Insights from a User-Centered Approach to Computerized Guidelines for Chronic Disease

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Barbara Jessica Shaten

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Karen M. Kuntz

June 2011

Acknowledgements

This Dissertation describes findings that emerged from the *HIT Asthma* project, funded by the Agency for Health Research and Quality (AHRQ), Contract # HHSA 290-06-0020, Task Order No. 5. The prime contractor was Denver Health and Hospital Systems; Minneapolis Medical Research Foundation – the research wing of Hennepin Faculty Associates, the physician practice plan serving Hennepin County Medical Center (HCMC) – was the primary subcontractor.

Dr. Scott Davies, Chief of Medicine at HCMC, created the context that made the emergence of the project possible. As HCMC made the decision to adopt an enterprise-wide commercially electronic health record (EHR) system, he anticipated that the data emerging from such a system could be used to improve the care and outcomes of patients with chronic disease, which disproportionately afflicted the urban poor patient population that HCMC served. He used his position at HCMC and political influence within the community to create the Center for Urban Health, charged to develop ways to make this happen. As the Research Director for the Center for Urban Health, I became aware of AHRQ ACTION contracts – funding mechanisms to test promising options for improved health services delivery within the delivery organizations themselves. By anticipating the promise of EHRs, and working to establish an infrastructure charged to manifest that promise, Dr. Davies made it possible for me to know about and to respond to the request for proposals (RFP) released by AHRQ, seeking demonstrations of health information technology to improve quality.

Dr. Kevin Larsen, Chief Medical Informatics Officer at HCMC, worked with me to manifest the vision that Dr. Davies had. When I told him about the AHRQ RFP, Kevin directed attention to asthma, because he knew there was a quality improvement initiative in asthma already underway, with a strong clinical champion, who already had encountered the limits of the enterprise-wide EHR system for supporting clinical decision making. The system superbly integrated information across the enterprise, but was only marginally successful at supporting complex clinical reasoning in chronic disease. With respect to both technology and electronic information management, this project was conceptually many steps ahead of the organization's state of thinking and development, posing challenges to both. Once the project was underway, Dr. Larsen established the necessary working relationships with three distinct

technology support teams at HCMC – the EHR implementation team, the team that managed servers and databases, and the team that managed and extracted data from the EHR system's analytic database.

Dr. Gail Brottman, Director of Pediatric Pulmonary Medicine at HCMC, was the asthma clinical champion. A demanding "user", she refused to accept anything less than what would give her patients the best asthma care. She was steadfast in her vision of the tool's functioning, and insisted on the features the computerized asthma decision tool had to have if physicians were going to use it. She scoured the asthma guidelines searching for answers to questions that I and the software developers had. She was the tool's inspiration; she led the effort to introduce it into HCMC, and demonstrated it to prospective users across Minnesota. She was the reason the eAAP (Electronic Asthma Action Plan) came to exist.

Dr. Angeline Carlson and Cherylee Sherry constructed the medication database for the eAAP. Starting with a simple Excel spreadsheet containing 300 rows, they kept adding as they painstakingly operationalized the abstract language in the EPR-3 (Third Expert Panel Report for the Diagnosis and Management of Asthma)[1] used to recommend asthma daily controller medications. The spreadsheet eventually grew to more than 35,000 rows, a graphic demonstration of how much operational detail went unmentioned in the guidelines. Dr. Donald Uden, Professor of Pharmacy at the University of Minnesota, helped resolve some of the ethical quandaries that Drs. Carlson and Brottman faced when operationalizing the abstract guideline language led to medication choices they felt were clinically questionable. Touch Thouk kept the office humming during these periods of high stress.

Faith Dohman, RN, nurse educator at Hennepin Faculty Associates, worked with patient focus groups to identify the best language and visual features of an effective asthma action plan. Jennifer Rodlund, the RN at HFA charged with clinical quality, convened a group of nurse colleagues to manually review patients' charts to determine which ones had paper asthma action plans scanned into their records. They performed these reviews voluntarily, without pay, because they were committed to improving patient care and they believed me when I told them that I needed the information to understand the impact of the eAAP on improving the rates with which their patients had asthma action plans. Dr. Brendon Cullinan, Director of Ambulatory Care Quality at HCMC, helped me devise a sampling strategy that would meet

both research and quality objectives, and was prepared to defend the use of nurse time for chart reviews if necessary.

Dr. Robert Grundmeier at the Children's Hospital of Philadelphia gave us the trick that enabled a simple invocation of the eAAP from local electronic health record system, with live data transfers from the patient's record into the tool. Without that trick we would have been forced to rely upon a stale database, with numerous performance issues that would have significantly jeopardized user acceptance.

Tim Michalski, president of Lighthouse Software Solutions, gave many more hours of himself and his staff than we had contracted for to write the code for the tool, because he was so excited by its promise. There were many midnight phone calls, after he had put his young children to bed, when he and I talked through the latest set of logical and user interface challenges. He patiently taught me how to use the wiki-based interface to document issues through "ticketing", and then patiently explained that not every ticket I placed could really be closed. He nearly single-handedly kept the eAAP alive once the contract ended with AHRQ in February 2010.

Mary Ann Jagodzinski, RN, allowed herself to be brought back from retirement so that she could wield her considerable authority among clinicians at HCMC to encourage them to use the eAAP. "If I can do this, you can do this," she would say to the physicians expressing resistance to new technology. She was more of a technological novice than they were, but it was the right thing to do, and it wasn't hard.

Mary Ann Kastorff, who had been trained in the use of the EHR system's analytic database, spent hours helping me understand its structure, contents, and sometimes whimsical naming conventions, so that I could frame sensible queries to extract data from it. Kerri Nelson, in Knowledge Management at HCMC, spent even more hours in explanation, and executed those queries. My friend and long-time colleague Ken Svendsen was always available on the other end of the phone when I was stymied yet again about how to do something in SAS. Glenn Trygstad helped configure the SAS datasets formed by merging data extracted from the EHR with data captured in the eAAP database. This was a not insignificant task, as any anyone familiar with the volume and complexity of EHR data can attest.

It is customary to briefly acknowledge project funders, but in this situation, AHRQ deserves more than a nod. Dr. Jon White, who managed the Health Information Technology portfolio at AHRQ, had the courage and audacity to fund a technology development project led by two physicians, a pharmacist, and a PhD-seeking health services researcher, none of whom had ever developed software before. Approximately six months after project inception, when it looked like we most certainly would fail, Dr. White revealed that he also had funded an experienced informaticist at Yale University who had the same objectives as we: Build a computerized decision support tool for asthma based on the EPR-3. Dr. White was curious to know what differences would emerge from projects with identical goals, one guided by an informatics (e.g. guideline-centric) approach, and the other guided by a physician (e.g. user-centric) approach. This was my first realization that the question was worth asking, and thus Dr. White at AHRQ originally posed the central question from which this Dissertation emerged.

Bob Mayes, our project officer at AHRQ, brought us back from the brink of failure numerous times. Any sensible project officer would have cancelled the contract each time we told him the latest set of challenges we faced. He did not. The challenges were more interesting to him that the solutions to them; it told him what physicians actually needed from both guidelines and technology to support the work they did.

A version of Chapter II (*Physicians, Guidelines and Cognitive Tasks),* was previously published in Evaluation and the Health Professions[2]. Dr. Steve Sussman, editor of the journal, offered much encouragement and useful advice. The article was derived from a report we had submitted to AHRQ, which itemized – at AHRQ request – the gaps between the support that working clinicians required and the structure and content of the EPR-3. Dr. Brottman and Dr. Carlson both contributed to the report, which I wrote. Contents of the report not presented in the Dissertation can be found in Appendix A.

Dr. Brottman and Dr. Carlson contributed to portions of Chapter III (*Effect of a Computerized Decision Support Tool on the Proportion of Asthma Patients with Asthma Action Plan*). Dr. Brottman represented the clinical perspective, and the importance of asthma action plans, while I represented the methodological perspective, and the promise of interrupted time series analysis for evaluating the impact of interventions on the care and outcomes of

chronic disease populations. Dr. Carlson navigated the differences in perspectives and helped me attain a balanced result.

Dr. Carlson served in multiple roles: project pharmacist, project evaluator, mediator between clinical and methodological points of views, and finally, as an external reviewer on my Committee. Other Committee members were Dr. Karen Kuntz, Dr. Beth Virnig, Dr. Douglas Wholey and Dr. Francois Sainfort, all from the Division of Health Policy and Management, within the School of Public Health at the University of Minnesota.

Karen Kuntz, my advisor, was willing to serve in this role though she barely knew me and the topic was only tangentially related to her field. Her calm and deliberate manner was a useful antidote to my frequent lapses into rhetoric and hyperbole. Beth Virnig served as the committee member representing the Ethics perspective, and offered helpful encouragement and suggestions for Chapter IV (*Ethical Issues in Computerized Decision Support for Chronic Disease Management*). She pushed me to write more crisply by pointing out that readers would more easily grasp the ideas if they didn't have to search for them in the midst of too many words. Doug Wholey graciously continued to serve on my Committee as this project inexorably encroached on the one that I originally had proposed as a thesis topic, one in which he had a passionate interest of his own. I could not write two completely different Dissertations at once, while keeping a full time job; something had to give and it eventually was the original topic. Dr. Wholey stayed though he could have gone. Francois Sainfort also graciously agreed to serve as a Committee member, and chaired the Dissertation defense.

I am grateful for the trust you've placed in my ability, and I hope I haven't disappointed it.

A note on names: My legal name and my name of registration with the University of Minnesota is Barbara Jessica Shaten, but I have published and worked under the name of Yiscah Bracha since 1998. This dissertation should be conceived as the work of Yiscah Bracha.

To my friends and family, thank you for your patience all these years as I declined one social invitation after another, in order to keep my focus on completing this PhD. I look forward to rejoining you and I hope you still accept me.

All flaws, oversights, excesses and failures remain completely my own.

To Loren.

ABSTRACT

For more than two decades, the medical informatics community has worked towards representing evidence-based guidelines in computer code, intended to be executed at the point of care. The purpose is to close the gap between evidence of best medical practices and the care that patients receive. Most informatics work has taken a "guideline-centric" approach, focusing primarily on the guidelines rather than the physicians who are the intended users.

The *HIT Asthma* project took a "user-centric" approach towards developing, implementing and evaluating the effects of a computerized decision support (CDS) tool for ambulatory asthma care. The user-centricity yielded findings that questioned the ability of guidelines to support medical work, raised ethical concerns, and challenged the epistemic foundations of the evidence which the guidelines are intended to impart. The approach also demonstrated how CDS tools for chronic disease could become prototypical technologies to support a "Healthcare System that Learns", a vision promulgated by the Institute of Medicine Roundtable on Evidence-Based Medicine.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER I. INTRODUCTION

I.A. THE INFORMATICS RATIONALE FOR COMPUTERIZED DECISION SUPPORT

For more than two decades, the medical informatics community has pursued an agenda that brings evidence-based guidelines to the point of care through computerized decision support (CDS) tools [3-7]. The purpose is to close the well-documented gap between evidence of best medical practices, and the care that patients actually receive [8-10]. The theoretical rationale is that physicians make decisions based on proximal information that is quick and easy to use [11]. During a time-pressured encounter, physicians will not search through guidelines transmitted through lengthy narratives; computer technology puts information at their fingertips, anticipating the questions they are likely to have.

Although some are aware that physicians will adopt technology only if it meets their needs[4], the CDS informatics development approach is predominantly "guideline-centric": At the center of the enterprise to build and implement a CDS tool are the practice guidelines and the recommended physician behaviors they contain [12]. An alternative approach is "user-centric": At the center of the enterprise to build a CDS tool is the physician user who will be deploying it. Empirical findings from the informatics literature, as well as theories developed in the fields of human factors and cognitive ergonomics, suggest that a user-centric approach is essential if users are to deploy the technologies that are being created for them.

This dissertation describes findings and insights that emerged from a user-centric approach to creating, implementing and evaluating the effect of a computerized decision support tool for ambulatory asthma care. As predicted by theory, the technology was quickly adopted by a subset of physician-users at the implementation site, and other physicians at different sites expressed strong interest in it, exhibiting an appetite for uptake not generally found for technologies that are developed without strong user involvement.

