sub-

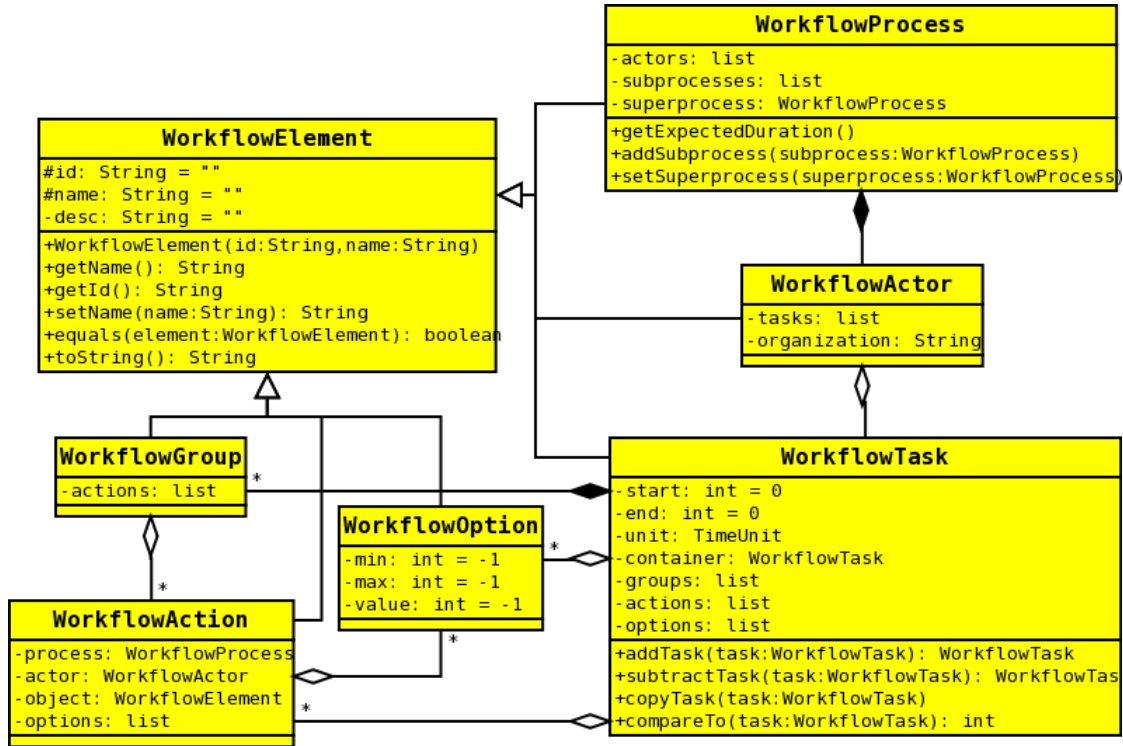


Figure C.1: **WorkflowElement** represents any workflow element and is the parent of all classes in this diagram. **WorkflowProcess** class keeps information regarding workflow processes. **WorkflowTask** stores task-specific operations. **WorkflowActor** depicts actors who can complete tasks. **WorkflowAction**, **WorkflowOption**, and **WorkflowGroup** are used to generate ARCs.

processes. **WorkflowTask** can contain another **WorkflowTask**, and addition and subtraction operations are available for this class. Although the associations are not shown in this diagram, **WorkflowAction** has three attributes referring to a process, an actor, and an object.

C.2 Process Hierarchy

Given the relationships between workflow elements and the XML files containing actual process data, MStart can now instantiate each process and construct a

hierarchy of processes. Figure C.2 depicts the class diagram of laboratory test processes described in Appendix B.

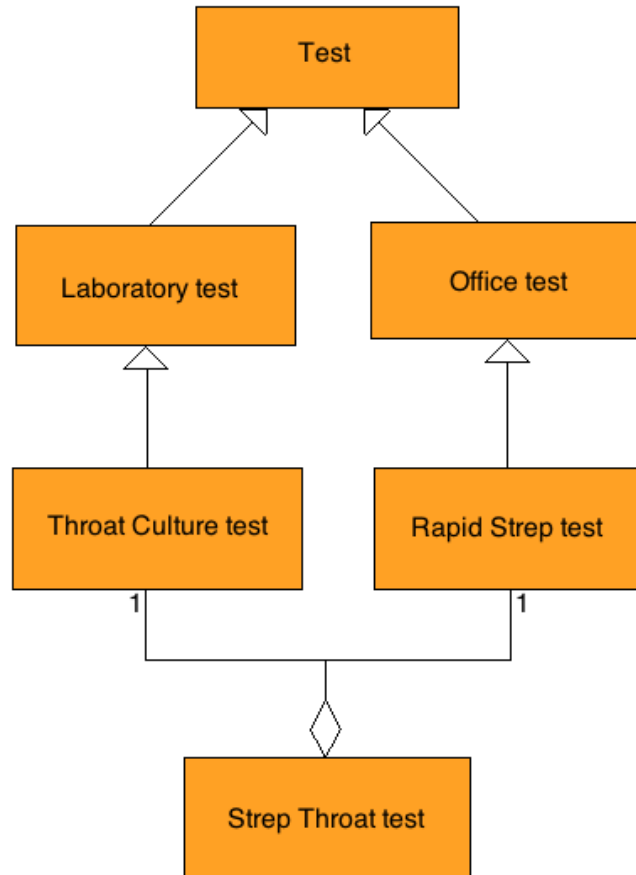


Figure C.2: Test object has two children, laboratory and office test. While a throat culture test descends from laboratory test, a rapid strep test is an office test. Throat culture and rapid strep tests are together ordered as one test, called strep throat.

After the workflow model file is read, the algorithm resolves “Generalization” (e.g. laboratory test inherits from simple linear test) and “Aggregation” (e.g. strep throat test contains a culture test) relationships shown in Figure C.1 for test processes. This information is then used to determine the actual steps in each test. For example, although a laboratory test model definition does not own a “provider

analyzes results” task (see Figure B.2), because a laboratory test inherits from simple linear test and a simple linear test possesses this task after the “patient gets the test done” task (Figure B.2), the algorithm determines that the “provider analyzes results” task should be the last task of a laboratory test.

C.3 Predicting Process Times

Appendix B explained how expected duration is calculated for all processes in a way that captures default durations. However, it overlooks the fact that there are weekends, holidays, etc. It does not take into account that some hours of the day, some weekdays, or some months might be busier than others. It also does not take into account past experiences. For this reason, a module is provided to take such special conditions into consideration.

Figure C.3 shows a simple XML file that lets a system administrator customize MStart for such occurrences. For example, the second argument value of function DATE coincides with the 3rd Monday of January, which is a Federal holiday in the United States known as the Martin Luther King, Jr. Day. Here, the system administrator may list possible cases that could affect duration computations with the `function` tag. The `argument` of a function is an independent variable that takes specific inputs or argument values. `Factor` implies how much that argument influences the time computation. In particular, given a date range, for each day, MStart can check against every function in the file to identify the relevant factors. MStart multiplies the factors altogether and the product is the speed of processing (in other words, how many units of work get done) so the higher it is, the faster things get done and vice versa. If no attribute matches for a given function, the factor is assumed to be 1 (default value, i.e. no effect). For instance, Friday,

```

<functions>
  <!-- federal holidays -->
  <function name="DATE">
    <argument factor="0">
      <value name="New Year's Day" month="1" date="1" />
      <value name="Martin Luther King, Jr. Day" rank="3" day="Monday"
        month="1" />
      <value name="Presidents Day" rank="3" day="Monday" month="2" />
      <value name="Memorial Day" rank="last" day="Monday" month="5" />
      <value name="Independence Day" month="7" date="4" />
      <value name="Labor Day" rank="1" day="Monday" month="9" />
      <value name="Columbus Day" rank="2" day="Monday" month="10" />
      <value name="Veterans Day" month="11" date="11" />
      <value name="Thanksgiving Day" rank="4" day="Thursday" month="11" />
      <value name="Christmas Day" month="12" date="25" />
    </argument>
  </function>
  <!-- some weekdays are slower/faster -->
  <function name="DAY_OF_WEEK" weight="1">
    <argument value="Saturday, Sunday" factor="0" />
    <argument value="Monday, Tuesday" factor="0.75" />
    <argument value="Wednesday, Thursday" factor="1" />
    <argument value="Friday" factor="1.5" />
  </function>
</functions>

```

Figure C.3: Each function lists some argument value(s) it can take. The weight and factor determine how much the argument counts towards duration estimations. Default value is 1.

January 1st, 2010 (New Year's Day) results in no work ($0 \times 1.5 = 0$). On the other hand, Monday, April 25, 2011 will end with $1 \times 0.75 = 0.75$ amount of work because certain tests may be difficult to obtain after the weekend if there is a high demand. To be able to finish 1 unit of work, a third of work time during Tuesday, April 26, 2011 is desired since 0.75 total amount of work could be fulfilled on Tuesday, April 26, 2011.

C.4 Data Objects

MStart also reads two XML files that enclose all actors and organizations seen in Figure C.4. For a laboratory test process, **organization** corresponds to a clinical facility and **actors** are medical workers. Critical to the simulation are the following facts: (i) organizations support some types of processes, and (ii) a test instance could be supported by multiple organizations. Actors supervise other actors so supervisors can be notified in case of a delay or error.

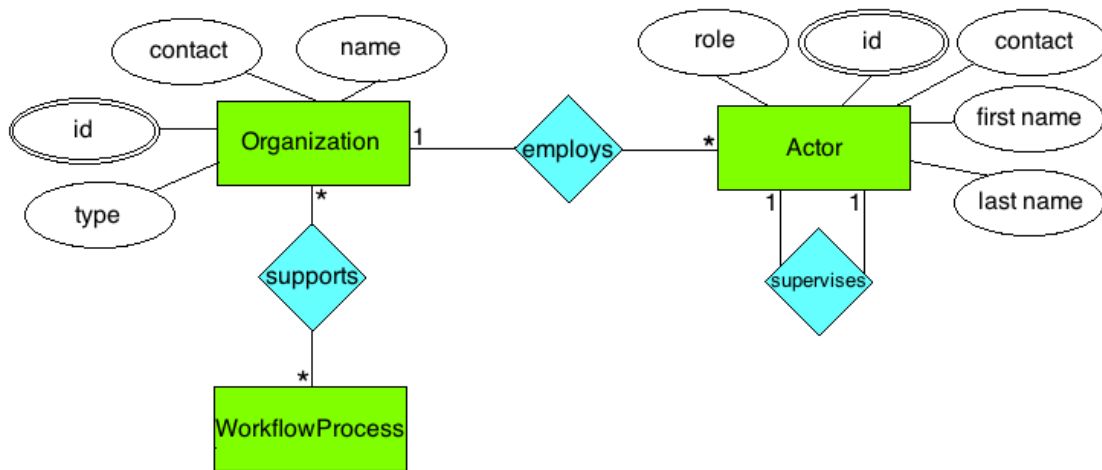


Figure C.4: A database of actors and organizations are input to the running application. This entity-relationship diagram depicts that an organization employs actors, organizations support multiple workflow processes, and an actor is supervised by another actor.

Besides actor and organization objects, MStart needs to retain events that happen at execution time. An **event** references an **order** and a **process** (Figure C.5). Every event depicts an order's current snapshot. At creation time of an event, a time for the test result is predicted using the configuration file settings and the computations described in Section C.3. Furthermore, **Order** class contains all the information about the lifecycle and status of a process (Figure C.5). A **log** is

created for every task completed during the actual processing of an `order`.

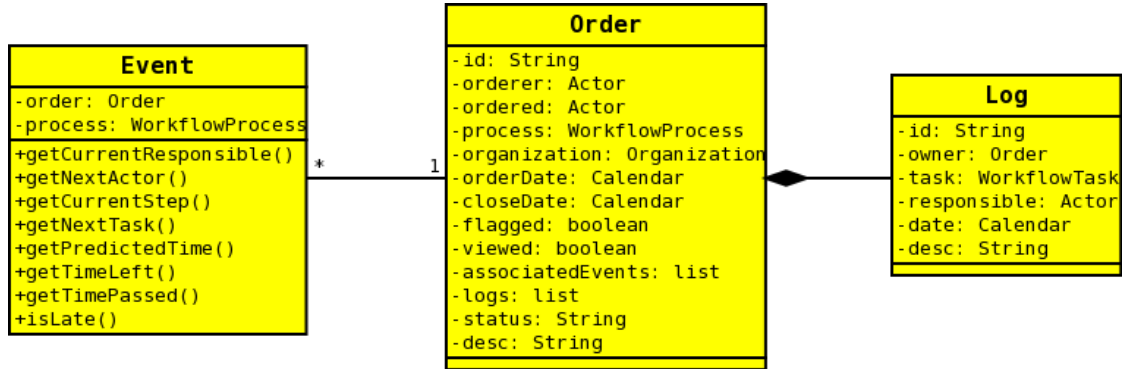


Figure C.5: During runtime, an order contains a list of logs and generates events.

C.5 Interface Generator

Interface Generator is responsible for generating the user interface of MStart. It creates three screens: (i) order, (ii) track, and (iii) complete.

Initially, actors are required to login and depending on the type of the actor, MStart enables/disables certain screens. Figure C.6 demonstrates how a logged in clinic assistant would see MStart interface. There is no ordering or complete screen. Assistants may only access pending orders and they can not follow-up on results.