User-centricity, however, led to unexpected insights of greater significance. Because physician-users were intimately involved in the tool's development, they became acutely aware of the inadequacy of the evidence upon which the guidelines were based. This awareness then created instances in which clinician/developers faced conflicts between their respect for the authority of the guidelines and the guideline developers, and their ethical obligations to patients. Although it was possible to resolve the ethical issues by using the CDS tool as a way to generate practice-based evidence, the contemporary "evidence hierarchy" in evidence-based medicine assigns greater epistemic value to evidence from randomized controlled trials (RCTs) [13]. This relative value system creates social pressure to use CDS tools for chronic disease simply as a way to disseminate evidence generated predominantly by RCTs, and discourages their use for generating evidence from practice. However, generating evidence from practice, and incorporating results directly into new versions of the tool, is exactly what was recommended by an Institute of Medicine (IOM) Roundtable on Evidence-Based Medicine, when it asserted that the nation needs a healthcare system that learns. [14].

I.B. EVALUATIONS OF THE IMPACT OF CDS AND THE ROLE OF USER ACCEPTANCE

I.B.1. Overview of computerized decision support (CDS)

CDS tools range widely in complexity. The most complex type, under consideration here, guides users through decision sequences that contain multiple contingencies and branching points, with decisions based on patients' responses to prior therapy. Seroussia et al [15] calls this "guided mode" support, to differentiate it from the simpler and more common "critiquing modes", often delivered through enterprise-wide electronic health record (EHR) systems.

Before considering complex guided support in more depth, it is useful to describe critiquing support, to understand the differences. Users experience the most basic level of critiquing support as reminders that pop up on screen; the mechanism employs data-driven algorithms built into the EHR, which "fire" automatically when a set of logical rules are satisfied. A more advanced level of critiquing support (often going by names such as "order sets", or "computerized physician order entry") gives users integrated sets of ordering

selections, and fires an alert if the user makes a selection that is contraindicated by existing data, or by pre-specified ranges [16, 17].

Alerts and reminders have been shown improve rates of performing single recommended actions for patients at risk, such as immunizing a child [18], ordering preventive services [19, 20], and reducing medication errors [21]. Computerized order entry systems have potential to reduce overuse of expensive diagnostic services [4] and to achieve consistency and standardization in ordering across the medical system that uses them.

While useful, these reminders, alerts, and order sets do not provide the kind of evidence-based guided decision support that physicians need to manage a chronic disease. Several common clinical scenarios illustrate the differences: A reminder system may alert a physician if an asthma patient is due for an updated asthma action plan; an order set may direct the physician towards ordering a daily controller medication; a warning may appear if the daily controller the physician orders is contraindicated by another medication order for that patient. But none of these systems will help the physician decide which daily controller medication to order and at what dose, given the patient's level of asthma control that has been achieved with current therapy. While a pulmonary specialist will likely be able to make these decisions, a primary care physician who treats multiple conditions may require knowledge support.

I.B.2. Complex CDS for managing chronic disease

An active medical informatics subfield has been developing guideline-based CDS tools for chronic disease for more than twenty years [3, 5, 22]. There are thousands of guidelines, with the number growing every year, and many informaticists have dedicated careers towards creating artificial languages intended to automatically convert guideline narratives into executable code [12, 23]. In a vision first articulated in 1982 [24], one day the codes will be broadcasted from central repositories and downloaded into the EHRs at local delivery sites.

I.B.3. The impact of CDS for chronic disease, and the imperative of user centricity

In 2009, Heselmans et al [25] performed a systematic literature review to determine the degree to which implementation of CDS tools for chronic disease is associated with clinicians' adherence to guideline recommendations, and/or patient outcomes. Unlike other systematic reviews which asked similar questions [26-29], but did not differentiate between simple critiquing vs. complex guided support, Heselmans et al only reviewed studies of complex systems supporting chronic disease, which were introduced into ambulatory care through existing electronic health record systems. They identified twenty-seven relevant studies, some of which compared "computerized guidelines" to their paper equivalents, and some of which compared "computerized guidelines" to "usual care" of no guideline implementation at all.

For adherence to guidelines, more than half the studies showed no difference between computerized guidelines and usual care, and for all reviewed studies, there were no differences between computerized and paper guidelines. Of the few studies that evaluated patient outcomes, none showed differences between computerized guidelines and either comparison group. Heselmans et al concluded that "There is little evidence at the moment for the effectiveness of an increasingly used and commercialised instrument such as electronic multidimensional guidelines".

The conclusion may be premature. Upon inspection, the absence of effect was almost always due to users rejecting the technology. Users complained that the timing of the system triggers was poor, that the systems were difficult to use, that the content was not relevant to delivering care [27], or that using the system required more time than physicians had available [22,30]. Quaglini et al found that physicians needed both administrative and decision support; if the system did not provide both, they perceived it as useless and ignored it all, including the guideline-based decision support [31]. Evaluations of the effectiveness of CDS tools thus

cannot be separated from evaluations of their usability, because if clinicians don't use them, questions about effectiveness are moot.

Requirements for usability and thus acceptance fall into three major categories. The first requirement is that the system be integrated into usual workflows [4, 6, 22, 28, 31-33]. When the normal workflow is directed by an enterprise electronic health record system, the software providing the support must either be part of the system, or reachable from it [4, 22]. A second requirement is that the system must offer content physicians find valuable [27]. In chronic disease management, content is specific to the patient's location on the disease trajectory. Bates et al identify this ability to identify patient status a vitally important frontier [32]. A third requirement is that the system support multiple functions, including administrative ones [31]. The user must invest both time and effort to invoke the system and follow its guidance. Given the volume of activities that physicians are required to perform, if they are unable to achieve multiple goals, the likelihood that they will seek only the decision support diminishes. Users need to experience the system as a "workflow manager" [34, 35] in addition to being a "knowledge communicator" [4].

To ensure that users accept a CDS system, some informaticists have argued that physicians be involved in tool design [4, 6], and some have even suggested that development should be driven by physician-user demand[4]. These are tentative conclusions based on empirical observations of what works and what does not. However, they are basic theoretical principles in the discipline of human factors, and its subfield of cognitive task design [36-42], applied to clinical decision making [43-46]. Only by consulting the user can the designer build a technology that reliably supports the user's cognitive work.

I.B.4. Summary: CDS for chronic disease and the imperative of user-centricity

Despite the imperative to understand and specifically design to meet user needs, most

CDS tools developed within the informatics community take a document-centric approach. The "customer" is not the user, but rather the academic health and health policy communities who want users (e.g. physicians) to behave the way evidence-based practice guidelines recommend. The purpose of the tool is described not in terms of meeting users' needs, but rather in terms of delivering guidelines to the point of care.

Users need relief from administrative burdens, answers to their questions, and pragmatic suggestions to improve patients' outcomes. These needs could be met by document-centric CDS tools if the recommendations in chronic disease practice guidelines did relieve administrative burdens, and did answer clinical questions arising during the ordinary delivery of care, and did offer pragmatic advice to improve patient outcomes.

But according to the IOM Roundtable on Evidence-Based Medicine, practice guidelines seldom achieve these objectives, primarily because the evidence upon which they are based is generated in a manner that is too distant from the realities of delivering care[14]. This means that CDS tools designed with a document-centric approach not only risk rejection because of clumsy usability, they also risk rejection because the content they offer does not satisfy users' needs for knowledge and administrative support.

I.C. A USER-CENTRIC CDS TOOL FOR AMBULATORY ASTHMA CARE

This dissertation is based upon findings that emerged during the *HIT Asthma Project.* A user-centric decision support tool for ambulatory asthma care was built to meet needs of physicians, nurses and clinical educators at Hennepin County Medical Center (HCMC), an urban safety net in Minneapolis Minnesota. HCMC cares for approximately 10,000 patients with asthma. Most ambulatory asthma care is delivered in nine clinics: four family medicine clinics located in the community, two pediatrics and two adult medicine clinics on the main downtown campus, and one pulmonary specialty clinic also located downtown. Project clinical

leaders included a pediatric pulmonary specialist, a general internist who also was the organization's Chief Medical Informatics Officer, a pharmacist, and a nurse-educator.

The pulmonary physicians needed more asthma-specific administrative and clinical decision support than the electronic health record (EHR) system could provide. The system was an integrated suite of products offered by one of the nation's leading vendors of EHRs. Like other enterprise-wide EHRs that have been built to optimize the workflow functions of contemporary healthcare delivery (e.g. scheduling, ordering, documentation of clinical conditions, billing)[47, 48], this system could provide reminders, alerts and order sets, but it could not meet user demands to produce an asthma action plan (AAP – Figure I.1) that was automatically populated with treatment selections individualized to the patient, with the visual properties that local clinical users required.

Asthma action plans, and care management plans in general, are cited by the clinical community as essential components of care for patients with chronic disease [49-51]. In clinic, such plans serve as "boundary objects" [52] between patient and physician, facilitating communication about home-based asthma management across linguistically disparate communities. Called "self-management education", clinicians believe that conversations facilitated by the plan empower patients to manage their conditions outside of the health care delivery site, and that they are essential for good outcomes.

The HCMC clinicians leading the HIT Asthma project wanted a tool that would produce an AAP individualized to the patient at the conclusion of an asthma encounter, with visual properties similar to the AAPs they already produced manually (Figure I.1), and they wanted the tool to support the decision-making necessary to populate the plan according to the patient's condition and the most recently available asthma guidelines.

Although the enterprise-wide EHR system could not provide the desired support, several

years earlier, the Minnesota Department of Health (MDH) had produced a small, technologically simple "niche product" for asthma that offered some decision support and did produce a visually appealing AAP. The MDH product had become obsolete due to updates and revisions to the guidelines upon which it was based[*]. Additionally, the original niche product only offered support for initiating asthma treatment; it did not offer support for adjusting treatment over time. It could be invoked from a desktop computer or server, but was not available from an enterprise EHR system.

Since the MDH had demonstrated that it was possible to build a tool with some of the desirable properties, the purpose of the HIT Asthma project was to build a similar tool with more advanced functioning. It would include recommendations from the most recent guidelines, it would provide support for adjusting treatment decisions over time, and it would be integrated into the workflow of an enterprise EHR system so that relevant information already present in the EHR would "blow into" the tool automatically. A one-page asthma action plan would appear at the end of each invocation, populated with patient information and the treatment decisions the physician had made.

Clinical leaders guided the tool's development; user acceptance tests were performed with primary care and pulmonary specialty physicians, nurses, clinic managers, respiratory therapists and pharmacists. A software developer wrote the code using scripts in widespread use for web applications. Workflow integration was achieved by enabling users to invoke the tool from within the EHR using the familiar "ordering" process.

To introduce it to users, the pediatric pulmonary specialist and an experienced nurse conducted orientation sessions at clinics that provided asthma care. In these sessions, they

[*] To use the IOM Roundtable terminology, there was no automatic feedback loop that updated the product with new evidence. Summaries of the evidence were updated every 5-7 years, through the guideline revision process.

emphasized not the technology, but rather the importance of patient self-management, facilitated by the asthma action plan. They introduced the software as a tool to help providers achieve these clinical goals.

I.D. OVERVIEW OF RESULTS FROM THE USER-CENTRIC APPROACH

I.D.1. The gap between guidelines and physicians' needs

To treat a patient with chronic disease, physicians must adjust therapies over time, using information about response to prior therapy, and information about new events. These are demanding cognitive tasks that require extensive clinical knowledge and complex reasoning.

Software developers do not know what the cognitive tasks are, and clinical experts perform them without being fully aware [53-55]. To provide necessary content to the software developers, the project's clinical leads consulted the guidelines for asthma care published by the National Asthma Education and Prevention Program (NAEPP)[1], searching for answers to typical questions posed by primary care physicians at the implementation site.

Chapter II in this Dissertation, *Physicians, Guidelines, and Cognitive Tasks,* frames their search as a case study that compares physician requirements for workflow and knowledge support to the structure and content of one of the most widely respected clinical guidelines in the US. Using methods developed in cognitive ergonomics, the chapter concludes that there are wide gaps between what physicians need to perform their jobs efficiently and effectively, and what the guidelines provide.