The Interface Generator visualizes events described in the previous subsection. For instance, Figure C.7 depicts two events as seen in MStart. A blood test starts as one order, and then branches into BMP and CBC events. Each event may be handled independently by different clinical staff, and finally, comes back to the ordering physician's office so that he or she can make decisions based either on individual reports or the overall result.

MSTART File

MSTART Simulator Assistant Jennifer Young checks if all the recent orders for patients coming today are back or not to track them down

Current Time: Sunday, March 13, 2011 11:26 AM Pause Continue

Riverside Clinic Assistant: Young, Jennifer Patient: Phillips, Sarah; Reed, Megan; Wood, Courtney Sign out

Track

Results to Review (9) All New Late Not Completed

Patient	Test	Ordered By	Order Date	Review By	Abnormality
Reed, Megan	Mammogram / P	Brown, Joe (P)	Mar 3, 2011 9:35 AM	Mar 10, 2011 9:35 AM	
Phillips, Sarah	TSH / F	Brown, Joe (P)	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Phillips, Sarah	Pap Smear / F	Brown, Joe (P)	Mar 8, 2011 5:39 AM	Mar 15, 2011 6:39 AM	⚠
Phillips, Sarah	MRI / F	Green, Bob (Res)	Mar 6, 2011 9:31 PM	Mar 13, 2011 10:31 PM	⚠
Wood, Courtney	MRI / F	Green, Bob (Res)	Mar 6, 2011 3:50 PM	Mar 13, 2011 4:50 PM	⚠
Wood, Courtney	X-Ray / F	Green, Bob (Res)	Mar 7, 2011 4:30 PM	Mar 14, 2011 5:30 PM	⚠
Reed, Megan	Pap Smear / F	Green, Bob (Res)	Mar 6, 2011 9:46 PM	Mar 13, 2011 10:46 PM	
Reed, Megan	Dietetics / P	Brown, Joe (P)	Mar 7, 2011 8:43 AM	Mar 14, 2011 9:43 AM	
Wood, Courtney	Blood (CBC) / P	Green, Bob (Res)	Mar 7, 2011 7:45 AM	Mar 14, 2011 8:45 AM	

Pending Test Results (4) All Late Not Completed

Patient	Test	Ordered By	Order Date	Result Due	Status
Phillips, Sarah	X-Ray	Green, Bob (Res)	3/7/11	3/13/11	Exam scheduled
Reed, Megan	Mammogram	Brown, Joe (P)	3/3/11	3/9/11	Image(s) interpreted and reported
Reed, Megan	Dietetics	Brown, Joe (P)	3/7/11	3/13/11	Preliminary Report prepared
Wood, Courtney	Blood (CBC)	Green, Bob (Res)	3/7/11	3/13/11	Preliminary Results recorded

Planned Tests (1)

Figure C.6: Tracking screen of the user interface of MStart as seen by an assistant, Jennifer Young, in Riverside Clinic. There is no order or complete screen. Tables show arrived, pending, and planned orders for patients, Sarah Phillips, Megan Reed, and Courtney Wood, who have visits today. Results that have returned to the clinic are listed at the top in “Results to Review” table, which is not interactive for assistants that can not follow-up. Orders that are in-progress and not returned yet are shown under “Pending Test Results” where assistants can see additional details with a click. Orders that are placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed). All tables are sorted by default. Color-coding indicates lateness and warning sign shows abnormal results.

Pending Test Results (6)					
Patient ▲	Test	Order Date ▲	Result Due ▲	Status	
Peterson, Audrey	Blood (BMP)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered	
Peterson, Audrey	Blood (CBC)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered	

Figure C.7: Audrey Peterson’s blood test is entered as one order but has two separate tests called BMP and CBC that can individually be processed. Therefore, tracking of these two events are separately shown.

The user is guided with ARCs to interact with these events. The interactive content of an ARC is generated directly from the model definition, i.e. depending on what actions are specified in the XML file, ARCs are populated with widgets like checkboxes, drop-down boxes, buttons, and/or lists.

The screenshot shows an ARC interface with the following elements:

- Two checkboxes: No Results and Enter Result.
- An 'Order' section containing:
 - Repeat Test
 - A numeric input field with the value '1' and up/down arrows.
 - Radio buttons for frequency: days, weeks, months (selected), and years.
- A 'Comments:' label above a large text input area.
- Three buttons at the bottom: a question mark button, 'Complete Later', and 'Complete'.

Figure C.8: Laboratory technicians can record preliminary results using ARCs, which are automatically generated from a process specification and list common possible actions.

On the other hand, Figure C.8 exemplifies possible actions of a laboratory technician recording preliminary results at the lab. The exam is shown in a popup alongside the actions. XML definition in Figure B.6 is used directly to generate this screen. “Complete Later” button keeps the data in his/her track list. “Complete” sends the results to some clinical staff who will take care of the next task(s) in the management of this lab test.

C.6 MStart Simulator

To demonstrate MStart in action, it can simulate the placement of orders and actor actions. The simulator advances time rapidly to show how the system would work in a real setting as orders become active. To allow for a deterministic execution and save time on inputting each order through the order screen, it makes use of two classes `OrderWriter` and `OrderReader`. `OrderWriter` class accepts an input file, a list of processes, actors, organizations, an initial time, and the desired number of orders as input. When executed, `OrderWriter` randomly generates orders and logs, and writes them to the given file (see an example in Figure C.9). When MStart is run, `OrderReader` class reads these orders. Logs are taken into account as the time comes, i.e. when the simulator advances time.

```
<orders>
  <order date="1299132731662" desc="" id="1000" orderer="10" process="106" responsible="49" status="in-progress">
    <log date="1299141046898" desc="" id="1061002" responsible="49" subprocess="106" task="1002"/>
    <log date="1299141191226" desc="" id="1061003" responsible="21" subprocess="106" task="1003"/>
    <log date="1299156881669" desc="" id="1061004" responsible="25" subprocess="106" task="1004"/>
    <log date="1299295722071" desc="" id="1061005" responsible="25" subprocess="106" task="1005"/>
    <log date="1299158230719" desc="" id="1061001" responsible="10" subprocess="106" task="1001"/>
  </order>
  <order date="1299164968810" desc="" id="1001" orderer="10" process="107" responsible="52" status="in-progress">
    <log date="1299166099505" desc="" id="1071002" responsible="52" subprocess="107" task="1002"/>
    <log date="1299168484351" desc="" id="1071003" responsible="21" subprocess="107" task="1003"/>
    <log date="1299209826505" desc="" id="1071004" responsible="24" subprocess="107" task="1004"/>
    <log date="1299226515098" desc="" id="1071005" responsible="24" subprocess="107" task="1005"/>
    <log date="1299210000000" desc="" id="1071001" responsible="10" subprocess="107" task="1001"/>
  </order>
  <order date="1299597753390" desc="" id="1002" orderer="10" process="105" responsible="61" status="in-progress">
    <log date="1299603940967" desc="" id="1051002" responsible="61" subprocess="105" task="1002"/>
    <log date="1299604126079" desc="" id="1051003" responsible="22" subprocess="105" task="1003"/>
    <log date="1299607582104" desc="abn" id="1051004" responsible="25" subprocess="105" task="1004"/>
    <log date="1299637582104" desc="abn" id="1051005" responsible="25" subprocess="105" task="1005"/>
    <log date="1299616279105" desc="abn" id="1051001" responsible="10" subprocess="105" task="1001"/>
  </order>
</orders>
```

Figure C.9: `OrderWriter` class randomly generates a list of orders and logs to be read by `OrderReader` class, which is used by the running MStart application. Date is represented with a long integer. Ids are assigned by `OrderWriter`. The id of the order is combined with the id of the task to create log ids. The numbers for responsible, subprocess, and task correspond to ids in actors and processes files. Desc field includes information about abnormality.

C.7 Summary

This appendix described the infrastructure behind MStart. While the example input files contain medical data, this software architecture produces a domain-independent system that can be widely used and easily modified to generate MStart applications for business, academic, or other processes. MStart converts the workflow model into a hierarchy of process definitions, which provide input for an Interface Generator when combined with a database of actors and organizations.

Appendix D

Table Design Guidelines

Tables have conventionally been used in various domains as a place to store large amounts of data as a reference for users to look up and compare values [44]. Table designs have been limited to support these two tasks. Apart from this, users manage their daily workflow through various tables [10]. These are interactive tables that allow users to perform operations on the items. This section will name a look-up table, which is extended with functionality, a rich table (see Figure D.1). Rich tabular displays generally consist of multiple rich tables that are related, as in Figure D.1.

Rich tables have rows, arranged vertically, which display items of the same type. Rows can be sortable by some criteria. Each item may have multiple attributes or fields that are shown in a column, arranged horizontally in a table. Rows and columns may be filtered to show desired items. When the table size does not accommodate table's all rows or columns, a scrollbar enables users to see the hidden parts of the table. Tables, rows, and columns may all have headers with descriptive titles. A column within a row is called a cell, which holds a value. Rows, columns, headers, and cells may be selectable, single- or double-clickable, or editable. Rows, columns, headers, and cells may reveal an explanatory tooltip on mouse hover.

Given a workflow of items, a rich tabular display is generated automatically with the following principles to assist users in finding the most critical information

The screenshot shows the MSTART application interface. At the top, there's a header with 'MSTART' and 'Physician Joe Brown reviews results from his inbox several times a day'. Below that, there's a 'Riverside Clinic' section with 'Provider: Brown, Joe'. The main content area is divided into three sections: 'Results to Review (17)', 'Pending Test Results (6)', and 'Planned Tests (10)'. The 'Results to Review' table has columns for Patient, Test, Order Date, Review By, and Abnormality. The 'Pending Test Results' table has columns for Patient, Test, Order Date, Result Due, and Status. Annotations include 'Show Pending Results' pointing to the 'Pending Test Results' table and 'Prioritize by Late and Lost Status' pointing to the 'Results to Review' table.

Patient	Test	Order Date	Review By	Abnormality	
Evans, Anna	MRI / F	Mar 3, 2011	1:13 AM	Mar 10, 2011	1:13 AM
Parker, Amy	MRI / P	Mar 2, 2011	11:43 PM	Mar 9, 2011	11:43 PM
Reed, Megan	Mammogram / P	Mar 3, 2011	9:55 AM	Mar 10, 2011	9:55 AM
Stewart, Kim	Mammogram / F	Mar 3, 2011	10:09 AM	Mar 10, 2011	10:09 AM
Bennett, Claire	Pap Smear / F	Mar 7, 2011	8:01 PM	Mar 14, 2011	9:01 PM
Cox, Pamela	Blood (BMP) / F	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM
Cox, Pamela	Blood (CBC) / F	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM
Gray, Bridget	Blood (BMP) / F	Mar 6, 2011	10:48 PM	Mar 13, 2011	11:48 PM
Gray, Bridget	Blood (CBC) / F	Mar 6, 2011	10:48 PM	Mar 13, 2011	11:48 PM
Cooper, Molly	X-Ray / F	Mar 8, 2011	10:22 AM	Mar 15, 2011	11:22 AM
Howard, Amanda	Blood (BMP) / P	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM
Howard, Amanda	Blood (CBC) / P	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM

Patient	Test	Order Date	Result Due	Status		
Morris, Monica	TSH	February 15, 2011	2:23 PM	February 21, 2011	2:23 PM	Sample drawn and sent
Parker, Amy	MRI	March 2, 2011	11:43 PM	March 8, 2011	11:43 PM	Image(s) interpreted and reported
Peterson, Audrey	Blood (BMP)	March 3, 2011	7:53 PM	March 9, 2011	7:53 PM	Test ordered
Peterson, Audrey	Blood (CBC)	March 3, 2011	7:53 PM	March 9, 2011	7:53 PM	Test ordered
Reed, Megan	Mammogram	March 3, 2011	9:55 AM	March 9, 2011	9:55 AM	Image(s) interpreted and reported
Sanders, Chelsey	Pap Smear	March 3, 2011	4:54 AM	March 9, 2011	4:54 AM	Specimen collected and sent

Figure D.1: Rich tables adhere to the design guidelines. Results that have returned are listed at the top in “Results to Review”, while orders that have not returned to the physician are shown under “Pending Test Results”. Orders that have been placed already but will take effect in the future can be accessed in the “Planned Tests” (collapsed here). All tables are sorted by default so as to visually aid users see important results at the top. Newly arrived results are yellow, late orders are orange, and not completed are red.

faster. While some of these principles may apply to tablets or smart phones with touch-based interactions, they are mainly developed for desktop interfaces that are controlled with a mouse device.

D.1 Data Arrangement

D.1.1 Columns/Rows:

1. Sort the table according to one or more column(s) by default, arranged vertically down (Evidence: High; Conformance: Mandatory). The designer

should make a list of important information the table is going to convey in a decreasing order of priority [3, 44, 87, 91]. Then, the table is sorted according to these criteria. For example, the most important information in the tables of Figure D.1 is whether an item is late since the decision makers are expected to complete their tasks with no or little delays. Thus, the table is sorted according to this information first. Then, it is sorted according to whether items have been viewed so that users can focus on items unseen before. The third sort criteria is whether something is abnormal and needs immediate attention. Finally, the items have to be grouped by similarity since users look at similar items together. Sorting increases perception because it is easier to see the most important data at the top of the table; cognition since the user is relieved from computing the ordering in their head to make sense of the items, and motor performance due to reducing the amount of scrolling needed to find the necessary information.

2. *Permit re-sorting of tables with a click on the column header (Evidence: High; Conformance: Mandatory).* While default sorting gives the most natural ordering of items, users should be able to modify the sorting easily and given an option to revert back to the default sorting [3, 86, 87, 94, 121]. This improves cognition in situations when different orders have to be considered. Having re-sorting as easy as a click increases motor performance.
3. *Avoid horizontal scrolling in the default view (Evidence: High; Conformance: Recommended).* It is useful to lay the tables out in a readable way initially [3, 87]. After the first sight, when users ask for more information, they should be able to access it. Availability of extra columns should be explicitly indicated on the table, e.g. the last column might have an arrow that instructs users

to click for more information [3] (Figure D.2). This improves perception, cognition, and motor performance.



Figure D.2: The rightmost column header indicates extra columns are available.

4. *Focus on the data itself (Evidence: Medium; Conformance: Mandatory).*

While the design of the table is important in conveying the data, the primary purpose of a table is to represent information [44]. The data itself should be the most prominent feature (as in Figure D.1). See Figure D.3 for a bad example, which styles all the elements in the window the same way. This facilitates perception, cognition, and motor performance.

Planned Tests		
Patient	Test	Ordered By
Kelly, Cecelia	X-Ray	Brown, Joe (Provider)
Price, Chloe	Urinalysis	Brown, Joe (Provider)

Figure D.3: Excel defaults to the same font style and size for the title of the table, column headers, and table contents. The table data is not easily differentiable.

5. *Use sort icons in column headers to communicate that the table is sortable; conventionally upward/downward arrow for ascending/descending values, while the arrow size indicates sort priority (Evidence: Medium; Conformance: Mandatory).* Once the table is sorted, it is important to provide immediate feedback to the user [87]. All this information can be conveniently communicated via arrows in the column headers. Arrow direction presents sort order while arrow size indicates priority (see Figure D.4). This greatly assists perception as well as cognition for comparison.
6. *Perform computations for users; value, derived from data, should be readily available in the cells (Evidence: Low; Conformance: Mandatory).* If some

Patient ▲	Test	Order Date	Review By 🔥	Abnormality ▼
-----------	------	------------	-------------	---------------

Figure D.4: Note the sort icons on column headers. Patient and Test columns have categorical values that are placed to the left of Order Date and Review By columns, which hold quantitative values.

values have to be calculated from the given data, these should be readily made available in a separate column (e.g. “Review By” or “Result Due” date in Figure D.1, neither of which occur in the data itself). This especially improves cognition. See Figure D.5 for example columns that require users to compute. However, caution should be given if the number of columns need to increase as this may introduce horizontal scrolling.

Expected...	Elapsed Time
7 d	2 d

Figure D.5: The deadlines need to be calculated based on current time, expected duration, and elapsed time.

7. *Reduce the number of columns whenever possible (Evidence: Low; Conformance: Recommended).* Due to the small size of visual memory and the difficulty of searching through complex information [121], the implementation should remove unnecessary columns as much as possible via preprocessing the data in tables. More importantly, unnecessary columns of data waste valuable screen space and enforce people to wade through information that they do not need, which wastes their time. This can increase perception by making more important columns pop out, cognition by allowing a quick understanding of what the data presents, and motor performance via decreasing horizontal scrolling.
8. *Remove a column that always has the same value to save space (Evidence: Low; Conformance: Recommended).* Although it is essential to keep some

columns, other columns might communicate information that can readily be seen on the display. For instance, the data of Figure D.1 contains “Ordered By” field. Within the physician view this is always the same so it is removed to de-clutter the table. Although this may decrease perception of this information, it increases the perception of other elements in the table. It also helps avoid horizontal scrolling in the default view, which may improve motor performance.

9. *Use endless scrolling when all results do not comfortably fit within one page (Evidence: Low; Conformance: Recommended).* To avoid loading time, sometimes designers choose to show only a predefined number of items into the table. When users want to see more data, each time they click a button, such as “More” (Figure D.6a) or “Next” (Figure D.6b), at the end of the table to load more items. These are both successful commercial examples. When the data is as critical as in the medical domain, adding that additional click after each and every one-page scroll is redundant. The table should permit endless scrolling when all results do not comfortably fit within one page (Figure D.1). This increases motor performance as it eliminates the clicks after a scroll per page.



Figure D.6: Buttons that show more items in a list.

10. *Combine columns when appropriate (Evidence: Low; Conformance: Recommended).* Instead of having a column per attribute, information may be aggregated in one column. Especially columns that can only take a prede-

finer set of values may be combined with other columns. For example, the test name and type of result (“Finalized” or “Preliminary”) are combined in the “Test” column (Figure D.7). This makes the data easier to scan, improving perception. It may have a negative effect on cognition especially if the individual column values have to be compared to each other. One way to circumvent this is to make the most important column the first part in the aggregated column so that any such comparison can be quickly done by simply looking at the beginning of the column.

Dietetics / P
Blood (BMP) / P
Blood (CBC) / P
Pap Smear / P

Figure D.7: Instead of spelling out ‘Preliminary’ for each order, the term is abbreviated to ‘P’ and combined with the order name.

D.1.2 Row Sequence:

1. *Put the most severe row at the top of the table while ensuring that the most important rows are still visible (Evidence: Medium; Conformance: Mandatory).* Tables with severity criteria should be sorted by this row such that the most severe cases appear at the top of the table [91]. It is important that the criteria used to define severity do not cause an overwhelming number of items to be flagged as such. When there are too many alerts, people learn to ignore them or turn them off [41]. The design should enforce the perception of all severe rows in decreasing importance. Otherwise, users need to scroll too much (i.e. poor motor performance). For example, the most severe case

in Figure D.8 is a not completed order and such instances are put at the top of the table while still displaying late orders.

Patient ▲	Test	Order Date ▲	Result Due 🕒	Status
Morris, Monica	TSH	February 15, 2011 2:33 PM	February 21, 2011 2:33 PM	Sample drawn and sent
Parker, Amy	MRI	March 2, 2011 11:42 PM	March 8, 2011 11:42 PM	Image(s) interpreted and reported
Peterson, Audrey	Blood (BMP)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered
Peterson, Audrey	Blood (CBC)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered

Figure D.8: The not completed order appears at the top, followed by late orders. In addition, same patient orders are grouped together.

2. *Group related rows together so they are close in proximity for comparisons (Evidence: Low; Conformance: Recommended).* Groups of rows that will be used for comparisons should be placed together. This could be an option that could be switched on and off. It results in better perception, cognition, and motor performance. For example, Figure D.8 groups results by patient name because clinicians tend to look at results per patient.

D.1.3 Column Sequence:

1. *Offer rearranging of columns (Evidence: High; Conformance: Recommended).* To change the default column order, support rearrangement [3, 87, 94, 121] with a drag-and-drop. Users can drag the column header and drop at their desired location to move any column. It enhances perception and cognition. In cases when rearrangements can become confusing, users should have the option to change back to the default arrangements.
2. *Place sets of categorical values to the left of the quantitative values associated with them (Evidence: Medium; Conformance: Mandatory).* Cells that contain different quantitative values associated with a categorical value should appear on the right [44]. These categorical values will be read first (in most languages that read from left to right) and assist in comparing the quanti-

tative values on the right. It promotes both perception and cognition as it facilitates comparisons. Note that the “Test” column is placed before the date columns in Figure D.4.

3. *Place columns containing data that should be compared close together (Evidence: Medium; Conformance: Mandatory).* If comparisons will be performed on one or more columns, such columns should appear one after another to prevent eye movements or scrolling on the window [44]. Cognition and motor skills are impacted by this guideline. The date columns in Figure D.4 are placed one after the other. If placing columns next to each other is not possible, placing them as close as possible without the need for scrolling is desired.
4. *Organize the most important columns on the left to permit reading in the conventional left-to-right order (Evidence: Medium; Conformance: Recommended).* Most languages are read from left to right so rows will be read in that order [87]. Keeping the most essential information on the left side of the window reduces eye movement on the window. Figure D.4 keeps the “Patient” and “Test” information on the left. This enhances perception and might also speed up motor performance if such columns require scrolling.

D.1.4 Related Tables:

1. *Use just enough space between tables to make them noticeable (Evidence: Medium; Conformance: Recommended).* When the tables are lined up vertically or horizontally, there is just enough empty space between tables [121]. This facilitates perception and cognition when tables contain similar data.
2. *Size the tables according to their frequency of usage (Evidence: Medium;*

Conformance: Recommended). The most frequently used table could have the largest size in the window to allow rapid access and interaction. The benefits are in terms of perception and motor skills according to Fitts' law [46]. The largest table of Figure D.1 is “Results to Review” while “Planned Tests” are collapsed because the most frequently used table is “Results to Review”. Users can customize this layout by dragging the split panes vertically as needed (Figure D.9). In addition to using size to create a hierarchy of importance, there may be other visual attributes to make tables more salient than others, such as borders of varying intensities.



Figure D.9: The split pane between the tables permit users to modify default table sizes. While the upward/downward arrows (leftmost) are to expand/collapse, dragging the dot (middle) resizes.

3. *Filter a table that is not used often to show only the important data or stretch the table with support for full-view or expansion on-demand (Evidence: Low; Conformance: Mandatory)*. When a table is not used often, such as “Pending Test Results”, it is filtered to show only the important data or collapsed like “Planned Tests” with expansion on-demand (Figure D.10). This improves perception and motor performance for more important tables but reduces perception and motor performance for less important tables.

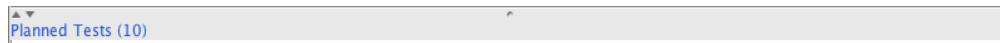


Figure D.10: The collapsed planned table can be expanded by clicking the upward arrow on the leftmost.

4. *Order the tables according to their frequency of usage (Evidence: Low; Con-*

formance: Recommended). Either vertically or horizontally, the order of tables could be assigned by their frequency of usage. For example, the most frequent table “Arrived Results” appears at the top while “Planned Tests” appears at the bottom in Figure D.1. This promotes perception and motor skills because the more accessible table is the most frequently used table (when scrolling is unnecessary).

5. *Keep table structure consistent from table to table (Evidence: Low; Conformance: Recommended)*. If all the tables have the same columns this is easier to achieve by using the same order for the columns. However, if the tables contain different columns or a different number of columns, one way to keep the structure consistent is to place the distinct columns at the rightmost end to allow for alignment of the same columns on the left side. Figure D.1 aligns “Patient”, “Test”, and “Order Date” columns on the left side for “Results to Review” and “Pending Test Results”. The advantages include faster perceptual and cognitive performance because the columns in different tables can be read and compared quickly.

D.2 Labeling

D.2.1 Style headers differently but ensure consistency (**Evidence: High; Conformance: Mandatory**)

Regardless of whether it is a column, row, or table header, the font style should be consistent within its own group but different from each other and the table data [121]. This particularly helps with perception because various information communicated through headers stands out against others. “Results to Review”,

“Pending Test Results”, and “Planned Tests” all have the same font style and a different style from column headers or table data (Figure D.1).

D.2.2 Give the table a descriptive title with a total row count (Evidence: High; Conformance: Mandatory)

A table should have a title placed at the top of the table that clearly describes in a couple of words what the table contains [3, 87, 91, 121]. Row counts can be appended to table titles. Tables that dynamically grow or allow filtering should automatically update their counts. These counts help perception and cognition. “Pending Test Results” (Figure D.11) shows the number of items after applying the filters.



Figure D.11: Table title contains row count (after filters are applied).

D.2.3 Keep the headers visible in the window at all times (Evidence: High; Conformance: Mandatory)

If the table becomes large, make the table headers stay visible in the window [86] as users scroll up/down and left/right a page to remind users of the rows and columns in the table (Figure D.12). This is useful for perception [91]. In addition, motor performance improves as this guideline eliminates scrolling to the beginning of the table. However, one row space is wasted for the header.

Patient ▲	Test	Order Date		Review By 🔥		Abnormality ▼
Cooper, Molly	X-Ray / F	Mar 8, 2011	10:22 AM	Mar 15, 2011	11:22 AM	⚠️
Howard, Amanda	Blood (BMP) / P	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM	⚠️
Howard, Amanda	Blood (CBC) / P	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM	⚠️
Phillips, Sarah	TSH / F	Mar 8, 2011	5:39 AM	Mar 15, 2011	6:39 AM	⚠️
Phillips, Sarah	Pap Smear / P	Mar 8, 2011	5:39 AM	Mar 15, 2011	6:39 AM	⚠️
Richardson, Allison	Blood (CBC) / P	Mar 7, 2011	11:07 PM	Mar 15, 2011	12:07 AM	⚠️
Bennett, Claire	Pap Smear / F	Mar 7, 2011	8:01 PM	Mar 14, 2011	9:01 PM	
Cox, Pamela	Blood (BMP) / F	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM	
Cox, Pamela	Blood (CBC) / F	Mar 8, 2011	9:34 PM	Mar 15, 2011	10:34 PM	
Gray, Bridget	Blood (BMP) / F	Mar 6, 2011	10:48 PM	Mar 13, 2011	11:48 PM	
Gray, Bridget	Blood (CBC) / F	Mar 6, 2011	10:48 PM	Mar 13, 2011	11:48 PM	
Gray, Bridget	X-Ray / P	Mar 7, 2011	4:01 PM	Mar 14, 2011	5:01 PM	

Figure D.12: Column headers are still visible when the table is scrolled down.