I.D.2. On the pragmatic effects of introducing the computerized tool

Chapter III, *Effect of a Computerized Decision Support Tool on the Proportion of Asthma Patients with Asthma Action Plans*, evaluates the impact of introducing the CDS tool at HCMC on the percentage of asthma patients with current AAPs. It also explores reasons why patients

with presumably similar clinical profiles might not have electronically-generated plans: Their physicians resisted the technology, and/or they weren't generating plans. Weekly rates were calculated for current AAPs, generated manually on paper, by the new tool, and overall. Interrupted time series analysis was used to assess the impact of introducing the tool.

Graphs of the series of observations over time were visually striking, and supported derivation of a theory that explained the contribution of the CDS tool to the likelihood that an asthma patient would have a current AAP. According to the theory, given similar clinical and care-seeking profiles, patients with physicians inclined to create AAPs would have them if support for physicians were available, and the physicians preferred computerized to paper support. Patients with physicians not inclined to create AAPs would not have them, even if physician support were available. If correct, a promising strategy to increase the percentage of patients with current AAPs is to target recalcitrant physicians and encourage them to produce plans, with success more likely if computerized support were available.

Although the data were insufficient to test the theory, they did show that physicians did not reject the technology, and they pointed towards useful areas of future inquiry.

I.D.3. Ethical challenges posed by faithful adherence to presenting guidelines on-screen

Chapter IV in this Dissertation, *Ethical Issues in Computerized Decision Support for Chronic Disease Management*, explores two ethically problematic situations that arose as a consequence of treating the tool primarily as a way to disseminate recommendations based on existing evidence. If the tool were to faithfully adhere to the guidelines, all recommendations appearing in the guidelines would be presented on screen, and only those recommendations appearing in the guidelines would be presented on screen. Ethical quandaries arose when the clinician/developers became aware of instances in which such adherence could harm some patients and deprive others of benefit. Technically, these

quandaries could have been resolved by presenting the questionable choices on screen and attaching tags explaining caveats. The chapter explains that in one instance, this resolution was socially acceptable but in the other instance, the same resolution was not. The difference in social acceptance was due to the privileged status granted to evidence generated by randomized controlled trials (RCTs).

The chapter reviews the ethical and epistemic objections to privileging RCTs, particularly as a design capable of generating evidence useful for primary care of patients with one or more chronic disease. The chapter suggests that if CDS tools were conceived not just as mechanisms for disseminating evidence obtained primarily from RCTs, but also were conceived as mechanisms for generating observational evidence arising from practice, not only could these ethical quandaries be resolved, the tool could serve as a prototypical technology necessary to achieve the IOM Roundtable vision of a "healthcare system that learns".

I.E. CONCLUSIONS EMERGING FROM THE USER-CENTRIC DEVELOPMENT APPROACH

The user-centric developmental approach accented guidelines' limitations. Although informaticists have long been aware of these gaps, such awareness has not stimulated the informatics community to reevaluate the basic goals of CDS tools. Unlike informaticists, the users whose daily professional activities would be influenced by the tool were much more inclined to be alarmed by these limits and thus to question the basic goal of the tool, particularly if limits in the guidelines created by gaps in existing evidence leads to ethically problematic features of the tool. In this Dissertation I argue that the basic goals of CDS tools for chronic disease must be expanded. Not only should they disseminate evidence, derived primarily from RCTs, they also should be to generate practice-based evidence.

FIGURE I.1 SAMPLE ASTHMA ACTION PLAN

My Asthma Action Plan

Name: **Abigail Test**
Date: **07/10/2009**
My Doctor or Clinic: **HCMC General Medicine**
My Doctor or Clinic Phone: **612-873-2300**

My Asthma Severity: **Severe Persistent**
My Peak Flow Number: **45**
Avoid your asthma triggers: **Bronchitis , Strong Odors , Cold Air ...**

GO

- I feel good
- No cough or wheeze
- Can work, sleep and play without asthma symptoms

My peak flow number is above **36**

Green Zone: Asthma in good control

1. Take your asthma control medicine every day:
   - **Fluticasone Inhaler 100 mcg (Flovent) 2 puffs twice a day**
   - **Zileuton Tablet ER 600 mg (Zyflo CR) 2 tablets twice a day**
2. If exercise triggers your asthma, take:
   - **Levalbuterol nebulizer solution, 1.25mg/0.5ml (Xopenex) 1.25mg/0.5ml**
   - 15 minutes before exercise or sports, and
   - during exercise if you have asthma symptoms
3. Spacer to use with inhaler:

Slow

I have any of these:
- I do not feel good
- Cough or wheeze
- Chest feels tight
- Wake up at night

My peak flow number is between **22 and 36**

Yellow Zone: Asthma getting worse

1. Keep taking your Green Zone medicines.
2. Start taking your rescue medicine:
   - **Levalbuterol nebulizer solution, 1.25mg/0.5ml (Xopenex) 1.25mg/0.5ml** every 20 minutes for up to 1 hour. Then every 4 hours for 24-48 hours.
3. If you do not return to the Green Zone in 12-24 hours, or you get worse, start taking your oral steroid medicine:
   - **Prednisone Tablet, 15.0 mg, Twice a day for 5 days.**
4. If you stay in the Yellow Zone for more than 12-24 hours, call your doctor.

Stop

911

I have any of these:
- I feel awful
- Medicine not helping
- Breathing getting harder
- Trouble walking or talking
- Nose opens wide to breathe

My peak flow number is below**22**

Red Zone - Medical Alert - Get help

1. Take your rescue medicine NOW:
   - **Levalbuterol nebulizer solution, 1.25mg/0.5ml (Xopenex) 1.25mg/0.5ml**
2. Take your oral steroid medicine NOW:
   - **Prednisone Tablet, 15.0 mg, Twice a day for 5 days.**
3. Call your doctor NOW.
4. If you are still in the Red Zone after 20 minutes, and you have not reached your doctor:
   - Take your rescue medicine again, and
   - Call 911 or go to the emergency room right away.

Clinic or provider for follow up: **HCMC General Medicine, 612-873-2300** When: **4 weeks**
Electronically signed by : **chill, MD**
Person given Asthma Action Plan and Trigger Control sheet: **The patient**

CHAPTER II. PHYSICIANS, GUIDELINES AND COGNITIVE TASKS

II.A. INTRODUCTION

Evidence-based clinical practice guidelines are implemented erratically if at all during the routine delivery of medical care[8, 9], a robust finding across fields such as knowledge translation[56-59], translational research[60], and implementation research[61]. In a normative attempt to close the gap between evidence and the delivery of medical care, most research has examined strategies to change physician behavior. Direct strategies target individual physicians, and include education, practice-based reminders and alerts, and incentives. Indirect strategies target the social and organizational contexts within which physicians work[62-65]; they have used social tactics such as relying on opinion leaders[66, 67], and tapping the power of practice communities (e.g. multidisciplinary practice teams including nurses, pharmacists, physicians) that collectively engage in collective "sense-making" about the applicability of guidelines for the patients in their care[68, 69]. While multiple interventions delivered simultaneously sometimes achieve results [70], especially when context is also addressed[71], no interventional strategy targeted towards physicians has shown resounding success[72].

If physicians are the interventional targets, it is reasonable to assume that those designing the interventions occupy the standpoint of guideline developers, who view physicians as recipients of the knowledge they wish to impart. This chapter takes the standpoint of practicing physicians, who view guidelines as potential resources to help perform work more effectively. Using asthma as an illustrative case, it compares physicians' tasks and cognitive requirements during a typical ambulatory encounter to the structure and content of the most widely respected asthma guidelines in the US. Analytic methods are derived from theories of human factors, and in particular, cognitive ergonomics [37], a perspective that informs the

design of artifacts and information systems intended to support cognitive tasks [36].

The analysis reveals gaps between resources physicians require and what guidelines provide, and leads to the conclusion that guidelines do not adequately support medical work.

II.B. METHODS

II.B.1 The Guidelines in the Case: Third Expert Panel Report (EPR-3)

The *EPR-3: Guidelines for the Diagnosis and Management of Asthma*, was published in August 2007 by the National Asthma Education and Prevention Program (NAEPP); NHLBI formally introduced it to the clinical community through a press release highlighting its importance[73]: "The NAEPP today issued the first comprehensive update in a decade of clinical guidelines for the diagnosis and management of asthma".

The EPR-3 is an illustrative case of evidence-based practice guidelines for two reasons:

1. Gaps between these guidelines and physician requirements cannot be dismissed as a consequence of poor guideline quality, because the EPR-3 is an exemplar of guideline quality, according to currently accepted criteria [74-79].

2. Low rates of adherence to these guidelines cannot be attributed to lack of awareness of them. Physician surveys conducted in the late 1990s showed that asthma guidelines enjoyed some of the highest rates of awareness [80] and familiarity [81]. Among surveyed pediatricians, more reported using guidelines for asthma than for any other clinical condition[82]. Yet when those surveys were conducted, adherence to asthma guideline recommendations was still low [83].

II.B.2. The project site as an illustrative case of guideline implementation

The clinical tasks and knowledge requirements were elicited from physicians participating in a project to create a computerized decision support tool for asthma. They accepted the value of the national asthma guidelines and were committed to implementing them. They actively communicated their informational and usability requirements through their

participation in the design of the computerized tool.

These physicians practiced at Hennepin County Medical Center (HCMC), a quasi-public safety net located in Minneapolis, Minnesota. Much of the care provided at HCMC was either uncompensated, or compensated through low rates from public medical assistance programs. As a result, financial resources were limited, and patient resources were limited as well. Illiteracy, unfamiliarity with English, mental illness, homelessness, substance abuse – all are common among the patients at the study site.

This site serves as an illustrative case because:

1. As noted, at the clinics providing asthma care, clinical leaders accepted the value and validity of the NAEPP guidelines. Failure to implement guideline recommendations at these clinics thus could not be attributed to cultural resistance stemming from epistemic and/or value-based challenges among leadership;
2. Site resources were limited; success at implementing guideline recommendations thus could not be attributed to an abundance of resources not replicable elsewhere.

II.B.3. Methods to analyze physicians' tasks and cognitive requirements

The tasks and cognitive requirements under analysis are those that primary care physicians encounter when they are physically in the exam room with a patient who already has been diagnosed with asthma. The immediate context is created when the patient enters the clinic (Figure II.1), framed by the resources provided and the constraints imposed by HCMC and the patient's social and cultural milieu. As noted, resources were limited. Also, like many other health care delivery sites in the US, compensation is triggered by the encounter, which introduced financial pressure to perform as many encounters as possible in a single day, necessarily reducing the amount of time available for each one.

Analysis pertains to the tasks the physician is expected to perform during the limited

amount of encounter time (innermost circle, Figure II.1) [*].

Methods to discern physicians' tasks employed principles of hierarchical task analysis [84, 85]. This framework decomposes complex tasks into subtasks, operations and actions, and then represents them graphically. The graph then serves as a model for the activities[86].

Physicians described their tasks by drawing pictures of the computer screens they wanted to see, one sheet of paper for each screen, placed in expected order. Content and sequence often depended on the outcome of underlying logical conditions, such as the presence of an existing record for an asthma encounter, or whether the patient was currently taking controller medications for asthma. These logical conditions were represented using the Microsoft product Visio, where hyperlinks to images of relevant screens were attached to each logical branching point. The Visio models gave users the opportunity to critique prospective software functioning, and ensured that the model faithfully represented their tasks.

While hierarchical task analysis was used to model the sequential sets of tasks appearing in physicians' workflows, cognitive task analysis [40-42] was used to model the cognitive requirements associated with an ambulatory encounter for asthma. Cognitive methods are particularly useful, and particularly challenging, for tasks that require extensive, but often tacit, subject-matter expertise [39, 40, 87, 88].

Subject-matter knowledge was elicited iteratively. Clinical experts communicated knowledge requirements using expert clinical vocabulary, often without operational detail. Software developers requested that detail, which the experts attempted to provide by searching the EPR-3. If operational detail was available in the EPR-3, they used that; if it was not, they translated their clinical vocabulary into operational terms the developers could use.

---

[*] Although physicians often are unable to complete these tasks in the exam room, these deviations from idealization are irrelevant here.