D.2.4 Align column headers with their associated data (Evidence: High; Conformance: Mandatory)

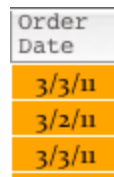
Headers and their associated data should be aligned correspondingly [86, 87]. It shows that the data is associated with the header, i.e. improves perception. It becomes handy when different alignments are used throughout the table. Note that “Patient” column header is left-aligned as well as all the patient names that appear below it (Figure D.13).

Patient ▲
Parker, Amy
Reed, Megan
Stewart, Kim
Richardson, Allison

Figure D.13: Patient column header and the patient names are left-aligned.

D.2.5 Avoid a header that is significantly wider than the data it is indicating by spreading such headers into two or more lines (Evidence: Medium; Conformance: Mandatory)

Headers that are significantly wider than the data impede horizontal scanning of cell values. Therefore, they should be spread into two or more lines to reduce the width of the column or avoid truncation [86] (Figure D.14). This positively impacts perception and motor skills for horizontal scrolling. However, this guideline may require more vertical scrolling (motor skills) as the new header covers an extra row.



Order Date
3/3/11
3/2/11
3/3/11

Figure D.14: Long headers are split into two or more lines

D.2.6 Indicate editable columns (Evidence: Low; Conformance: Mandatory)

Include a pen icon (e.g. “Result Due” column header in Figure D.15) or a visual indicator next to the header or each cell to point out that the value could be edited in place. If an entire column is editable moving the icon to the header saves space and minimizes distractions from data. This guideline is for perceptual performance.

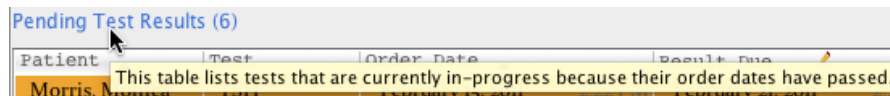


Result Due
February 21, 2011 2:22 PM
March 8, 2011 11:42 PM

Figure D.15: Result Due column header has a pen icon to indicate the cells below it can be edited.

D.2.7 Show a tooltip for the title that describes the table's function (Evidence: Low; Conformance: Recommended)

While the title is compact to leave room for the table contents, the title should display on hover a tooltip a more comprehensive explanation (in one or two sentences) of the table's function (Figure D.16). It could even include concrete examples for clarification. This mostly assists in cognition of the table. The duration of the tooltip on screen can be increased depending on the length of the description.



Patient	Test	Order Date	Result Due
Morris, M			

Figure D.16: A tooltip for the table title describes the table's function.

D.2.8 Do not truncate column headers; break long headers by full words whenever possible, otherwise split in the middle with a hyphen (Evidence: Low; Conformance: Recommended)

Column headers should not be truncated because they communicate information that applies to the entire table. If there are many columns, users might not remember the meaning. Use familiar or otherwise clear abbreviations as they save space. If there are no abbreviations with clear meanings, split long headers by full words in two or more lines. When this is not possible (i.e. there is one long word), split the word in the middle with a hyphen (Figure D.17). This affects perception.

However, this guideline may require more vertical scrolling (motor skills) as the new header covers an extra row.

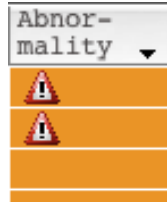


Figure D.17: Column header is split with a hyphen when the data itself takes little space or the column is dragged.

D.3 Settings & Help

D.3.1 Provide custom filtering on-demand (Evidence: High; Conformance: Mandatory)

When users do not remember what they are searching for, filters become practical. For more important and frequently used features, filters could appear on the main display (e.g. “Late” and “Not Completed” radio buttons in Figure D.18) while for more complex features, this could be performed on a separate panel that does not distract from the table (Figure D.18). Filters can be selected from radio buttons when only one selection is possible or from checkboxes when multiple selections are possible [91, 121]. This guideline promotes perception since filtered values will no longer distract from the desired values and motor performance as it reduces scrolling.

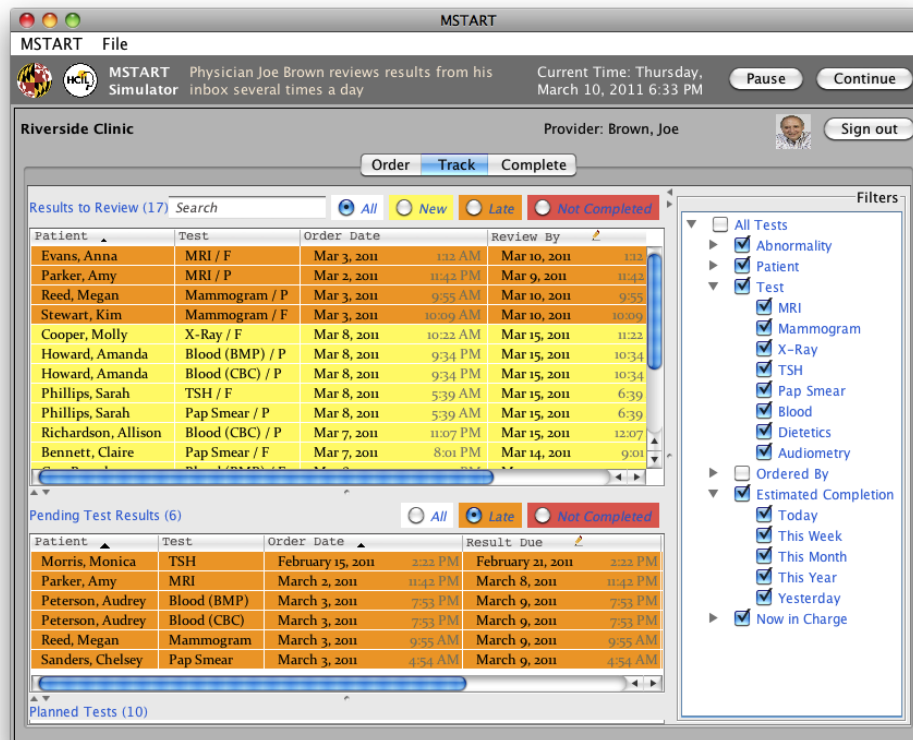


Figure D.18: The panel on the right is opened (by clicking collapse/expand arrows at the top left of the panel) to allow for custom filtering. Users can customize their filter selections from here.

D.3.2 Allow settings to be saved and loaded (Evidence: Medium; Conformance: Recommended)

Although tables support views based on user roles, users should also be able to make changes, which persist automatically, on the default table preferences [87]. Next time when the table is displayed again, the table should load with the saved settings. In addition, users should be allowed to revert to default settings at any time. This could be done through a preference pane. This guideline has to do with perception and motor skills because the default settings might not be optimized

for certain users.

D.3.3 Do not let tooltips go outside of screen space (Evidence: Low; Conformance: Mandatory)

Tooltips stay at the point of mouse location for a short amount of time but if the mouse is close to a screen corner, some parts of tooltips might leave the screen space. To prevent this, the designer has to slide the tooltips such that they are still close to the mouse location but do not partially disappear. It supports perception because the tooltip is in sight; cognition as the description is understandable; and motor skills since users do not have to move the mouse or the application window to read the tooltip.

D.3.4 Provide help in a separate window (Evidence: Low; Conformance: Mandatory)

While tooltips guide users, a help window that compiles together all descriptions in one-page is still beneficial. Tooltips stay on the screen for a short period and users might prefer to read or search them in a separate window. In addition, novice users would like to start by reading a help menu to learn more. Because a window stays on screen longer than a tooltip, this guideline serves perception and cognition. Although this may hinder motor performance because users have to open up a new window from the menu, mouse movements on the screen are reduced significantly which might result in better performance overall.

D.3.5 Derive filter values from current table entries, not all database entries (Evidence: Low; Conformance: Recommended)

One option for creating filter values is to provide a list of all entries in the database, i.e. there will be filter values with no matches in the table. To limit the number of possibilities users have to deal with (such as unavailable filter options), filters can be derived from the current table values. When there are modifications in the table, filters should be updated accordingly. ‘Test’ filter in Figure D.18 lists only test values that appear in the tables, not all possible tests. This is effective for perception because users can focus on only relevant data and motor skills because there are fewer filters to scroll and click.

D.3.6 Group filter values by range if possible (Evidence: Low; Conformance: Recommended)

One way to reduce possible filters is grouping them into ranges when columns represent integers or dates. ‘Estimated Completion’ filter in Figure D.18 has ‘Yesterday’, ‘Today’, ‘This Week’, ‘This Month’, ‘This Year’ rather than each day. Due to the same reasons from the previous guideline, it helps with perception and motor skills.

D.3.7 Support search for large tables (Evidence: Low; Conformance: Recommended)

In large tables, users should be allowed to search by keywords when they know what they are looking for. One possibility is a fuzzy search, which returns results similar to the given keyword. Even for exact keyword searches, there needs to be a suggestion in case of misspellings. Figure D.19 has a search box at the top of

the tables. This guideline should be applied for perception and motor skills.



Figure D.19: Large tables need a search option.

D.3.8 Show a description for columns in a tooltip (Evidence: Low; Conformance: Recommended)

If there is ambiguous data, such as columns with the same type of values, tooltips on header hovers can display their description (Figure D.20). The advantage is better cognition of the columns meaning.

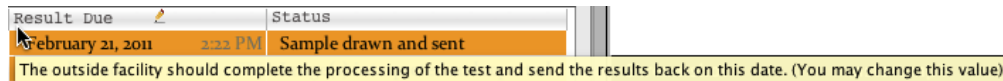


Figure D.20: Tooltips for headers provide more information about the column.

D.4 Delineation

D.4.1 Feature light white space between the rows, no heavy gridlines (Evidence: High; Conformance: Mandatory)

To be able to scan the table efficiently from top to bottom, some delimiters need to be between the rows of the table. This helps the eye to differentiate one row from another. Otherwise, users are confused when they reach the middle of the table. While one option is to alternate the fill colors (in a striped style) for the table entries [3, 44], this is not applicable when the table is color-coded. Featuring light white space between the rows [91, 141] serves the same purpose (Figure D.21b). If delimiters dominate the table content (e.g. thick lines in Figure D.21a, long white

spaces) then the data occupies less space and is harder to scan. Perception and motor performance is slower with extra space due to eye movement and scrolling. Light white space means a color that is faint, i.e. just visible enough to do the job and no more. White against a dark background would stand out too much, just as black against a light background does.

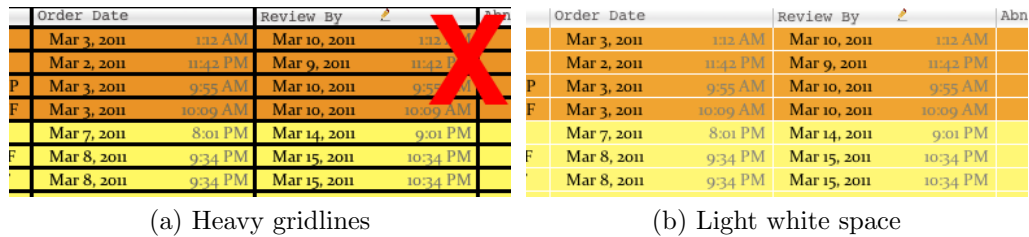


Figure D.21: Heavy gridlines versus light white space between the rows.

D.4.2 Allow sufficient space between columns to clearly separate them but no more (Evidence: High; Conformance: Mandatory)

To scan an entry efficiently from left to right, delimiters should be used between columns (Figure D.21). This helps users read different information regarding the entry. Otherwise, users can not understand the attributes of an item. On the other hand, excessive space impedes horizontal scanning [86,141]. Perception and motor performance suffer in such cases due to eye movements and scrolling. Cognition may be affected in a negative way if users perform operations on more than one column (e.g. comparing values in two columns that are far apart require temporary memorization).

D.4.3 Calculate initial column widths from data but offer resizing of columns (Evidence: Medium; Conformance: Recommended)

Instead of assigning a default equal width to each column (Figure D.22a), compute widths from the given data for the table. This lays out the data with no truncations or excessive spacing so that users do not need to adjust column widths (Figure D.22b). This, in turn, assists in perception and motor performance because readability is increased and there is no need to drag and resize columns. Offer resizing on demand [3, 94].

Patient	Test	Ordered	Alert	Abnormality
Parker, Amy	MRI / P	Brown, Joe...	03/03/2011...	Normal
Reed, Megan	Mammogra...	Brown, Joe...	03/04/2011...	Normal
Stewart, Kim	Mammogra...	Brown, Joe...	03/04/2011...	Normal
Evans, Anna	MRI / F	Brown, Joe...	03/04/2011...	Normal
Gray, Bridget	Blood (CBC)...	Brown, Joe...	03/07/2011...	Normal
Phillips, Sa...	Pap Smear...	Brown, Joe...	03/08/2011...	Abnormal
Gray, Bridget	Blood (BM)...	Brown, Joe...	03/08/2011...	Normal

(a) Equal column widths

Patient	Test	Ordered By	Alert Date ▲	Abnormality
Parker, Amy	MRI / P	Brown, Joe (P)	03/03/2011@21:53	Normal
Reed, Megan	Mammogram / P	Brown, Joe (P)	03/04/2011@01:26	Normal
Stewart, Kim	Mammogram / F	Brown, Joe (P)	03/04/2011@03:15	Normal
Evans, Anna	MRI / F	Brown, Joe (P)	03/04/2011@22:28	Normal
Gray, Bridget	Blood (CBC) / F	Brown, Joe (P)	03/07/2011@20:18	Normal
Phillips, Sarah	Pap Smear / P	Brown, Joe (P)	03/08/2011@09:43	Abnormal
Gray, Bridget	Blood (BMP) / F	Brown, Joe (P)	03/08/2011@13:46	Normal

(b) Column widths calculated from data

Figure D.22: Column widths that are equal versus calculated from data.

D.4.4 Define a min and max width for each column depending on the data it presents; these widths should be user changeable (Evidence: Medium; Conformance: Recommended)

Columns should not be allowed to shrink to the extent that cell values are not readable anymore or they should not be stretched out to the extent they impede with horizontal scanning (see Figure D.23 for a bad example). Such examples negatively affect both perception and motor performance. Default min and max widths could be calculated from the possible cell values [87] and could be changed by users.

Cooper, Moll...	X-Ray / F	Mar 8, 2011	10:22 AM
Howard, Am...	Blood (BMP) / P	Mar 8, 2011	9:34 PM
Howard, Am...	Blood (CBC) / P	Mar 8, 2011	9:34 PM
Phillips, Sara...	TSH / F	Mar 8, 2011	5:39 AM
Phillips, Sara...	Pap Smear / P	Mar 8, 2011	5:39 AM
Richardson,...	Blood (CBC) / P	Mar 7, 2011	11:07 PM
Bennett, Clai...	Pap Smear / F	Mar 7, 2011	8:01 PM

Figure D.23: While the first column is not readable in full, second column impedes horizontal scanning.

D.4.5 Do not stretch tables to fill available space; only the last column may be stretched, if not right-aligned, to align with other tables (Evidence: Medium; Conformance: Recommended)

To fill out extra space on the window, table size should not be stretched out in a way that harms users' visual scanning [86]. Table size should instead be dependent on the data it presents. The last column of a table could be stretched to align with another table's size. This is only possible if the last column is not right-aligned. If a right-aligned column is stretched, this might interfere with horizontal scanning and thus, perception.

D.4.6 Add some padding to columns if there is room; allow user-defined values (Evidence: Medium; Conformance: Recommended)

All values should stand out against the ones in other cells. While avoiding excess space between columns aids in horizontal scanning, some padding separates different columns [121]. Placing values in adjacent cells too close together impedes perception and cognition (Figure D.24a). In particular, columns that have the same type of data should be distinguishable to allow for comparison (Figure D.24b). In contrast, less spacing and padding might provide better motor performance due

to little need for scrolling. But, if cells or rows are clickable, motor performance improves with enough spacing and padding because they become easily clickable, thereby preventing mis-clicks.

Parker, Amy	MRI	Parker, Amy	MRI
Peterson, Audrey	Blood (P/W)	Peterson, Audrey	Blood (BMP)
Peterson, Audrey	Blood (CBC)	Peterson, Audrey	Blood (CBC)

(a) No padding

(b) Padding

Figure D.24: Columns without and with padding.

D.4.7 Add some padding to rows for easy visual scanning; allow this to be set by the user (Evidence: Medium; Conformance: Recommended)

First of all, row height should be set so that the text fits perfectly [87] (Figure D.24a). To scan down the rows, each row needs some space between to be able to read it easily [121] (Figure D.24b). While scanning down, users are looking at one column, which has the same type of data. However, adding padding results in smaller fonts, which decreases readability for many users. Maintaining font size will reduce the number of items displayed, which will increase short term memory load if users want to compare list items, and will increase navigation time as scrolling is increased. Based on the previous guideline, this guideline improves perception, cognition, and motor performance.

Parker, Amy	MRI	Parker, Amy	MRI
Peterson, Audrey	Blood (BMP)	Peterson, Audrey	Blood (BMP)
Peterson, Audrey	Blood (CBC)	Peterson, Audrey	Blood (CBC)

(a) No padding

(b) Padding

Figure D.25: Rows without and with padding.

D.5 Formatting

D.5.1 Orientation

Avoid text orientations other than horizontal, left to right (Evidence: Medium; Conformance: Mandatory). Since most languages are read in left to right order, vertical or diagonal orientations interfere readability [44]. Fast reading improves cognition. In addition, various text orientations result in different scrolling patterns.

D.5.2 Alignment

1. *Ensure consistency in the alignment of similar data (Evidence: Medium; Conformance: Mandatory).* Similar data types should be aligned in a consistent manner throughout the table [121]. For example, if a column that can take a set of values is left-aligned, another column that can take a similar set of values should also be left-aligned. Note that ‘Patient’ and ‘Test’ columns take text values and both are left-aligned (Figure D.26). This guideline attends to perceptual and cognitive performance because users can see and understand that different columns have similar data.
2. *Align the time and date markers for all numerals in a column (Evidence: Medium; Conformance: Recommended).* Just like decimal separators, date

Patient ▲	Test
Parker, Amy	MRI / P
Reed, Megan	Mammogram / P
Stewart, Kim	Mammogram / F

Figure D.26: Patient and Test columns with text values are both consistently left-aligned.

and time markers (colon, period, AM/PM indicator) in a column have to be aligned for fast perceptual and cognitive performances [121]. Note that this means a consistent number of digits for months, days, and years (Figure D.27). Some space might need to be wasted in the cell to allow for alignment.

Order Date	
Mar 2, 2011	11:42 PM
Mar 3, 2011	9:55 AM
Mar 3, 2011	10:09 AM
Mar 7, 2011	11:07 PM
Mar 7, 2011	11:07 PM

Figure D.27: Order Date is double-aligned, meaning date values are left-aligned and time is right-aligned.

3. *Bottom-align column headers (Evidence: Low; Conformance: Mandatory).*

When all columns are one line only, top and bottom alignment are essentially the same. When some column headers are split into two or more lines, one line headers can either be top or bottom aligned with the rest (Figure D.28). Top alignment will create a distance between the header and the cells below it so bottom alignment is a better choice in this case. This guideline has to do with the perception of the column headers.

4. *Align numeric (time) values to the right while keeping all other values in the table left-aligned (Evidence: Low; Conformance: Recommended).* Time values, which depict numerical data, are aligned to the right for rapid com-

Patient ▲	Test	Order Date	Review By 🕒	Abnormality ▼
Evans, Anna	MRI / F	3/3/11	Mar 10, 2011 11:12 AM	
Parker, Amy	MRI / P	3/2/11	Mar 9, 2011 11:42 PM	
Reed, Megan	Mammogram / P	3/3/11	Mar 10, 2011 9:55 AM	

Figure D.28: Headers are bottom-aligned when they are split in two or more lines.

parison. Note that this means date values are left-aligned (Figure D.27). All other cells that include text remain left-aligned since text is read from left to right. This guidelines enhances cognitive performance.

5. *Left-align columns with icons for horizontal continuity (Evidence: Low; Conformance: Recommended).* One way to align columns with an icon or a single character is center-alignment since the content will not take much space (Figure D.29). This introduces space on both ends of the cell and interferes with eye movement. Moving such columns to the far right end and left-aligning for horizontal continuity boosts perception. If the column is in the middle of the table, then center-alignment might be a better choice.



Figure D.29: Abnormality column makes use of center-alignment. Thus, it introduces extra spaces on the left and right end of the column.

6. *Double-align the day and time in a date column (Evidence: Low; Conformance: Recommended).* A column that conveys more than one type of information can utilize double-alignment to simplify the complex data format.

One such column type is a date column that contains both day and time information. To allow for the comparison of time values and still use a consistent alignment for days, days can be left-aligned while the times are right-aligned (Figure D.27). This helps the perception and cognition of these cell values. Note that different fonts are also used to communicate the double-alignment as the adjacent columns might be associated with the left or right-aligned data.

D.5.3 Number and Date Precision

1. *Round values displayed in table cells where it is not misleading to do so (Evidence: Medium; Conformance: Mandatory).* Numeric values that have too much precision could be simplified for readability by rounding [86]. For instance, time values could be hidden when there is no room in the date column (Figure D.31c). Perceptual, cognitive, and indirectly motor performance are affected by this guideline. This partially reduces the perception of the precision but increases the perception of other table elements. Cognition deteriorates with less precision but may result in less scrolling.
2. *Use abbreviation for those values that have the same substring (Evidence: Medium; Conformance: Recommended).* Abbreviating text in a column that repeatedly contains a particular substring can significantly reduce the cell size. Figure D.7 writes ‘P’ for preliminary results rather than the full word, ‘Preliminary’. However, a list of the abbreviations should be available to users [121]. It could be in a menu or shown alongside with the tables. If the list of abbreviations is visible, it is helpful for perception. Abbreviations may hinder cognitive performance but reduce the amount of scrolling significantly

if the original text is too long.

3. *Truncate values if abbreviation does not apply, but use tooltips for showing details of values that do not fit in the cell (Evidence: Medium; Conformance: Recommended).* Abbreviation does not always apply in cases such as numerical values, when there is no universal agreement on the term, or if abbreviation might confuse the meaning. In such cases, truncation could be done by inserting an ellipsis at the end [3] but tooltips should still show details of values that do not fit in the cell (Figure D.30). While truncation reduces perception and cognition, tooltips can help with these. Truncations have both pros and cons for motor performance by reducing scrolling but introducing an extra hover over.

Mamm...	Mar 3, 2011
Mamm...	Mar 2, 2011
X-Ray / ...	Mammogram / P
Blood (...)	Mar 8, 2011

Figure D.30: After users change the column width, cells that do not fit in the given width get truncated. The full cell value appears in a tooltip.

4. *Gradually show more precision if space permits or users seek (Evidence: Low; Conformance: Recommended).* When columns are rounded, users should be able to see the full precision when they want to and if there is enough space on the window. This is usually implemented with a drag of column delimiters. More precision can be shown gradually (perhaps with a transitioning effect) via dragging (see Figure D.31a, D.31b, D.31c). This smoothens the perception and cognition of the precision. One danger is motor performance deteriorates as columns' sizes increase and horizontal scrollbars appear.

Order Date	Order Date	Order Date
March 3, 2011 1:12 AM	3/3/11 1:12 AM	3/3/11
March 2, 2011 11:42 PM	3/2/11 11:42 PM	3/2/11
March 3, 2011 9:55 AM	3/3/11 9:55 AM	3/3/11

(a) Max (b) Mid (c) Min

Figure D.31: As users drag the column header to the left, less information is presented gradually.

5. *Keep the precision consistent from column to column (Evidence: Low; Conformance: Recommended).* If there are multiple columns of the same type in a table, initially display them using the same precision. For example, the Results table uses a short month format (e.g. Mar 3, 2011) format in both ‘Order Date’ and ‘Review by Date’ columns (Figure D.32a), while the Pending table employs the longer month format (March 3, 2011) in both ‘Order Date’ and ‘Result Due’ columns (Figure D.32b). This does not mean as users interact with one of these columns, the others should be updated accordingly. In fact, users may prefer not to change them simultaneously. In addition, this guideline should be applied with care as one column may need to be seen at a particular precision while another column at a different precision. This guideline assists in perceptual performance.

Order Date	Review By	Order Date	Result Due
Mar 2, 2011 11:42 PM	Mar 9, 2011 11:42 PM	February 15, 2011 2:23 PM	February 21, 2011 2:23 PM
Mar 3, 2011 9:55 AM	Mar 10, 2011 9:55 AM	March 2, 2011 11:42 PM	March 8, 2011 11:42 PM
Mar 3, 2011 10:09 AM	Mar 10, 2011 10:09 AM	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM
Mar 7, 2011 11:07 PM	Mar 15, 2011 12:07 AM	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM
Mar 7, 2011 11:07 PM	Mar 15, 2011 12:07 AM	March 3, 2011 9:55 AM	March 9, 2011 9:55 AM
Mar 7, 2011 8:01 PM	Mar 14, 2011 9:01 PM	March 3, 2011 4:54 AM	March 9, 2011 4:54 AM

(a) Results (b) Orders

Figure D.32: By default, both tables make use of consistent precisions for columns with same type.

6. *Split cells into two lines when a value is too long after abbreviation and truncation (Evidence: Low; Conformance: Recommended).* After abbreviation and truncation, if the cell still does not fit, one option is to split the row into two lines (Figure D.33). Such an action has an effect on the entire row and introduces extra vertical space in the table view, which pushes possible rows from the bottom of the table out of the current page. Thus, this guideline should be used with caution. For example, it is preferable for rows that have more than some threshold number of long columns and should not be used for more than another threshold number of rows in one table page. This guideline considers perception, cognition, and motor skills. While perception and cognition is increased for the particular cell value, the perception of other rows in the current view is decreased. Also, it introduces more scrolling (hence poor motor performance).

Patient ▲	Test	Order Date ▲	Result Due 🕒	Status
Morris, Monica	TSH	February 15, 2011 2:22 PM	February 21, 2011 2:22 PM	Sample drawn and sent
Peterson, Audrey	Blood (BMP)	March 3, 2011 7:53 PM	March 9, 2011 7:53 PM	Test ordered

Figure D.33: Because the value in the last column does not fit, the cell is split into two lines.