II.B.4. Methods to analyze the structure and content of the EPR-3

Because the EPR-3 is a narrative document, its use entails reading sequences of sentences, paragraphs, and sections. The sequential order reflects the artifact's structure. It is represented with an ordered list of its main and first two sub-section headings, along with the number of pages each sub-section consumes. Number of pages is assumed to represent the amount of information the artifact provides for the topic under analysis. Actual content was determined by project physicians, who searched the EPR-3 for the operational specificity the software developers required, noting every page where required content was found.

II.C. RESULTS

II.C.1. Physicians' Workflow Requirements

II.C.1.i. Tasks and activities

The sequence of screens the user encounters while executing the software is shown in Figure II.2., which also models the flow of clinical tasks during the asthma encounter. More detail associated with each task is provided in Figure II.3. The activity sequence begins and ends with documentation, and documentation takes place throughout the encounter. Anecdotal reports suggest that many of the physicians used the documentation activity as an opportunity to provide patient education, by inviting the patient to view the screen with them, explaining what the elements on the screen mean.

At the highest level, once the identity of the patient with asthma has been confirmed, the remaining task sequence is straightforward: Diagnose the nature, intensity and complexity of the patient's asthma; determine the appropriate level of therapeutic aggressiveness given the patient's history of responses to prior therapy; select and order therapies; complete the asthma action plan and review it with the patient.

II.C.1.ii. Knowledge requirements

Figure II.4 lists the pieces of clinical knowledge necessary to complete an ambulatory encounter for asthma so that it adheres to guideline recommendations. The requirements include knowledge of the diagnostic criteria to determine the patient's asthma severity or control, which in turn include specific symptom reports, and the results from objective lung function tests. The requirements also include detailed knowledge about several classes of pharmaceutical products. Finally, requirements include knowledge about conditions under which it is appropriate to change the level of treatment intensity, when to refer a patient to a specialist, how frequently the patient should be seen for asthma.

II.C.2. Structure and content of the EPR-3, and its relation to knowledge requirements

II.C.2.i. Structure

The EPR-3 contains approximately 440 pages of text and diagrams. The introductory material and the first two sections, consume 58 pages (13% of the total). The content in these sections is not required by physicians while they are delivering care. When the introductory sections are added to references and other parenthetical remarks, approximately half the content of the EPR-3 is not useful while the physician is delivering care.

Sections 3-5 in the EPR-3 are relevant to care. Section 3 addresses "… the four essential components of asthma care, namely: assessment and monitoring, patient education, control of factors contributing to asthma severity, and pharmacologic treatment". Section 4 covers the long-term management of asthma. Section 5 deals with managing asthma exacerbations.

The EPR-3 does not support a number of necessary clinical tasks. It does not address documentation activities, and it does not give explicit advice about what the patient should do in the case of an extreme exacerbation. It does not address situations in which the patient's resources limit the ability to execute its recommendations.

The EPR-3 does not have an index, thus it is not possible to search by subject. If viewing the EPR-3 electronically, it is possible to search for character strings, but the most salient topics may generate hundreds of instances of the string. The Table of Contents is highly detailed, but topics are listed in order of appearance, rather than alphabetically; its subject headings employ academic rather than clinical vocabulary, which do not direct the user to specific topics. For example, the Table of Contents does not help locate a topic of highest importance to physicians: How to determine the patient's current treatment step.

II.C.2.ii. Content.

This section compares working clinicians' questions to the knowledge content of the EPR-3, limiting analysis to core clinical decisions: How aggressively to treat the patient, and what daily controller medications to use to achieve that target aggressiveness.

With respect to level of aggressiveness, the EPR-3 employs the terminology of "stepwise management". It recommends that the clinician initiate treatment at a step corresponding to the patient's level of (untreated) asthma severity, increasing the step if the patient's asthma is not controlled, and potentially decreasing the step if the patient's asthma remains controlled at the current step. The EPR-3 refers to six steps, from lowest to highest aggressiveness.

For each step, the EPR-3 describes preferred and alternative treatment plans[*]. A treatment plan is comprised of one or more pharmacologic elements; the EPR-3 names 10 unique elements in its section on daily controller medications. The cornerstone element for most treatment plans is an inhaled corticosteroid (ICS); plans corresponding to higher steps have higher doses of the ICS and/or supplement the ICS with another element.

Given the four potential doses of the ICS (none, low, medium, high), and the presence or

[*] Within each step, there are three sets of plans, one for each of the three patient age groups: Age 0-4 years, age 5-11 years, and age 12+ years. Sometimes more than one preferred plan is listed; sometimes there is no alternative plan listed. The EPR-3 does not explain the difference between preferred and alternative plans.

absence of any of the other nine elements, there are 2048 logically possible treatment plans (4 x 2^9). Of these, the EPR-3 explicitly represents 20-24 of them, differentiated by step and patient age. Except when it refers to an element by proprietary name, the EPR-3 does not operationalize plans by listing commercially available products that fulfill them.

This presentation of step, plan and combination of pharmacological elements does not provide sufficiently actionable knowledge. For example, the EPR-3 recommends that the clinician increase the treatment step if the patient's asthma is not being controlled at the current step. But unless the current step has been documented, the clinician would not know what it is. This was one of the most important physician requirements; both Maviglia et al[22] and Bates et al [32]deemed it a formidable challenge in providing CDS for chronic disease.

The EPR-3 addresses this issue in only two brief paragraphs. It suggests that the physician match the patient's medications to the step in which they appear as elements of a treatment plan. But as noted above, the EPR-3 only mentions 20-24 plans out of the 2048 logically possible medication combinations that a clinician could see in actual practice[*], and thus offers a "current step" for only 1-2% of logically possible clinical realities. There was, for example, no step in the EPR-3 corresponding to a medium-dose ICS, a long-acting beta agonist (LABA) and Montelukast, which physicians used quite commonly.

Furthermore, the EPR-3 was unclear with respect to what 'currently taking' means. In one paragraph, 'currently taking' is defined in terms of patient report; in the other, 'currently taking' is defined in terms of physician belief. Because patient report may differ from physician belief, how to proceed with clinical therapy will differ as well.

The narrative, figures and tables in the EPR-3 use the terms 'low', 'medium', and 'high'

[*] Initially, project physicians dismissed most of these combinations as clinically ridiculous, never expected to see them in practice, and questioned the need to support them. Yet viewing a "ridiculous" combination often prompted them to recall a situation in which they had observed it. Eventually they accepted that if the combination was logically possible, then it could be observed, and if it could be observed, support was necessary.

doses of ICS, although commercially available ICS medications are seldom labeled this way. Many ICS products come in only two strengths, which their manufacturers call low-medium or medium-high. The EPR-3 acknowledges that ICS doses vary by product and delivery device, but provides mapping functions for only a subset. In these functions, the EPR-3 expresses doses in terms of mcg/day of the product's active ingredient. Clinicians and patients express doses in terms of puffs (e.g. take two puffs twice a day). Some of the dosing functions that the EPR-3 expressed in terms of mcg/day could not be achieved with commercially available products, given the mcg per puff. For some products, the EPR-3 high dose was classified as an overdose by the manufacturer. The EPR-3 did not discuss the common clinical situation in which an ICS and a long-acting beta agonist (LABA) is packed together in the same product, so that a high dose of the ICS could deliver an overdose of the LABA.

II.C.2.iii. Terminology

The EPR-3 often employed vague and/or ambiguous terminology, capable of being interpreted multiple ways. The Introduction suggests that the Expert Panel employed such terminology deliberately, using it to communicate the strength of its commitment to a recommendation. The consequence, however, is that it provides no direct guidance. For example, on page 287, the report states, "For patients who have *very poor* asthma control, consider increasing treatment by two steps, a course of oral corticosteroids, or both." (emphasis in the original). The report offers three suggestions to respond to very poor asthma control, and does not recommend any one of the three over another.

II.D. DISCUSSION

This analysis was guided by theories in cognitive ergonomics, which stimulates investigations about the likelihood that an artifact will help or hinder the user at work. Investigations guided by this theoretical frame have found that supportive artifacts are those

which the user experiences as intuitive, efficient and pleasurable to use, while providing all the information they require. Artifacts that hinder work either do not provide necessary information, or create burdens while providing it, such as forcing the user to divert attention away from core tasks and activities [38, 89, 90].

If the EPR-3 is treated as an artifact, theory suggests that it is unlikely to support medical work. Its relevant content is often dispersed among different subsections throughout the report, and without an index, the content is difficult to find. It employs vague vocabulary lacking actionable operational detail, and/or academic vocabulary that neither physicians nor patients use. The presentation mode is thus a barrier for practicing clinicians, who typically seek answers to questions by consulting immediately available proximate sources that are quick to use [11]. Furthermore, even when sources are proximate and available, clinicians will only search if the expected value of the outcome is higher than the expected value of other activities that compete for the same limited time [91]. From the perspective of the working clinician, the EPR-3 information value is low, because it is not expressed at the necessary level of operational detail. The detail it does contain is often incomplete (e.g., it assigns steps to only 2% of the medication combinations that could be observed; it provides dosing ranges for only a subset of commercially available ICS products), it is sometimes internally inconsistent (e.g., it gives two different definitions of 'medications the patient currently is taking'), and uses vocabulary that is difficult to translate into clinical practice (e.g., it expresses ICS doses in mcg rather than puffs per day). It contains many instances of what McDonald and Overhage memorably call "weasel words" [92]: vague or ambiguous expressions that carry little informational value, because they can be interpreted multiple ways.

Use of the EPR-3 could hinder medical work. Studies of medical expertise have shown that the cognitive processing strategy most closely associated with diagnostic accuracy [43, 93, 94]

uses forward reasoning from observations to hypotheses, while relying on tacit subject-matter knowledge to filter out irrelevant observations and cues [95]. The presence of many non-salient observations and cues can disrupt these successful cognitive processes [96], even among experts. Thus extraneous information may disrupt the reasoning of medical generalists and/or novices, precisely the audience the EPR-3 wishes to reach.

During the course of a typical working day, a primary care physician confronts multiple clinical conditions, for which there may be a voluminous literature addressing therapeutic strategies. Clinical practice guidelines were conceived as a solution to the problem that no individual physician could possibly stay current with this literature and the evidence it describes [97, 98]. But if guidelines are represented as lengthy un-indexed reports, filled with information that is not relevant to the immediate task of delivering care, with the relevant information that it does contain scattered throughout the document in sometimes contradictory ways, using academic rather than clinical vocabulary, without the operational detail that physicians require, theory suggests that we should not wonder why physicians don't use them. We should wonder why they ever do.

These predictions from theory are supported by evidence. When asked why they do not use guidelines, physicians frequently say that they are confusing, inconvenient, and/or difficult to use[99]. In one survey, 43% of physicians responding cited ambiguity and lack of clarity in guideline recommendations as a reason for ignoring them[100]. Experiments have shown that rates of physician adherence to recommendations increase when they are expressed using specific rather than vague terminology [101-103].

The cognitive breach between clinical work and lengthy narrative documents is precisely why the medical informatics community has established an active sub-discipline focused on representing clinical guidelines through computer code, executable at the point of care [3, 5,

12, 104]. Reviews of the effectiveness of such systems to improve adherence to guidelines and patient outcomes have generated mixed results [26-29], although many do not differentiate between simple "critiquing" compared to more complex "guided" support. None of these reviews addressed the knowledge gaps between the narrative guidelines and what is required by executable code, although these gaps have been well-documented in the technical informatics literature [22, 30, 105-107].

Both informaticists writing code and practicing clinicians delivering care must draw from a knowledge base, which is expressed through the very same narratives that convey evidence-based practice guidelines. Therefore, informaticists must interpret the narratives as working clinicians must do. While both are frustrated by vagueness and lack of operational detail, the consequences differ. If an individual clinician encounters gaps, he or she will fill them with tacit knowledge acquired through experience, along with tacit understanding of the context in which the knowledge is to be applied. If the clinician is unable or unwilling to do this, he will abandon the text. The effect of interpretation and/or abandonment is limited to that one individual, a physician who presumably is charged with final clinical responsibility.