D.5.4 Font

1. *Select a font family and size that is legible (Evidence: High; Conformance: Mandatory).* The font style and size should be readable [44]. Both sans (e.g. Arial, Helvetica) and serif (e.g. Times) typefaces are good for this, however, the tiny embellishments in serif fonts need extremely high definition to avoid looking blurry. While the font size may depend on the screen space, size of 12 is the most legible [91, 141]. This guideline aims at increasing perceptual

and cognitive performance.

2. *Ensure consistency in the typeface (Evidence: Medium; Conformance: Mandatory)*. While distinct elements are distinguished by their typeface styles, similar elements in the window need to have the same typeface to show the relationship [121]. This guideline is for perceptual and cognitive performance.
3. *Make table values prominent with a font style that is easily distinguishable from others used throughout the interface (Evidence: Medium; Conformance: Recommended)*. The main focus of a window with a table is the contents of the table. Thus, table data should be the most distinguishable factor on such a window. To achieve this, designers can choose darker font colors, a bigger font size, a more readable font family than the rest of the widgets on the window [121]. This complements perception and cognition of the information in the table.
4. *Do not drop to a smaller font to fit a table on the window (Evidence: Low; Conformance: Recommended)*. Font sizes must not be altered unless users set them manually. Automatic modification of the font interferes with perception and cognition of table contents. Any changes should be done with transition effects. In this case, the risks are higher. When the font size changes, the table layout will be recomputed again and this may require even more changes.

D.5.5 Color and Icons

1. *Change color saturation to make important information stand out (Evidence: Medium; Conformance: Recommended)*. If there are levels of severity in the color-coded information (Figure D.34), saturation can be employed [121].

This helps with perception and cognition of the color coded data. For people with color deficiencies, this might be difficult to see.

Evans, Anna	MRI / F	Mar 3, 2011	1:12 AM	Mar 10, 2011	1:12 AM
Parker, Amy	MRI / P	Mar 2, 2011	11:42 PM	Mar 9, 2011	11:42 PM
Reed, Megan	Mammogram / P	Mar 3, 2011	9:55 AM	Mar 10, 2011	9:55 AM
Stewart, Kim	Mammogram / F	Mar 3, 2011	10:09 AM	Mar 10, 2011	10:09 AM

Figure D.34: Use of saturation in color-coding for levels of severity.

2. *Use icons whenever possible (Evidence: Medium; Conformance: Recommended).* By cutting down the necessary reading, distinctive and meaningful icons can dramatically improve the speed at which a user can comprehend the information, i.e. better cognitive performance [3]. Icons can especially be used for boolean values such as when some condition evaluates to true (see ‘Abnormality’ column in Figure D.1). On the other hand, people can only remember a few icons and a screen with too many different icons can become overwhelming.
3. *Provide a color legend to show and filter based on the coding (Evidence: Low; Conformance: Mandatory).* The legend shows color coding and allows filtering. While the color coding displays important information in the table, the legend shows the meaning of different colors for cognition. In addition to an external filter panel, the legend allows quick filtering to enhance motor performance for this important criteria. Figure D.35 presents a color legend example that allows filtering with radio buttons. The color background decreases text legibility by decreasing contrast.



Figure D.35: Color legend shows and filters based on color coding.

4. *Use color coding to show grouping (Evidence: Low; Conformance: Recommended).* Most tables do not utilize colors, i.e. white background with black font. When rows are marked and grouped based on an important criteria, color coding the rows can communicate this information. It advances perception and cognition skills. To help universal accessibility, font style (such as boldface) can be used instead of color. This guideline should be reserved for very critical information that requires attention. In Figure D.1, the grouping is based on urgency and the colors also indicate this because it is the most crucial information in the table.

Bibliography

- [1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*, 1st ed., ser. Human-Computer Interaction. London, UK: Springer Verlag, 2011.
- [2] “Allscripts Enterprise EHR,” <http://www.allscripts.com/en/solutions/ambulatory-solutions/ehr/Show/ProductSelect/AllscriptsEnterpriseEHR/Overview.html>, Allscripts Healthcare Solutions, Inc., accessed: 01/01/2013.
- [3] *Apple Human Interface Guidelines*, Apple Computer, Inc. [Online]. Available: <http://developer.apple.com/>
- [4] C. R. Aragon, S. S. Poon, G. S. Aldering, R. C. Thomas, and R. Quimby, “Using Visual Analytics to Maintain Situation Awareness in Astrophysics,” in *IEEE Symposium on Visual Analytics Science and Technology*, ser. VAST’08, October 2008, pp. 27–34.
- [5] J. S. Ash, M. Berg, and E. Coiera, “Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors,” *Journal of the American Medical Informatics Association*, vol. 11, no. 2, pp. 104–112, 2004.
- [6] N. Baig, R. E. Myers, B. J. Turner, J. Grana, T. Rothermel, N. Schlackman, and D. S. Weinberg, “Physician-Reported Reasons for Limited Follow-Up of Patients with a Positive Fecal Occult Blood Test Screening Result,” *The American Journal of Gastroenterology*, vol. 98, no. 9, pp. 2078–2081, September 2003.

- [7] D. Barreau and B. A. Nardi, “Finding and Reminding: File Organization from the Desktop,” *SIGCHI Bull.*, vol. 27, pp. 39–43, July 1995.
- [8] S. Bäumler, M. Balsler, A. Dunets, W. Reif, and J. Schmitt, “Verification of Medical Guidelines by Model Checking – A Case Study,” in *Proceedings of the 13th International SPIN Workshop*, ser. Lecture Notes in Computer Science, A. Valmari, Ed. Berlin: Springer, 2006, vol. 3925, pp. 219–233.
- [9] V. Bellotti, B. Dalal, N. Good, P. Flynn, D. G. Bobrow, and N. Ducheneaut, “What a To-do: Studies of Task Management Towards the Design of a Personal Task List Manager,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’04. New York, NY, USA: ACM, 2004, pp. 735–742.
- [10] —, “What a To-do: Studies of Task Management Towards the Design of a Personal Task List Manager,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’04. New York, NY, USA: ACM, 2004, pp. 735–742.
- [11] O. Bergman, R. Boardman, J. Gwizdka, and W. Jones, “Personal Information Management,” in *CHI’04 extended abstracts on Human factors in computing systems*, ser. CHI EA’04. New York, NY, USA: ACM, 2004, pp. 1598–1599.
- [12] J. Bertin, *Semiology of Graphics: Diagrams, Networks, Maps*, W. Berg, Ed. Madison, WI: University of Wisconsin Press, 1983.
- [13] “The Big Table Issue,” <http://www.8164.org/the-big-table-issue/>.
- [14] A. E. Blandford and T. R. G. Green, “Group and Individual Time Management Tools: What You Get is Not What You Need,” *Personal Ubiquitous Comput.*, vol. 5, pp. 213–230, January 2001.

- [15] E. A. Boohaker, R. E. Ward, J. E. Uman, and B. D. McCarthy, "Patient Notification and Follow-up of Abnormal Test Results: A Physician Survey," *Archives of Internal Medicine*, vol. 156, pp. 327–331, 1996.
- [16] A. Bottrighi, L. Giordano, G. Molino, S. Montani, P. Terenziani, and M. Torchio, "Adopting Model Checking Techniques for Clinical Guidelines Verification," *Artificial Intelligence in Medicine*, vol. 48, pp. 1–19, January 2010.
- [17] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, and F. Y. eds., "Extensible Markup Language (XML)," W3C, Specification, November 2008, version 1.0 (Fifth Edition).
- [18] C. Campbell and P. Maglio, "Supporting Notable Information in Office Work," in *CHI'03 extended abstracts on Human factors in computing systems*, ser. CHI EA'03. New York, NY, USA: ACM, 2003, pp. 902–903.
- [19] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Eds., *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999.
- [20] S. K. Card, A. Newell, and T. P. Moran, *The Psychology of Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1983.
- [21] P. Carraro and M. Plebani, "Errors in A Stat Laboratory: Types and Frequencies 10 Years Later," *Clinical Chemistry*, vol. 53, no. 7, pp. 1338–1342, 2007.
- [22] A. Cass, A. Lerner, E. McCall, L. Osterweil, J. Sutton, S.M., and A. Wise, "Little-JIL/Juliette: A Process Definition Language and Interpreter," in *Proceedings of the 2000 International Conference on Software Engineering*, 2000, pp. 754–757.

- [23] S. Christov, B. Chen, G. S. Avrunin, L. A. Clarke, L. J. Osterweil, D. Brown, L. Cassells, and W. Mertens, “Formally Defining Medical Processes,” *Methods of Information in Medicine*, vol. 47, no. 5, pp. 392–398, 2008.
- [24] D. Chugh and M. H. Bazerman, “Bounded Awareness: What You Fail to See Can Hurt You,” *Mind Society*, vol. 6, no. 1, pp. 1–18, 2007. [Online]. Available: <http://www.springerlink.com/index/10.1007/s11299-006-0020-4>
- [25] M. Correll, D. Albers, S. Franconeri, and M. Gleicher, “Comparing Averages in Time-Series Data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’12. New York, NY, USA: ACM, 2012, pp. 1095–1104.
- [26] D. R. Cox and D. Oakes, *Analysis of Survival Data*. Chapman & Hall/CRC, 1984.
- [27] P. Cram, G. E. Rosenthal, R. Ohsfeldt, R. B. Wallace, J. Schlechte, and G. D. Schiff, “Failure to Recognize and Act on Abnormal Test Results: The Case of Screening Bone Densitometry,” *Joint Commission Journal on Quality and Patient Safety*, vol. 31, no. 2, pp. 90–97, 2005.
- [28] E. B. Cutrell, M. Czerwinski, and E. Horvitz, “Effects of Instant Messaging Interruptions on Computing Tasks,” in *CHI’00 extended abstracts on Human factors in computing systems*, ser. CHI EA’00. New York, NY, USA: ACM, 2000, pp. 99–100.
- [29] M. Czerwinski, E. Horvitz, and S. Wilhite, “A Diary Study of Task Switching and Interruptions,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’04. New York, NY, USA: ACM, 2004, pp. 175–182.
- [30] “Data Table,” <http://patternry.com/p=data-table/>.

- [31] L. Dazzi, C. Fassino, R. Saracco, S. Quaglini, and M. A. Stefanelli, "Patient Workflow Management System Built on Guidelines," in *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, 1997, pp. 146–150.
- [32] W. De Pauw and S. Heisig, "Zinsight: A Visual and Analytic Environment for Exploring Large Event Traces," in *Proceedings of the 5th International Symposium on Software Visualization*, ser. SOFTVIS'10. New York, NY, USA: ACM, 2010, pp. 143–152.
- [33] M. den Hengst, M. McQuaid, and J. Zhu, "Impact of time-based visualization on situation awareness," in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, ser. HICSS'04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 1–10.
- [34] D. J. Dixon, "User interface design guidelines for charts, tables and dashboards," Mayo Clinic Global Products and Services, Tech. Rep., September 2010.
- [35] H. Dong, D. Greenest, G. Fleishert, and I. Kohane, "An Alert System for ED Lab Test Results," in *Proceedings of the AMIA Symposium*, 1998, p. 994.
- [36] S. M. Dovey, D. S. Meyers, R. L. Phillips, L. A. Green, G. E. Fryer, J. M. Galliher, J. Kappus, and P. Grob, "A Preliminary Taxonomy of Medical Errors in Family Practice," *Quality & Safety in Health Care*, vol. 11, no. 3, pp. 233–238, 2002.
- [37] R. Elandt and N. L. Johnson, *Survival Models and Data Analysis*. John Wiley and Sons, 1980.
- [38] N. C. Elder, T. R. McEwen, J. M. Flach, and J. J. Gallimore, "Management of Test Results in Family Medicine Offices," *The Annals of Family Medicine*, vol. 7, no. 4, pp. 343–351, July–August 2009.

- [39] N. C. Elder, T. R. McEwen, J. M. Flach, J. J. Gallimore, and H. Pallerla, “The Management of Test Results in Primary Care: Does an Electronic Medical Record Make a Difference?” *Family Medicine*, vol. 42, no. 5, pp. 327–333, May 2010.
- [40] N. C. Elder, M. V. Meulen, and A. Cassedy, “The Identification of Medical Errors by Family Physicians During Outpatient Visits,” *The Annals of Family Medicine*, vol. 2, no. 2, pp. 125–129, March–April 2004.
- [41] M. R. Endsley, B. Bolté, and D. G. Jones, *Designing for Situation Awareness: An Approach to User-Centered Design*. Boca Raton, FL: Taylor & Francis Group, 2003.
- [42] A. Faulring, B. Myers, K. Mohnkern, B. Schmerl, A. Steinfeld, J. Zimmerman, A. Smailagic, J. Hansen, and D. Siewiorek, “Agent-Assisted Task Management that Reduces Email Overload,” in *Proceedings of the 15th international conference on Intelligent user interfaces*, ser. IUI’10. New York, NY, USA: ACM, 2010, pp. 61–70.
- [43] S. Fertig, E. Freeman, and D. Gelernter, ““Finding and Reminding” Reconsidered,” *SIGCHI Bull.*, vol. 28, pp. 66–69, January 1996.
- [44] S. C. Few, *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, 2004.
- [45] J. A. Fisteus, L. S. Fernández, and C. D. Kloos, “Applying Model Checking to BPEL4WS Business Collaborations,” in *Proceedings of the 2005 ACM Symposium on Applied Computing*, ser. SAC’05. New York, NY, USA: ACM, 2005, pp. 826–830.
- [46] P. M. Fitts, “The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement,” *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.