The informaticist, by contrast, is developing code to be disseminated across numerous health care delivery sites, which will inform and guide the decisions of numerous clinicians. Code cannot contain the same gaps as text because executing computers, unlike humans, will not be able to fill them with tacit knowledge acquired through experience. The code must specify exactly what the computer is supposed to do under each and every logically possible circumstance. If it is incomplete, ambiguous and/or internally inconsistent, the computer will either refuse to execute it, or it will crash trying.

To prevent this from happening, software engineers demand precise, operationally defined, logically complete, unambiguous instructions. Because the narratives in which

guidelines are expressed typically do not contain statements of this kind, the engineers or those directing them must provide that operational detail.

The EPR-3 authors state that vagueness should be interpreted as weak commitment, because of a paucity of evidence, or inconsistency in the evidence that exists. In such instances, guideline authors employ vague vocabulary to encourage physicians to use clinical judgment. However, the actual effect of imprecise vocabulary may be to discourage physicians from consulting guidelines, and to encourage software engineers to assume clinical authority.

II.E. SUMMARY, LIMITATIONS AND CONCLUSIONS

This case study examined a well-respected set of guidelines for asthma care, comparing them to the workflow of primary care physicians treating diagnosed asthma patients at an urban safety net. Using the theoretical framework of cognitive ergonomics, the analysis shows that when guidelines are represented as lengthy narrative reports, they are unlikely to support physicians' medical work.

Although it would be valuable to perform a similarly structured analysis for other conditions, the results are likely to be similar, because others have expressed similar difficulties when working with guidelines for heart failure [105, 106], hyperbilirubinemia in newborns [107], diabetes [108], hypertension [109], and cardiac care [110]. It also would be valuable to perform a similarly structured analysis documenting the experience of other groups that attempted to convert the EPR-3 into computerized decision support.

Assuming the results from this case study can be generalized to other guidelines, and to the workflow requirements of physicians at other sites, the conclusion is that those who are concerned about the gap between evidence and care consider the perspective of working physicians. Most interventions intended to close the gap target physicians as the agents responsible for the gap. Physicians have cognitive limits, and face significant time and

knowledge constraints during the actual work of delivering medical care. If guidelines are intended to overcome these natural cognitive limits, they must be represented with artifacts that physicians can actually use. Using the well-respected EPR-3 as an illustrative case, this analysis shows why physicians can't use guidelines in their current representational form.

FIGURE II.1. CONTEXT OF THE PHYSICIAN-PATIENT ENCOUNTER ABOUT ASTHMA

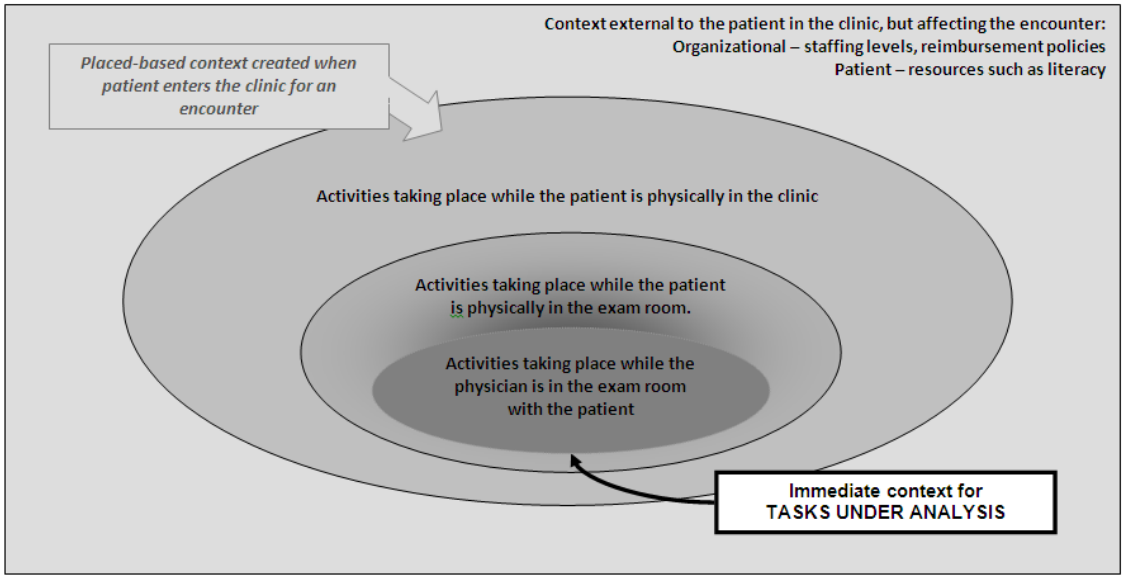


FIGURE II.2. TASKS PERFORMED DURING AN AMBULATORY ENCOUNTER FOR ASTHMA

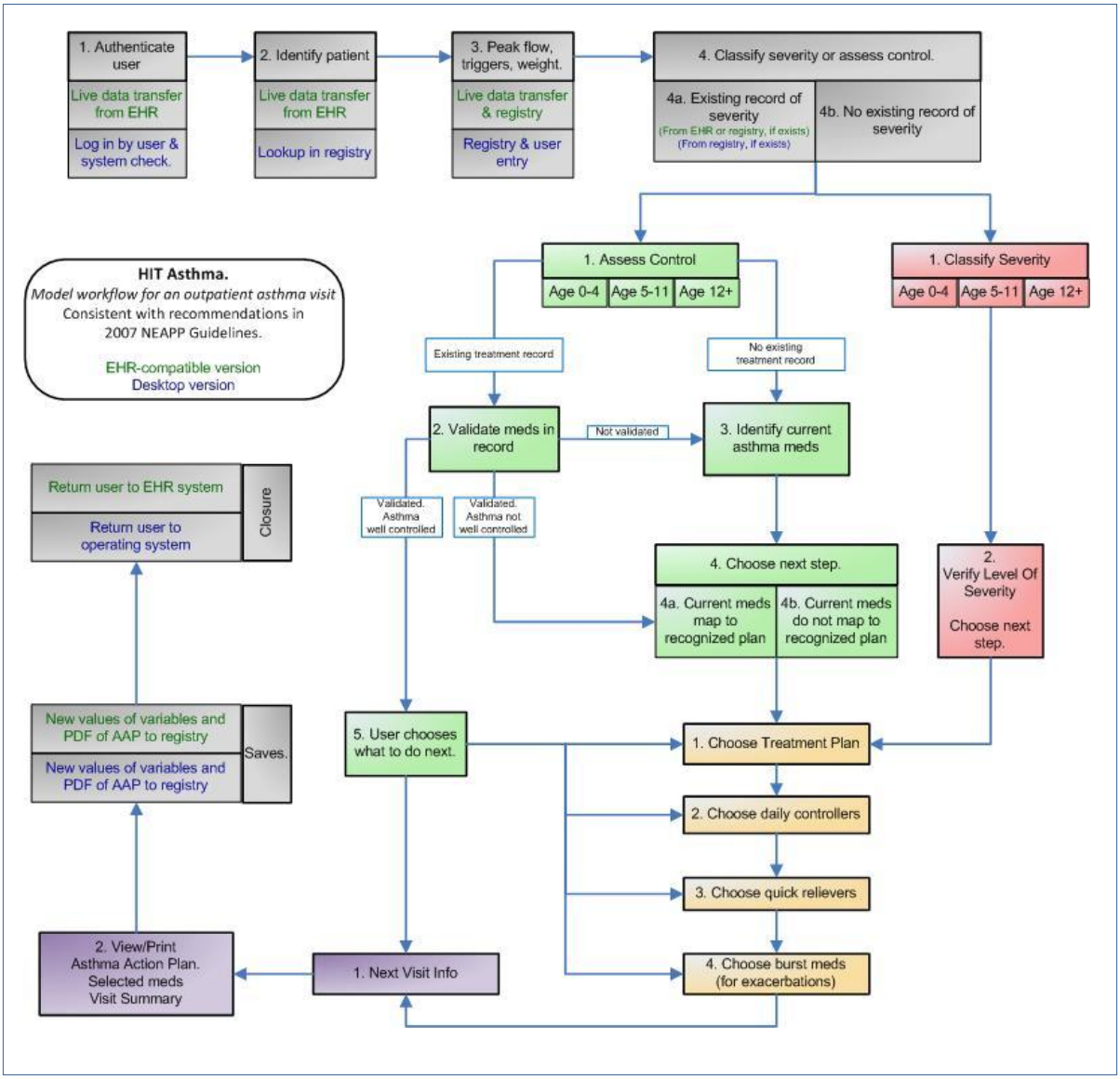


FIGURE II.3. TASKS PERFORMED IN THE EXAM ROOM DURING AN AMBULATORY ENCOUNTER FOR ASTHMA

Perform initial documentation.
- Verify patient identity
- Review medications
- Record complaints
- Review reason for visit

Determine purpose of visit. Asthma?
- New visit or followup visit for asthma?
  - If new visit → Severity? Triggers? Duration?
  - If followup → Control?
- Symptom reports and lung function results?
- Aggressiveness of current treatment?
  - Current meds?
  - Meds ←→ Aggressiveness?
  - Taking the meds as prescribed?
- Make primary clinical decisions:
  - Aggressiveness of treatment now?
  - Which daily controller meds?
  - Consider:
    - Adverse consequences
    - Patient resources and receptivity
    - Cost. Insurance? Co-pays? Refills?
    - Daily burden.
    - Patient beliefs around efficacy
- Make secondary clinical decisions:
  - Which rescue meds?
  - Instructions in case of exacerbation?
  - Next followup visit?
- Review with patient; modify if necessary

Place orders
- Medications
- Further tests if necessary
- Next visit

Perform final documentation
- For the patient
  - Asthma action plan (review with patient)
  - After-visit summary (review with patient)
- In the chart
  - Record and/or modify asthma in problem list
  - Compose and place progress note in chart
  - Compose encounter summary

FIGURE II.4. KNOWLEDGE REQUIREMENTS FOR AN AMBULATORY ENCOUNTER FOR ASTHMA (font saturation proportional to degree of completion in the EPR-3)

- Predicted peak flow based on patient's age, height and sex
- **Questions to ask and responses to record, to assess impairment from asthma**
- **How to convert symptom reports, lung function results, and results from Asthma Control Test to measures of asthma severity and control**
- **Step to initiate treatment based on measure of asthma severity**
- **Step to modify treatment based on measure of asthma control, current step, and whether patient was adhering to treatment plan.**
- **How to infer the current step from patient's reported use of daily controller medications.**
- **For each inhaled corticosteroid (ICS), the mcg/day associated with a low, medium, and high dose, differentiated by patient age.**
- **For each inhaled corticosteroid delivered by canister, number of puffs per day associated with low, medium and high dose, differentiated by patient age.**
- **Options for daily controller treatment plan, based on patient age and treatment step.**
- Commercially available medications, including brand, route of administration, strength, and instructions for use, fulfilling all possible treatment plans.
- **Commercially available medications, including route of administration and strength, that could be used as rescue or quick reliever medications.**
- **Commercially available holding chambers, based on previously selected routes of delivery for daily controller and rescue medications.**
- **Commercially available medications that could be used as systemic oral corticosteroids**.
- Number of mg/day for an oral corticosteroid that constitutes an overdose, based on patient's age and weight.
- Starting dose for an oral corticosteroid.
- **Follow-up time, based on patient age, level of asthma severity or control, and adherence to current treatment plan.**
- Instructions in the "yellow" and "red" zones of the printed asthma action plan.

CHAPTER III. EFFECT OF A COMPUTERIZED DECISION SUPPORT TOOL ON THE PROPORTION OF ASTHMA PATIENTS WITH ASTHMA ACTION PLANS

III.A. INTRODUCTION

Asthma is a common chronic disease affecting 23.3 million people in America and the cause of significant morbidity. Asthma patients who receive self-management education show improved health outcomes [49-51], which are enhanced when they also receive a written, individualized asthma action plan (AAP)[49] containing features found to be effective. Those features include prescribed daily controller medications, with dosing and instructions for use, along with instructions for recognizing and responding to exacerbations [111]. National asthma guidelines from several countries unequivocally recommend AAPs [1, 112, 113]. However, the rates with which primary care physicians provide [114-116] asthma action plans are low, and patient surveys show that few possess them [117, 118].

To produce the necessary content for an AAP, the provider must make a series of complex clinical decisions, beginning with an ascertainment of the patient's level of asthma severity or control. Given the burden of the disease, the provider must decide how aggressively to treat using pharmacological management with daily doses of asthma "controller" medications. Because all medications carry risks of untoward iatrogenic effects, the physician must find a balance between therapy aggressive enough to achieve control but no more aggressive than necessary. Scores of pharmacological products are commercially available, each with its own mechanism of delivery and per unit dose – knowing which product to use, with what delivery mechanism, and at what dose, can all be daunting to a primary care physician who sees numerous conditions each day. Physicians have reported uncertainty about the treatment

decisions necessary to populate an AAP; they also have expressed uncertainty about the appropriate visual format in which to present the content of the plan [117, 119]. The time required to create an individualized AAP imposes an additional barrier [117, 120-122], particularly in the pressured "assembly-line" context of ambulatory care in busy clinics.

We attempted to address these knowledge and time barriers with a guideline-based[1] computerized decision support (CDS) tool. Through a series of 6-8 intuitive screens taking less than 5 minutes to complete, the tool helps the user assess the patient's level of asthma severity or control, offers guideline-based recommendations of therapeutic aggressiveness (e.g. "treatment step"), choices of treatment plans to achieve the selected aggressiveness, and choices of commercially available medications to fulfill the plan. The tool inserts users' choices into a one-page, individualized AAP, which appears on the final screen. The clinician can print and review the AAP with the patient in clinic, and the patient can take a copy home.

To improve the rates with which patients received complete AAPs at Hennepin County Medical Center (HCMC), an urban safety net in Minneapolis MN, we conducted an intervention that included 1) education about the importance of AAPs, 2) an introduction to the tool, and 3) instructions for invoking the tool from HCMC's enterprise-wide electronic health record (EHR) by "ordering" an AAP. Placing an order was a common task well-supported by the enterprise EHR. This paper addresses two questions: 1) What was the impact of the intervention on the rates with which pediatric and adult asthma patients at HCMC had current and complete AAPs? 2) If asthma patients were not receiving electronically generated AAPs, was it because providers resisted the computerized technology, or because they were not completing AAPs?

III.B. METHODS

III.B.1. Implementing the asthma CDS tool

After an initial 3-week pilot at two pediatrics clinics[*], the tool went live throughout the enterprise on July 10, 2009. A pediatric pulmonary specialist and an experienced nurse conducted orientation and training sessions at primary care clinics from July through December 2009. They emphasized aligning asthma care with the guideline recommendations, highlighting the role of asthma self-management education, and identifying the written asthma care plan as a critical component. They introduced the CDS tool as user-friendly software to help generate the plan. After group sessions, the nurse-champion offered help to users in the clinics and to answer questions about the tool.

III.B.2.Analysis

III.B.2.i. Study design

We employed interrupted time series to analyze the weekly rates with which asthma patients had current AAPs created using paper templates, the electronic tool, and overall, for a two-year period "interrupted" approximately halfway by the introduction of the tool. Rates were analyzed for adult patients with visits at two neighborhood family medicine clinics, and for pediatric patients at the same neighborhood clinics plus two pediatric clinics on the main hospital campus. Results were generated from different age groups, receiving care at different sites, to explore context effects and to assess generalizability. Outcomes over time showed the effect of the intervention against pre-existing secular trends.

[*] During the pilot, a flaw in the database was discovered. It was corrected, but data collected during the pilot were lost. This is accounted for in the analysis using methods described below.

III.B.2.ii. Operational definition of outcome

The primary outcome was the weekly percentage of asthma patients with current asthma action plans (AAP), defined as being created within a year of the observation date. A secondary outcome was the percentage of asthma patients with "complete" AAPs, or AAPs containing all features deemed effective. The software ensured that all electronically generated plans were complete; a small sample showed that only 5% of manually-generated plans using a paper template contained all effective elements.

At each weekly observation date, the denominator ("patients at risk") is the number of patients known to have asthma who had at least one visit to HCMC within the previous year. Approximately nine thousand patients with asthma were identified from the EHR using methods described in Appendix B. The numerator at each weekly observation date is the number of patients at risk with at least one AAP dated within the preceding year. Patients with electronically-generated plans were identified from the database of the CDS tool. Patients with paper-generated plans were identified by manually reviewing the electronic charts for instances of paper plans that had been scanned into the record. Reviews were performed for a targeted sample of 899 patients; of these, 553 pediatric patients age 5-11 years, and 268 adult patients age 21 or more years, from two pediatrics clinics and two family practice clinics, met criteria for being at risk of receiving an asthma action plan, and were included in the analyses.

III.B.2.iii. Statistical Methods

To determine the effects of the intervention compared to secular trends, the weekly population prevalence rates for AAPs were modeled using the Autoregressive Integrated Moving Average (ARIMA) methods developed by Box and Tiao[123] and refined since [124].

Observations *y* at time *t* are represented as over-time "random noise" that appears after removing drifts or trends, adjusted by ways in which observations and/or shocks to the series at one moment in time are perpetuated through the series over time.

If a known "interruption" to the series takes place during the observation period, its immediate and extended effects upon the series are modeled as well. Hypotheses about the nature of the interruption effects are modeled with parameters that represent the effects of the interruption, where models also contain terms representing background noise, or secular trends. Parameters representing the effect of the interruption are tentatively retained if 1) compared to noise, their values attain conventional levels of statistical significance, and/or 2) their inclusion significantly improves the overall model fit, and/or 3) their inclusion generates a more random pattern among the residuals. Because a number of different models can be produced that meet one or all of these criteria, candidate models are further evaluated on the basis of how parsimonious they are, and the degree to which their interpretation is consistent with the visual display of the observations shown graphically over time.

Complete specification of effects includes identification of "lags", which refer to the number of intervening time periods between the appearance of an effect and its reappearance over regular intervals of time. Complete specification of the interruption effects includes:

1. Whether the effect has a gradual or abrupt onset;
2. The magnitude of the initial effect (the parameter ω). If the initial effect is gradual, its magnitude is represented as a change in slope. If the initial effect is abrupt, its magnitude is represented as the size of the step from the level of the series before and after the interruption took place. The larger the absolute value of the ω parameter(s), the greater the magnitude of the initial effect. Multiple ω parameters can be fit in a model, representing different lags over which the effect is observed.

3. The attenuation of the effect over time, represented by the parameter δ. Attenuation is assessed along two dimensions: 1) steadiness (or by contrast, variable/oscillating) is evaluated through the sign of the δ parameter(s); 2) speed of attenuation is evaluated through the magnitude of the absolute value(s) of the δ parameter(s). Positive values of δ are interpreted as steady attenuation effects, while negative values for δ parameter(s) are interpreted as an oscillating attenuation effects. Absolute values of δ parameters approaching 0 represent rapidly attenuating effects. Absolute values of δ parameters approaching 1 represent delayed attenuating effects. As with the ω parameters that represent initial effects, multiple δ parameters for attenuation can appear, each representing attenuations over a different lag.

A more mathematical treatment of these concepts is given in Appendix B, along with the statistical methods used to select the most parsimonious, best-fitting models for each age-clinic series. All model identifications, estimation and diagnostic checking were performed using the PROC ARIMA commands within SAS version 9.2.

III.C. RESULTS

For patients whose charts were reviewed, Figure IV.1 shows the percentage with current AAPs generated by hand (paper), electronically and overall, for each age-site combination. Once the tool became available, the percentage of patients with electronically generated AAPs steadily increased, with few differences among clinics in the uptake rates (Figure IV.2).

The effect of introducing the tool on the overall prevalence of AAPs depended on the prevalence prior to introducing the tool. When pre-intervention AAP prevalence rates were low, they began to climb almost immediately after the tool was introduced, an effect that is visually apparent on the graphs. Such striking visual results do not require sophisticated ARIMA modeling because the nature of the effect is unequivocal. When pre-intervention AAP

prevalence rates were higher, it is more difficult to discern an effect of the tool, and visual interpretation of the graphs is not definitive. It is in these latter situations that ARIMA modeling is particularly useful, because the models are able to identify effects of the interruption that may not be visual.

Table III.2 shows, for each of the six weekly series, the ω and δ coefficients representing the initial (ω) and attenuating (δ) effects of the intervention (and in the case of the pediatrics clinics, the effects of the 3-week pilot during which data were lost), p-values for the coefficients, and overall goodness-of-fit statistics, for the most parsimonious, best-fitting, interpretable ARIMA models that generated residuals most closely approximating white noise.

The table is introduced by first interpreting results for adults at FM2 where the graphs demonstrate visually that the intervention had an effect. Inclusion of the intervention term improves the model fit; both numerator and denominator coefficients are statistically significant. “Delay of 3 weeks” means that the effect begins to manifest three weeks after the intervention date. The ω coefficient for the initial effect is 2.4, meaning that there was an increase of 2.4% each week in the percentage of adult patients at FM2 with current asthma action plans. The δ coefficient for the attenuation effect is 0.26; it appears at lag 3. According to theory, as the absolute value of the δ coefficient approaches zero, the initial effect dampens more rapidly [125]; thus in this series, the dampening effect occurs soon. Lag 3 means that the dampening percolates through the system at 3-week intervals.

Models for the remaining series are interpreted similarly. Because the effect is visually evident for adults and children at FM2 and for adults at FM1, the remainder of this section interprets the results for the three series in which visual inspection is insufficient to detect an effect: Children at the two pediatrics clinics and at FM1.

For children at FM1, the ω coefficient for the initial intervention effect is negative, but is not statistically significant. The δ coefficient for the attenuation effect is significant ($p<0.001$) and also close to 1, meaning that attenuation of the initial negative effect is slow; the effect persists. However, in the absence of a statistically significant initial effect, it is difficult to interpret an attenuating effect. A conservative conclusion is that for children at FM1, the intervention either had no effect, or that the effect was negative.

At PED1, there is a negative effect attributed to the records lost during the pilot period. The overall model fit improves with the intervention term. The ω coefficient for the initial intervention effect is not statistically significant; δ coefficients at lag 2 and lag 7 for attenuating effects are significant, suggesting that an effect, if there is one, is highly complex[125], oscillating at short time periods (lags=2), and attenuating over a longer time period (lag=7), although the latter effect is weak ($p=0.077$). Again a conservative conclusion is that for children at this clinic, the intervention had no effect.

At PED2, the ω coefficient for the initial intervention effective is positive, and statistically significant ($p=0.012$), meaning that the percentage of patients with AAPs increased after the tool was introduced. The δ coefficients at lags 2 and 3 for the attenuating effects are negative and statistically significant, which means that the effect oscillates [125]. An interpretation is that at this clinic, the availability of the tool resulted in increased percentages of patients receiving AAPs, although the persistence of the effect is unclear.

III.D. DISCUSSION.

The project's clinical leaders believed that written asthma action plans, individualized to the patient [126], that physicians provide as a component of self-management education [127], facilitate patient self-management, which in turn improves patient outcomes.

A quality goal at HCMC was for all its asthma patients to have current AAPs, and for all of the AAPs to be generated by the electronic tool. The organization embraced the latter goal for a number of reasons. Tool-generated AAPs reliably contained all the components deemed necessary (in a small sample of paper-template plans completed manually, less than 5% contained all components). Using the tool exposed clinicians to guideline-recommended treatment. Because use of the tool was easily ascertained by querying its underlying database, it would be pragmatically easier to fulfill external reporting requirements regarding the percentage of asthma patients with current AAPs.

If HCMC asthma patients did not have tool-generated AAPs, the question was whether providers were declining to use the technology, or whether they were declining to create AAPs. The answer to this question would influence the strategy to attain the locally normative goal that all asthma patients would have current asthma action plans generated by the tool. If patients had current paper-generated plans, but not ones generated by the technology, the strategy would focus on encouraging users to adopt the technology. If patients did not have plans at all, generated either through paper or technology, the strategy would focus on encouraging users to create plans.

The results show no resistance to technology. For patients who had AAPs before the tool was introduced, rates of paper plans dropped as rates of electronic plans increased, suggesting that paper was being abandoned in favor of computerized technology. For patients who didn't have AAPs before the tool was introduced, rates of electronically-generated AAPs began to increase as soon as the tool became available.