- [47] S. Foresti, J. Agutter, Y. Livnat, S. Moon, and R. Erbacher, “Visual Correlation of Network Alerts,” *Computer Graphics and Applications, IEEE*, vol. 26, no. 2, pp. 48–59, March-April 2006.
- [48] J. Fox, N. Johns, C. Lyons, A. Rahmzadeh, R. Thomson, and P. Wilson, “PROforma: A General Technology for Clinical Decision Support Systems,” *Computer Methods and Programs in Biomedicine*, vol. 54, no. 1-2, pp. 59–67, 1997.
- [49] Y. Gil and V. Ratnakar, “Towards Intelligent Assistance for To-Do Lists,” in *Proceedings of the 13th international conference on Intelligent user interfaces*, ser. IUI’08. New York, NY, USA: ACM, 2008, pp. 329–332.
- [50] J. Greenbaum and M. Kyng, Eds., *Design at Work: Cooperative Design of Computer Systems*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 1992.
- [51] “Guidelines for Designing Tables,” <http://understandinggraphics.com/design/data-table-design/>.
- [52] J. Heer, N. Kong, and M. Agrawala, “Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time-Series Visualizations,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’09. New York, NY, USA: ACM, 2009, pp. 1303–1312.
- [53] J. Hickner, D. G. Graham, N. C. Elder, E. Brandt, C. B. Emsermann, S. Dovey, and R. Phillips, “Testing Process Errors and Their Harms and Consequences Reported from Family Medicine Practices: A Study of the American Academy of Family Physicians National Research Network,” *Quality and Safety in Health Care*, vol. 17, no. 3, pp. 194–200, 2008.
- [54] J. M. Hickner, D. H. Fernald, D. M. Harris, E. G. Poon, N. C. Elder, and J. W. Mold, “Issues and Initiatives in the Testing Process in Primary Care

- Physician Offices,” *Joint Commission Journal on Quality and Patient Safety*, vol. 31, no. 2, pp. 81–89, February 2005.
- [55] H. Hochheiser and B. Shneiderman, “Dynamic query tools for time series data sets: Timebox widgets for interactive exploration,” *Information Visualization*, vol. 3, no. 1, pp. 1–18, March 2004.
- [56] J. Horsky, G. D. Schiff, D. Johnston, L. Mercincavage, D. Bell, and B. Middleton, “Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions,” *Journal of Biomedical Informatics*, vol. 45, no. 6, pp. 1202–1216, 2012.
- [57] E. Horvitz, “Principles of Mixed-Initiative User Interfaces,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’99. New York, NY, USA: ACM, 1999, pp. 159–166.
- [58] E. Horvitz, J. Apacible, and M. Subramani, “Balancing Awareness and Interruption: Investigation of Notification Deferral Policies,” in *User Modeling 2005*, ser. Lecture Notes in Computer Science, L. Ardissono, P. Brna, and A. Mitrovic, Eds. Springer Berlin / Heidelberg, 2005, vol. 3538, pp. 433–437.
- [59] G. Hripcsak, P. D. Clayton, R. A. Jenders, J. J. Cimino, and S. B. Johnson, “Design of a Clinical Event Monitor,” *Computers and Biomedical Research*, vol. 29, no. 3, pp. 194–221, June 1996.
- [60] C. J. Hsiao, D. K. Cherry, P. C. Beatty, and E. A. Rechtsteiner, “National Ambulatory Medical Care Survey: 2007 Summary,” National Health, Stat. Report 28, September 2010.
- [61] S. T. Iqbal and B. P. Bailey, “Effects of Intelligent Notification Management on Users and Their Tasks,” in *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, ser. CHI’08. New York, NY, USA: ACM, 2008, pp. 93–102.

- [62] S. T. Iqbal and E. Horvitz, “Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’07. New York, NY, USA: ACM, 2007, pp. 677–686.
- [63] —, “Notifications and Awareness: A Field Study of Alert Usage and Preferences,” in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ser. CSCW’10. New York, NY, USA: ACM, 2010, pp. 27–30.
- [64] E. L. Kaplan and P. Meier, “Nonparametric Estimation from Incomplete Observations,” *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958.
- [65] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Eds., *Mastering The Information Age - Solving Problems with Visual Analytics*. Eurographics, November 2010. [Online]. Available: <http://www.vismaster.eu/book/>
- [66] G. M. Klass, *Just plain data analysis: finding, presenting, and interpreting social science data*. Rowman & Littlefield Pub Inc., 2008.
- [67] J. Kohlhammer and D. Zeltzer, “DCV: A Decision-Centered Visualization System for Time-Critical Applications,” in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, October 2003, pp. 3905–3911.
- [68] —, “Towards a Visualization Architecture for Time-Critical Applications,” in *Proceedings of the 9th international conference on Intelligent user interfaces*, ser. IUI’04. New York, NY, USA: ACM, 2004, pp. 271–273.
- [69] A. Komlodi, J. R. Goodall, and W. G. Lutters, “An Information Visualization Framework for Intrusion Detection,” in *CHI’04 extended abstracts*

- on Human factors in computing systems*, ser. CHI EA'04. New York, NY, USA: ACM, 2004, pp. 1743–1746.
- [70] M. A. Krall and D. F. Sittig, “Subjective Assessment of Usefulness and Appropriate Presentation Mode of Alerts and Reminders in the Outpatient Setting,” in *Proceedings of the American Medical Informatics Association Symposium*, 2001, pp. 334–338.
- [71] S. Kriglstein and S. Rinderle-Ma, “Change Visualizations in Business Processes – Requirements Analysis,” in *GRAPP/IVAPP*, 2012, pp. 584–593.
- [72] M. Krstajic, E. Bertini, and D. Keim, “CloudLines: Compact Display of Event Episodes in Multiple Time-Series,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2432–2439, December 2011.
- [73] G. J. Kuperman, J. M. Teich, D. W. Bates, F. L. Hiltz, J. M. Hurley, R. Y. Lee, and M. D. Paterno, “Detecting Alerts, Notifying the Physician, and Offering Action Items: A Comprehensive Alerting System,” in *Proceedings of the American Medical Informatics Association Annual Fall Symposium*, 1996, pp. 704–708.
- [74] H. Lam, A. E. Kirkpatrick, J. Dill, and M. S. Atkins, “Effective Display of Medical Laboratory Report Results on Small Screens: Evaluation of Linear and Hierarchical Displays,” *International Journal of Human-Computer Interaction*, vol. 21, no. 1, pp. 73–89, 2006.
- [75] S. Lauesen, M. Kuhail, K. Pantazos, S. Xu, and M. B. Andersen, “Extending Applications with Visualization,” Submitted to InfoVis 2012, 2012.
- [76] J. Lin, E. Keogh, and S. Lonardi, “Visualizing and Discovering Non-Trivial Patterns in Large Time Series Databases,” *Information Visualization*, vol. 4, no. 2, pp. 61–82, July 2005.

- [77] S. Ling and H. Schmidt, "Time Petri Nets for Workflow Modelling and Analysis," in *2000 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 4, 2000, pp. 3039–3044.
- [78] X. Luo, F. Tian, W. Liu, D. Teng, G. Dai, and H. Wang, "Visualizing time-series data in processlines: Design and evaluation of a process enterprise application," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC'10. New York, NY, USA: ACM, 2010, pp. 1165–1172.
- [79] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker, "Process mining in healthcare – a case study," in *Proceedings of the First International Conference on Health Informatics*, 2008, pp. 118–125.
- [80] A. C. Marcus, C. P. Kaplan, L. A. Crane, J. S. Berek, G. Bernstein, J. E. Gunning, and M. W. McClatchey, "Reducing Loss-to-Follow-up among Women with Abnormal Pap smears: Results from a Randomized Trial Testing an Intensive Follow-up Protocol and Economic Incentives," *Medical Care*, vol. 36, no. 3, pp. 397–410, 1998.
- [81] M. L. Matthews, L. Rehak, A.-L. Lapinski, and S. McFadden, "Improving The Maritime Surface Picture With A Visualization Aid To Provide Rapid Situation Awareness Of Information Uncertainty," in *IEEE Toronto International Conference – Science and Technology for Humanity*, ser. TIC-STH, September 2009, pp. 533–538.
- [82] T. R. McEwen, N. C. Elder, and J. M. Flach, "Creating Safety in Primary Care Practice with Electronic Medical Records Requires the Consideration of System Dynamics," *Journal of Healthcare Engineering*, vol. 2, no. 1, pp. 87–96, 2011.

- [83] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, “LiveRAC: Interactive Visual Exploration of System Management Time-Series Data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’08. New York, NY, USA: ACM, 2008, pp. 1483–1492.
- [84] J. McNames, D. Abercrombie, D. Turner, and T. Shannon, “Semiconductor Manufacturing Process Visualization,” in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 5, October 2003, pp. 4940–4944.
- [85] “OpenVista CIS,” <http://medsphere.org/community/project/openvista-cis>, Medsphere, accessed: 01/01/2013.
- [86] *Design Guidance – Displaying Graphs and Tables*, Microsoft, 2008. [Online]. Available: <http://www.mscti.net/DesignGuide/TablesGraphsDisplay.aspx>
- [87] *Windows Vista User Experience Guidelines*, Microsoft, Inc., 2010. [Online]. Available: <http://msdn.microsoft.com/en-us/library/aa511485.aspx>
- [88] C. Mikkelsen, J. Johansson, and M. Rissanen, “Interactive Information Visualization for Sensemaking in Power Grid Supervisory Systems,” in *15th International Conference on Information Visualisation (IV)*, July 2011, pp. 119–126.
- [89] J. W. Mold, D. S. Cacy, and D. K. Dalbir, “Management of Laboratory Test Results in Family Practice: An OKPRN Study,” *Journal of Family Practice*, vol. 49, no. 8, pp. 716–717, August 2000.
- [90] H. J. Murff, T. K. Gandhi, A. K. Karson, E. A. Mort, E. G. Poon, S. J. Wang, D. G. Fairchild, and D. W. Bates, “Primary Care Physician Attitudes Concerning Follow-up of Abnormal Test Results and Ambulatory Decision Support Systems,” *International Journal of Medical Informatics*, vol. 71, no. 2–3, pp. 137–149, 2003.

- [91] *Research-based Web Design and Usability Guidelines*, National Cancer Institute, Department of Health & Human Services, 2006. [Online]. Available: http://usability.gov/guidelines/guidelines_book.pdf
- [92] R. Noumeir, “Radiology Interpretation Process Modeling,” *Journal of Biomedical Informatics*, vol. 39, no. 2, pp. 103–114, April 2006.
- [93] M. I. Oppenheim, R. J. Mintz, A. G. Boyer, and W. W. Frayer, “Design of a Clinical Alert System to Facilitate Development, Testing, Maintenance, and User-Specific Notification,” in *Proceedings of the American Medical Informatics Association Symposium*, 2000, pp. 630–634.
- [94] *Java Look and Feel Design Guidelines*, Oracle, 2001. [Online]. Available: <http://www.oracle.com/technetwork/java/hig-137787.html#1010484>
- [95] L. Palen, “Social, Individual and Technological Issues for Groupware Calendar Systems,” in *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, ser. CHI’99. New York, NY, USA: ACM, 1999, pp. 17–24.
- [96] K. Pantazos, S. Tarkan, C. Plaisant, and B. Shneiderman, “Exposing Delays in Multi-Step Processes by Retrospective Analysis,” HCIL, UMD, Tech. Rep. 27, September 2012.
- [97] —, “Promoting Timely Completion of Multi-Step Processes – A Visual Approach to Retrospective Analysis,” HCIL, UMD, Tech. Rep. 9, April 2012.
- [98] “Ultimate Guide to Table UI Patterns,” <http://www.jankoatwarpspeed.com/post/2010/02/26/table-ui-patterns.aspx>.
- [99] T. H. Payne and J. Savarino, “Development of a Clinical Event Monitor for Use with the VA Computerized Patient Record System and Other Data Sources,” in *Proceedings of the American Medical Informatics Association Symposium*, 1998, pp. 145–149.