In groups where patients initially had paper AAPs, there was either no change in overall rates, or only a modest increase, with the introduction of the tool. If the same clinician cared

for the same patient panel over time, clinicians already creating plans may have switched from paper to electronic technology, but for clinicians not already creating plans, only some may have adopted the technology.

There is a general, parsimonious theory that explains these results: The likelihood that a patient in these clinics will receive an AAP depends on the clinician's inclination to create one, the presence of support (either a paper template or an electronic tool), and the "susceptibility" of the patient. For a susceptible patient (e.g., a patient with the appropriate asthma clinical profile, absent confounding or competing conditions), a clinician inclined to create an AAP will do so if support is available, preferring computerized technology to a paper template. A clinician not inclined to create one will not do so even if support is available.

Investigations that directly test this theory require data about AAP generation at the level of the physician, which are unavailable at this time. Were those data available, the theory would be sustained if the following theory-derived hypotheses were supported:

- In the clinics where paper AAP templates were available, rates of AAPs among patients within physician panels would be similar before and after introduction of the tool.
- In the clinics where paper AAP templates were available, persistence in paper-based AAPs after introduction of the tool would be limited to patients receiving care from certain physicians.
- In the clinics were paper AAP templates were not available, and in which introduction of the tool was associated with a steady rise in the percentage of patients with current AAPs, investigation at the physician level would show that the increase is due to increased percentages of patients with AAPs within certain physicians' panels, with no

new patients receiving AAPs in other physicians' panels.

Reasons for physician inclination to create AAPs, while interesting, is beyond the scope of this research at this time.

Three published reports also evaluated the effects of interventions on the percentage of patients with AAPs [128-130]; in two, the intervention included computerized guidelines. All three studies employed clustered randomized designs, and compared outcomes in intervention and control groups at a single post-intervention point in time. Effects in all three studies were equivocal, but for different reasons.

In one study, unexpectedly large differences between intervention and control groups in AAP baseline rates rendered post-intervention effects impossible to detect[129]. In another study, practices randomized to computerized guidelines at the point of care didn't use the system as expected, which investigators attributed to "…the challenge of integrating the systems into clinical encounters where busy practitioners manage patients with complex, multiple conditions"[128]. In the third, intervention effects were observed in suburban but not urban pediatric practices, differences which investigators could not explain[130].

The results in this study, from four clinics and two age groups, showed different initial rates of AAPs among patients, different trends over time, and different effects over time from introducing the CDS tool. Because the study design explicitly considered context and evaluated results over time, these findings could be explained by the same basic theoretical principles. Studies which fail to account for prior clinic and/or patient differences, and only examine outcomes at single moments in time, may draw incorrect conclusions and/or overlook subtle but important effects.

This study could have been strengthened by including more age groups and more clinics

and a longer post-intervention observation period. The tool's underlying database did not permit analysis at the level of the physician, which precluded the ability to test hypotheses about physician-based rates for AAPs. Revisions of the tool and its database will permit these tests in the future.

The results and theoretical derivation suggest that computerized decision support by itself will not encourage physicians to produce AAPs, but it will make it more likely if the physician already is inclined to do so. A QI strategy to increase AAP rates in asthma patients is to help physicians inclined to create them by providing electronic support, and to target physicians not inclined to create them with interventions intended to change their plan-creation tendencies.

FIGURE III.1. WEEKLY PERCENTAGES OF ASTHMA PATIENTS WITH CURRENT ASTHMA ACTION PLANS. BY PATIENT AGE AND CLINIC WHERE VISITS TOOK PLACE

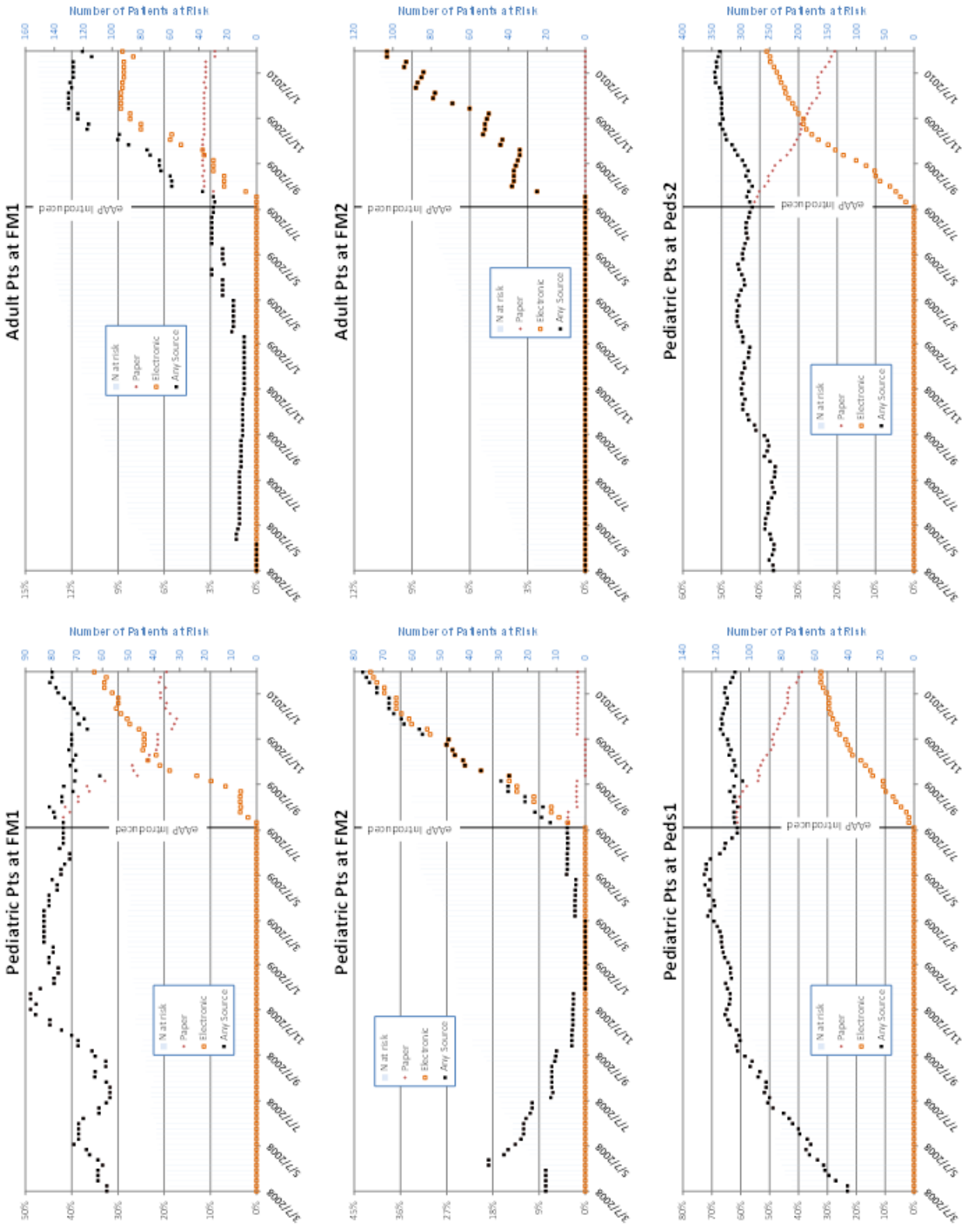
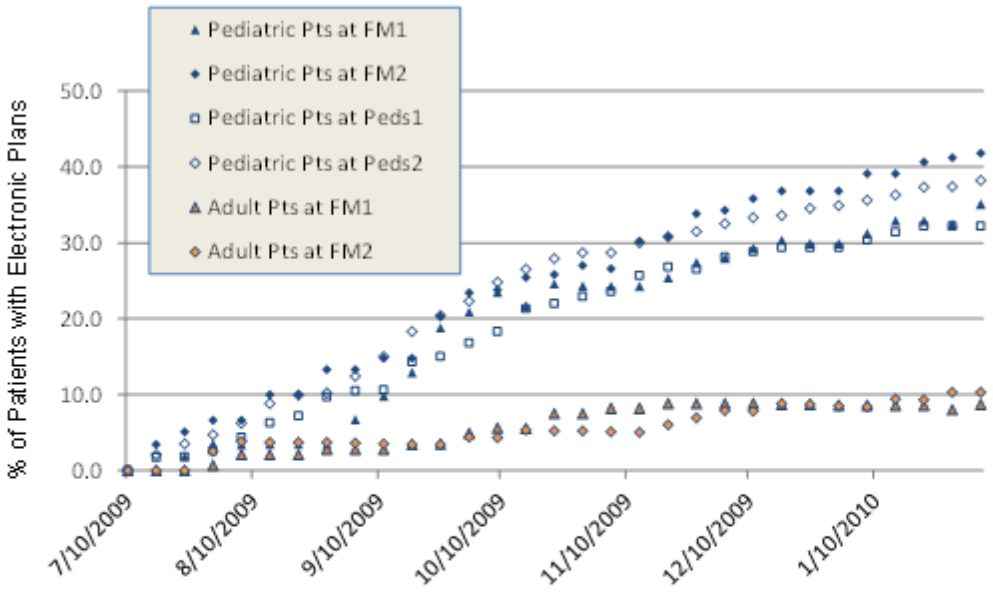


FIGURE III.2.
PERCENT OF HCMC ASTHMA PATIENTS* WITH ELECTRONIC ASTHMA ACTION PLANS
By Age Group and Clinic Where Visits Took Place.
March 2008 – February 2010.



* Patients known to have asthma on or before the observation date, with at least one visit for any reason within a year of the observation date.

TABLE III.1
NUMBER OF ASTHMA PATIENTS WITH AT LEAST ONE VISIT FROM MARCH 2007 – MARCH 2010 AT AN INCLUDED SITE AND NUMBER WITH CHARTS REVIEWED

Clinic where at least one visit took place*	School age (5-14 years)			Adult (21+ years)		
	N	Charts reviewed		N	Charts reviewed	
		N	%		N	%
FM1	304	93	30.6	908	184	20.3
FM2	134	79	59.0	225	117	52.0
Peds1	296	161	54.4	4	0	0.0
Peds2	1268	450	35.5	31	0	0.0

* Patients with visits at more than one site are included in the counts of all sites where visits took place.

TABLE III.2
ARIMA MODELS FOR WEEKLY POPULATION PREVALENCE RATES FOR AT-RISK ASTHMA PATIENTS WITH CURRENT ASTHMA ACTION PLANS

1a. Patients Seen at Family Medicine Clinic 1 (FM1)

Terms tested for effect	Patients Age 5-11 at Time of Visit: Term inclusion		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: Coeff	Term for interruption: p-val	Patients Age > 21 at Time of Visit: Term inclusion		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: Coeff	Term for interruption: p-val
Effects of the "interruption"	Delay: 4 wks	ω_0	--	--	-0.6014	0.49	Delay: 3 wks	ω_0	--	--	0.8590	<0.001
		δ_1	--	--	0.9344	<0.001		δ_1	--	--	0.7923	<0.001
Overall goodness of model fit	AIC		188.9		158.0		AIC		43.0		27.6	
	SBC		196.4		168.0		SBC		56.1		43.7	
Comments	The noise model contains an autoregressive term at lag=1 that does not reach conventional levels of statistical significance but is included because absence generates autocorrelation in the residuals. The term for the initial effect at lag 0 (ω_0) is included because model fit deteriorates without it.						Model contains six level-shifting outliers and one additive outlier. Without an intervention term, there is an additional level-shifting outlier at observation 23; its inclusion renders the denominator term insignificant and reduces model fit. Final model is shown without it.					