- [100] M. Peleg, A. A. Boxwala, O. Ogunyemi, Q. Zeng, S. Tu, R. Lacson, E. Bernstam, N. Ash, P. Mork, L. Ohno-Machado, E. H. Shortliffe, and R. A. Greenes, “GLIF3: The Evolution of a Guideline Representation Format,” in *Proceedings of the American Medical Informatics Association Symposium*, 2000, pp. 645–649.
- [101] B. Perez and I. Porres, “Verification of Clinical Guidelines by Model Checking,” in *Proceedings of the 21st IEEE International Symposium on Computer-Based Medical Systems*, ser. CBMS’08, June 2008, pp. 114–119.
- [102] J. L. Peterson, *Petri Net Theory and the Modeling of Systems*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1981.
- [103] M. S. Pfaff, J. L. Drury, G. L. Klein, L. More, S. P. Moon, and Y. Liu, “Weighing Decisions: Aiding Emergency Response Decision Making via Option Awareness,” in *Technologies for Homeland Security (HST), 2010 IEEE International Conference on*, November 2010, pp. 251–257.
- [104] R. L. Phillips, L. A. Bartholomew, S. M. Dovey, G. E. Fryer, T. J. Miyoshi, and L. A. Green, “Learning from Malpractice Claims about Negligent, Adverse Events in Primary Care in the United States,” *Quality and Safety in Health Care*, vol. 13, no. 2, pp. 121–126, 2004.
- [105] C. Plaisant, D. Heller, J. Li, B. Shneiderman, R. Mushlin, and J. Karat, “Visualizing Medical Records with LifeLines,” in *CHI 98 conference summary on Human factors in computing systems*, ser. CHI’98. New York, NY, USA: ACM, 1998, pp. 28–29.
- [106] C. Plaisant, R. Levy, and W. Zhao, “BizView: Managing Business and Network Alarms,” HCIL, UMD, Technical Report 22, November 1995.
- [107] E. G. Poon, T. K. Gandhi, T. D. Sequist, H. J. Murff, A. S. Karson, and D. W. Bates, ““I Wish I Had Seen This Test Result Earlier!”: Dissatisfac-

- tion with Test Result Management Systems in Primary Care,” *Archives of Internal Medicine*, vol. 164, no. 20, pp. 2223–2228, 2004.
- [108] E. G. Poon, J. S. Haas, A. L. Puopolo, T. K. Gandhi, E. Burdick, D. Bates, and T. A. Brennan, “Communication Factors in the Follow-up of Abnormal Mammograms,” *Journal of General Internal Medicine*, vol. 19, no. 4, pp. 316–323, 2004.
- [109] E. G. Poon, G. J. Kuperman, J. Fiskio, and D. W. Bates, “Real-time Notification of Laboratory Data Requested by Users through Alphanumeric Pagers,” *JAMIA: Journal of the American Medical Informatics Association*, vol. 9, no. 3, pp. 217–222, 2002.
- [110] E. G. Poon, S. J. Wang, T. K. Gandhi, D. W. Bates, and G. J. Kuperman, “Design and Implementation of a Comprehensive Outpatient Results Manager,” *Journal of Biomedical Informatics*, vol. 36, no. 1–2, pp. 80–91, 2003.
- [111] A. Pretorius and J. van Wijk, “Multiple Views on System Traces,” in *IEEE Pacific Visualization Symposium 2008*, March 2008, pp. 95–102.
- [112] M. Riveiro and G. Falkman, “Evaluating the Usability of Visualizations of Normal Behavioral Models for Analytical Reasoning,” in *Seventh International Conference on Computer Graphics, Imaging and Visualization (CGIV)*, August 2010, pp. 179–185.
- [113] G. G. Robertson, T. Chilimbi, and B. Lee, “AllocRay: Memory Allocation Visualization for Unmanaged Languages,” in *Proceedings of the 5th International Symposium on Software visualization*, ser. SOFTVIS’10. New York, NY, USA: ACM, 2010, pp. 43–52.
- [114] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human - Computer Interaction*, 3rd ed. Wiley Publishing, 2011.

- [115] C. L. Roy, E. G. Poon, A. S. Karson, Z. Ladak-Merchant, R. E. Johnson, S. M. Maviglia, and T. K. Gandhi, “Patient Safety Concerns Arising from Test Results that Return after Hospital Discharge,” *Annals of Internal Medicine*, vol. 143, no. 2, pp. 121–128, 2005.
- [116] G. D. Schiff, S. Kim, N. Krosnjar, M. F. Wisniewski, J. Bult, L. Fogelfeld, and R. A. McNutt, “Missed Hypothyroidism Diagnosis Uncovered by Linking Laboratory and Pharmacy Data,” *Archives of Internal Medicine*, vol. 165, pp. 574–577, 2005.
- [117] M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz, “Cardiogram: Visual Analytics for Automotive Engineers,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’11. New York, NY, USA: ACM, 2011, pp. 1727–1736.
- [118] M. M. Shabot, M. LoBue, and J. Chen, “Wireless Clinical Alerts for Physiologic, Laboratory and Medication Data,” in *Proceedings of the American Medical Informatics Association Symposium*, 2000, pp. 789–793.
- [119] Y. Shahar, S. Miksch, and P. Johnson, “The ASGAARD Project: A Task-specific Framework for Application, Critiquing of Time-oriented Clinical Guidelines,” *Artificial Intelligence in Medicine*, vol. 14, no. 1-2, pp. 29–51, 1998.
- [120] B. Shneiderman, “The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations,” in *Proceedings of the IEEE Symposium on Visual Languages*, September 1996, pp. 336–343.
- [121] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 5th ed. Boston, MA, USA: Addison-Wesley, 2010.

- [122] H. Singh, L. A. Petersen, and E. J. Thomas, "Understanding Diagnostic Errors in Medicine: A Lesson from Aviation," *Quality and Safety in Health Care*, vol. 15, pp. 159–164, 2006.
- [123] H. Singh, E. J. Thomas, S. Mani, D. Sittig, H. Arora, D. Espadas, M. M. Khan, and L. A. Petersen, "Timely Follow-up of Abnormal Diagnostic Imaging Test Results in an Outpatient Setting: Are Electronic Medical Records Achieving Their Potential?" *Archives of Internal Medicine*, vol. 169, no. 17, pp. 1578–1586, 2009.
- [124] H. Singh, E. J. Thomas, D. F. Sittig, L. Wilson, D. Espadas, M. M. Khan, and L. A. Petersen, "Notification of Abnormal Lab Test Results in An Electronic Medical Record: Do Any Safety Concerns Remain?" *The American Journal of Medicine*, vol. 123, no. 3, pp. 238–244, March 2010.
- [125] M. W. Smith, D. R. Murphy, A. Laxmisan, B. Reis, D. F. Sittig, and H. Singh, "Developing Reliable Systems for Tracking and Follow-up of Abnormal Test Results in Electronic Health Records," 2011, demo Oral Abstract Presentation at Society for Medical Decision Making. [Online]. Available: <https://smdm.confex.com/smdm/2011ch/webprogram/Session1467.html>
- [126] P. C. Smith, R. Araya-Guerra, C. Bublitz, B. Parnes, L. M. Dickinson, R. van Vorst, J. M. Westfall, and W. D. Pace, "Missing Clinical Information During Primary Care Visits," *JAMA: The Journal of the American Medical Association*, vol. 293, no. 5, pp. 565–571, 2005.
- [127] D. H. Solomon, H. Hashimoto, L. Daltroy, and M. H. Liang, "Techniques to improve physicians' use of diagnostic tests: A new conceptual framework." *The Journal Of The American Medical Association (JAMA)*, vol. 280, no. 23, pp. 2020–7, 1998.

- [128] R. Stabina, “Quantitative Data Graphics: Best Practices of Designing Tables and Graphs for Use in Not-for-Profit Evaluation Reports,” Master’s thesis, University of Oregon, June 2005.
- [129] S. J. Steindel and P. J. Howanitz, “Physician Satisfaction and Emergency Department Laboratory Test Turnaround Time Observations Based on College of American Pathologists Q-Probes Studies,” *Archives of Pathology*, vol. 125, no. 7, pp. 863–871, July 2001.
- [130] S. Sung, V. Forman-Hoffman, M. C. Wilson, and P. Cram, “Direct Reporting of Laboratory Test Results to Patients by Mail to Enhance Patient Safety,” *Journal of General Internal Medicine*, vol. 21, no. 10, pp. 1075–1078, 2006.
- [131] “Designing Tables 101,” <http://uxmatters.com/mt/archives/2009/09/designing-tables-101.php>.
- [132] P. C. Tang, D. Fafchamps, and E. H. Shortliffe, “Traditional Medical Records as a Source of Clinical Data in the Outpatient Setting,” in *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, 1994, pp. 575–579.
- [133] S. Tarkan, L. Franklin, C. Plaisant, and B. Shneiderman, “Design Guidelines for Ensuring Timely Management of Medical Orders,” HCIL, UMD, Tech. Rep. 25, September 2012.
- [134] S. Tarkan, C. Plaisant, B. Shneiderman, and A. Z. Hettinger, “Ensuring Timely Clinical Lab Test Result Management: A Generative XML Process Model to Support Medical Care,” HCIL, UMD, Tech. Rep. 15, June 2011.
- [135] —, “Reducing Missed Laboratory Results: Defining Temporal Responsibility, Generating User Interfaces for Test Process Tracking, and Retrospective Analyses to Identify Problems,” in *Proceedings of the AMIA Symposium*, Washington, DC, USA, 2011, pp. 1382–1391.

- [136] A. ten Teije, M. Marcos, M. Balsler, J. van Croonenborg, C. Duelli, F. van Harmelen, P. Lucas, S. Miksch, W. Reif, K. Rosenbrand, and A. Seyfang, “Improving Medical Protocols by Formal Methods,” *Artificial Intelligence in Medicine*, vol. 36, pp. 193–209, March 2006.
- [137] P. Terenziani, L. Giordano, A. Bottrighi, S. Montani, and L. Donzella, “Model Checking for Clinical Guidelines: An Agent-based Approach,” in *Proceedings of the American Medical Informatics Association Annual Symposium*, 2006, pp. 289–293.
- [138] R. S. Thompson, E. M. Rantanen, W. Yurcik, and B. P. Bailey, “Command Line or Pretty Lines?: Comparing Textual and Visual Interfaces for Intrusion Detection,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’07. New York, NY, USA: ACM, 2007, pp. 1205–.
- [139] “15 Tips for Designing Terrific Tables,” <http://designshack.net/articles/css/15-tips-for-designing-terrific-tables>.
- [140] J. Training, “Allscripts Video,” <http://www.youtube.com/watch?v=kHN1Mu3T0R0>, accessed: 01/01/2013.
- [141] T. Tullis and S. Fleischman, “Tabular Data: Finding the Best Format,” Intercom, June 2004.
- [142] “Ultimate Guide to Table UI Patterns,” <http://designingwebinterfaces.com/ultimate-guide-to-table-ui-patterns>.
- [143] W. M. P. van der Aalst and A. H. M. ter Hofstede, “YAWL: Yet Another Workflow Language,” *Information Systems*, vol. 30, no. 4, pp. 245–275, 2005.
- [144] W. M. P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, 1st ed. Springer Publishing Company, Incorporated, 2011.

- [145] C. van Walraven and C. D. Naylor, "Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits." *The Journal of the American Medical Association*, vol. 280, no. 6, pp. 550–558, 1998.
- [146] M. van Wijk, M. Mosseveld, and J. van der Lei, "Design of a decision support system for test ordering in general practice: Choices and decisions to make." *Methods of Information in Medicine*, vol. 38, no. 4-5, pp. 355–61, 1999.
- [147] M. M. Wagner, M. Pankaskie, W. Hogan, F. C. Tsui, S. A. Eisenstadt, E. Rodriguez, and J. K. Vries, "Clinical Event Monitoring at the University of Pittsburgh," in *Proceedings of the American Medical Informatics Association Symposium*, 1997, pp. 188–192.
- [148] T. Wahls, "Diagnostic Errors and Abnormal Diagnostic Tests Lost to Follow-Up: A Source of Needless Waste and Delay to Treatment," *The Journal of Ambulatory Care Management*, vol. 30, no. 4, pp. 338–343, 2007.
- [149] T. Wahls, T. Haugen, and P. Cram, "The Continuing Problem of Missed Test Results in an Integrated Health System with an Advanced EMR," *Joint Commission Journal on Quality and Patient Safety*, vol. 33, no. 8, pp. 485–492, 2007.
- [150] T. L. Wahls and P. M. Cram, "The Frequency of Missed Test Results and Associated Treatment Delays in A Highly Computerized Health System," *BMC Family Practice*, vol. 8, no. 1, pp. 32–39, 2007.
- [151] C. V. Walraven, V. Goel, and B. Chan, "Effect of population-based interventions on laboratory utilization: A time-series analysis." *The Journal Of The American Medical Association*, vol. 280, no. 23, pp. 2028–2033, 1998.
- [152] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning Temporal Data by Sentinel Events: Discovering Patterns

- in Electronic Health Records,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’08. New York, NY, USA: ACM, 2008, pp. 457–466.
- [153] C. Ware, *Information Visualization, Second Edition: Perception for Design (Interactive Technologies)*, 2nd ed. Morgan Kaufmann, April 2004.
- [154] M. Weber, M. Alexa, and W. Muller, “Visualizing Time-series on Spirals,” in *IEEE Symposium on Information Visualization 2001*, 2001, pp. 7–13.
- [155] WFMC, “Workflow Management Coalition Workflow Standard: Workflow Process Definition Interface – XML Process Definition Language (XPDL),” Workflow Management Coalition, Lighthouse Point, FL, USA, Technical Report WFMC-TC-1025, 2002.
- [156] S. A. White, “Business Process Modeling Notation (BPMN),” Object Management Group, Specification, 2004, version 1.0.
- [157] S. Whittaker and C. Sidner, “Email Overload: Exploring Personal Information Management of Email,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ser. CHI’96. New York, NY, USA: ACM, 1996, pp. 276–283.
- [158] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, “Lifeflow: visualizing an overview of event sequences,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI’11. New York, NY, USA: ACM, 2011, pp. 1747–1756.
- [159] T. R. Yackel and P. J. Embi, “Unintended Errors with EHR-based Result Management: A Case Series,” *JAMIA: Journal of the American Medical Informatics Association*, vol. 17, no. 1, pp. 104–107, 2010.

- [160] J. Zhao, F. Chevalier, and R. Balakrishnan, “KronoMiner: Using Multi-Foci Navigation for the Visual Exploration of Time-Series Data,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 1737–1746.
- [161] M. H. Zweig and G. Campbell, “Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine,” *Clinical Chemistry*, vol. 39, no. 4, pp. 561–577, 1993.