1b. Patients Seen at Family Medicine Clinic 2 (FM2)

Terms tested for effect	Patients Age 5-11 at Time of Visit: How the term is included in the model for the series		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: Coeff	Term for interruption: p-val	Patients Age > 21 at Time of Visit: How the term is included in the model for the series		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: Coeff	Term for interruption: p-val
Interruption terms	Delay: 1 wk.	ω_0	--	--	3.3063	0.003	Delay: 3 weeks	ω_0	--	--	2.4045	<0.001
		δ_1	--	--	0.7162	<0.001						
		δ_2	--	--	-0.4827	0.06		δ_3	--	--	0.2617	0.04
		δ_3	--	--	0.7077	<0.001						
Overall goodness of model fit	AIC		178.2		138.3		AIC		50.1		46.0	
	SBC		180.1		144.7		SBC		53.9		51.3	

1c. Pediatric Patients (age 5-11) Seen at Pediatrics Clinic 1 or 2

Effect of:	Patients with Visits to Peds Clinic 1: How the term is included in the model for the series		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: Coeff	Term for interruption: Coeff	Patients with Visits to Peds Clinic 2: How the term is included in the model for the series		No term for interruption: Coeff	No term for interruption: p-val	Term for interruption: β	Term for interruption: p-val
Intervention	Delay: 1 week.	ω_0	--	--	-2.2542	0.12	Delay: 1 week.	ω_0	--	--	1.8922	0.01
		δ_1	--	--	-0.9120	<0.001		δ_1	--	--	-0.7356	<0.001
		δ_7	--	--	0.6749	0.08		δ_3	--	--	-0.6432	0.004
Pilot (lost data) term (see appendix IIc for detail)	Delay: 1 week	ω_0	-2.5744	0.099	-3.7014	0.02	Delay: 1 week	ω_0	-0.9546	0.158	-0.4179	0.54
								δ_1	0.1973	0.770	0.9602	0.003
Overall goodness of model fit	AIC		208.7		164.9		AIC		128.9		116.9	
	SBC		210.6		171.5		SBC		132.6		125.6	

CHAPTER IV. ETHICAL ISSUES IN COMPUTERIZED DECISION SUPPORT FOR CHRONIC DISEASE MANAGEMENT

IV.A. INTRODUCTION

This chapter identifies ethical quandaries that emerged during a project to build a guideline-based computerized decision support (CDS) tool for ambulatory asthma care, where the recommendations and clinical content were derived from the Third Expert Panel Report (EPR-3): Guidelines for the Diagnosis and Treatment of Asthma [1]. The EPR-3 adheres to currently prevalent standards of guideline quality [77, 131, 132] and is held in high esteem. Ethical quandaries emerged when project's clinical leaders attempted to adhere to the EPR-3, by presenting all its recommendations on the tool's screen, and by only presenting on the screen recommendations that appeared in the EPR-3. Instances arose in which such faithful adherence had the potential to harm some patients, and to deprive others of benefit.

IV.B. ETHICAL ISSUES WHEN CDS TOOLS ARE "EVIDENCE-DISSEMINATORS"

When developing a guideline-based computerized decision support tool from a document-centric point of view, the purpose of the tool is to re-represent guideline recommendations in the form of executable code. Guideline developers base their recommendations on evidence. Like other authors of guidelines considered to be "high-quality", the authors of the EPR-3 explicitly rank levels of evidence, using prevailing standards of quality[133].

A. Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Requires substantial numbers of studies involving substantial numbers of participants.

B. RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

C. Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

D. Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

Although the EPR-3 states that most of its recommendations are based on evidence falling into categories B and C, because few of the studies the panel reviewed employed designs other than RCTs, it may be more accurate to say that the evidence is predominantly concentrated in categories B and D.

The Expert Panel communicates using a variety of human-readable formats, ranging from extensive narrative discussion to a condensation of key points in succinct tables. Because extensive narration is difficult to digest during the delivery of clinical care, most re-representations of guideline recommendations take the form of the succinct tables. For example, the complete discussion of daily medication management consumes more than 100 pages in the report, which the EPR-3 summarizes in six half-page tables. Each table condenses recommendations for initiating or continuing treatment, for three age groups (0-4, 5-11, 12+), over six “steps”, or levels of therapeutic aggressiveness. Within each step is a set of tersely worded treatment plans, expressed as combinations of one or more pharmacological agents. To illustrate, Figure IV.1 is the EPR-3 table for managing asthma in patients age 5-11 years.

The tables only are useful to practicing clinicians who know how to populate a “treatment plan” with commercially available products that fulfill it. Because most primary care physicians do not have this knowledge, the CDS tool developed in the HIT Asthma project helped the user populate a selected plan by presenting in drop-down lists the commercially available products,

routes of administration, strengths and instructions for use that would fulfill it.

IV.B.1. Ethical challenges associated with representing every EPR-3 plan.

Approximately 90% of the more than 1200 studies cited in the EPR-3 were randomized controlled trials, or RCTs. Very few RCTs enroll babies and young children, age 0-4, so there was a paucity of direct evidence about daily medication management for either efficacy or effectiveness in this vulnerable sub-population. The Expert Panel based its recommendations for treatment in this age on extrapolations from RCTs conducted in older children and adults. In the narrative explaining its reasoning, the Expert Panel encourages readers to use judgment when treating patients this age, and to think of individual treatment as a therapeutic trial: Give a limited dose, watch closely to see what happens, and then adjust.

However, in the condensed table of recommendations, this nuanced discussion is absent; what remains are lists of treatment plans derived from extrapolation. Direct conversion from the tables to the CDS screen presented users with treatment plan choices for which there was no direct evidence of either efficacy or effectiveness. And direct conversion from the selected plans to drop-down lists of products that fulfilled them presented users with choices of commercially available products that did not have FDA approval for use in this age.

As a pediatric pulmonary specialist, the lead clinician/developer knew from experience that these on-screen recommendations could be ineffective or even harmful if adopted by a physician accustomed to treating adults in general primary care, and the narrative in the guidelines confirmed her views. Yet to faithfully represent all recommendations in the tables, the CDS tool would have to present such choices on-screen, however clinically questionable.

IV.B.2. Ethical challenges associated with not representing choices the EPR-3 didn't address.

A commonly recommended treatment plan in the EPR-3 was the combination of an

inhaled corticosteroid (ICS) and a long-acting beta agonist (LABA). Among clinicians at the project site and across the site's metropolitan area, there was broad agreement that adding Montelukast to an ICS-LABA combination conferred additional benefit for asthma patients who also had respiratory allergies[*]. The EPR-3 was silent about the Montelukast-ICS-LABA treatment plan. There were no studies in its evidence tables that addressed the question, and none of the studies were restricted to asthma patients who had respiratory allergies.

Omitting support for this plan on the computer screen was not necessarily indifferent or benign. Providers who believed in the Montelukast-ICS-LABA plan would continue to use it, with or without tool support. The tool's clinician/developers were concerned that once their colleagues realized that the tool did not support a therapy which they believed offered benefit, they would conclude that the tool was useless and refuse to use it at all, thus depriving themselves and all their patients from its potentially helpful guidance[†]. Alternatively, they might use the tool to produce an ICS-LABA treatment plan, and write in the Montelukast by hand, sending the patient home with a document that differed from the electronic health record, causing potential confusion or even harm.

IV.C. EPISTEMOLOGICAL PREFERENCES AND RESOLUTION OF ETHICAL QUANDARIES

Ethical quandaries about presenting choices on-screen for patients age 0-4 were handled by flagging the troublesome selections with warning boxes such as "this treatment has not been tested in children this age", or "this product has not been approved for use in children

[*] Montelukast is frequently used alone to treat respiratory allergies.

[†] Invocation of the tool required a change in workflow during the time-pressured ambulatory environment. Workflow changes tend to be adopted in an all-or-none fashion, because there isn't time to think about which workflows to use given the particular characteristics of the patients being seen.

this age". The solution faithfully represented all guideline-recommended options, and deflected responsibility for handling problematic selections to the physician-user.

It was technically straightforward to do the same for the promising treatment plan in widespread local use, but which the EPR-3 didn't mention. The promising plan could have been presented on the screen, along with a warning box such as, "This plan is not mentioned in the guidelines, but local clinicians believe it works for patients with respiratory allergies". As above, the on-screen choices would have faithfully communicated guideline recommendations, tagging a choice not so recommended with a warning that deflected responsibility to the clinical user.

Unlike the technical solution to the first problem of missing evidence, the same technical solution to different problem of missing evidence was not acceptable. Early collaborators at the state's Department of Health objected on the grounds that "The tool has to follow the guidelines"; anything that wasn't in the guidelines could not be included in the tool. Project sponsors and clinical leads objected because the intent was to disseminate the tool to other sites. If the locally popular but unmentioned plan disseminated with the tool, this would be equivalent to promoting (or at least supporting) a treatment for which there was no evidence meeting the standards that the guidelines developers expressed.

A technical solution to the latter concern was to equip the tool with an administrator's function that would enable a local clinical authority (such as a site's medical director, or director of ambulatory care quality) to select which treatment plans would appear on the screen. Plans recommended in the guidelines would appear as defaults; plans selected by the administrator would be flagged as such. Provider choices of treatment plans and patient outcomes could then be analyzed to identify promising treatments emerging from

experimentation in clinical practice.

Ethically, this solution was defensible if the CDS tool was conceived not just as a disseminator of evidence generated primarily from RCTs, but also as a collector of evidence generated from clinical practice. The ethical justification uses the same principles that support randomized trials (potential risks to some are outweighed by the promise of benefit to all), and also the principles that support Medicare's policy of "coverage with evidence development"[134, 135][*] (potentially beneficial treatment will not be withheld, but individuals who avail themselves of it participate in the production of evidence).

Project leaders, sponsors and early collaborators at the state's Department of Health rejected this argument on the grounds that such evidence about effectiveness would not have been generated by an RCT. Their objection was consistent with the hierarchy of evidence detailed in the EPR-3. According to that hierarchy, it was epistemically defensible to base recommendations for babies and young children on findings from RCTs conducted in older children and adults; it was epistemically less defensible to equip the tool to collect practice-based outcomes associated with treatment that had not been studied in an RCT.

This hierarchy and the decision rules it supports ignore patient risk and/or benefit. There was a non-trivial risk that treating babies the same as adults could harm the babies. There was little risk of harm from the Montelakast-ICS-LABA treatment plan. Montelakast already was an approved treatment for respiratory allergies; ICS-LABA was an approved and commonly recommended treatment for asthma. It made clinical sense to use the three-way combination to treat patients who had both respiratory allergies and asthma, and anecdotally, benefit had

[*] These policies offer coverage for promising but untested technologies, under the condition that beneficiaries receiving such coverage enroll themselves in studies that evaluate effectiveness.

been observed. There was a greater risk associated with not supporting it with the tool.

In other words, granting epistemic preference to the RCT over other forms of evidence had the potential to cause harm to some patients and to deprive others of benefit.

IV.D. AN EPISTEMOLOGICAL CRITIQUE OF THE RCT

Philosophers of science [136-138], health policy scholars [14], ethicists [139, 140] and physicians [141-143] collectively have generated a number of concerns about the epistemically privileged status that the RCT enjoys among advocates of evidence-based medicine, particularly in the case of chronic disease [139, 144]. Their concerns fall into two related categories. One set addresses the limited applicability of RCT findings to the delivery of ordinary medical care to ordinary patients; these are concerns about external validity [145]. Although low external validity often is framed as the necessary price to be paid for high internal validity [146], a second set of epistemologically more sophisticated concerns addresses the ability of RCT methodology to address questions about treatments extended over time, in which the interaction between context and treatment cannot be ignored. These concerns are germane for chronic disease, which are axiomatically characterized by chronicity.

A third concern, derived from the first two, is that by privileging evidence from RCTs, the only questions that tend to be entertained are ones that can be addressed with RCTs. Since the results have limited applicability to primary care or to chronic disease, most of the questions that primary care physicians have about their patients with chronic disease cannot be answered with existing evidence [147].

RCTs struggle with external validity because the more tightly controlled are the circumstances under which the trial is performed, the more narrow the range over which results can be applied. Control reduces within-group variability, so that between-group effects

are easier to detect with sample sizes small enough to be economically feasible. Control is achieved by homogenizing study populations with strict inclusion and exclusion criteria, and by ensuring that both participants and providers adhere strictly to study protocols.

In ordinary clinical care, patient populations are diverse, containing individuals who do not resemble the homogenized populations selected for a trial. One British study found that 90% of patients with asthma would not have been eligible for the RCTs that tested the efficacy of the medications they were using [148]; similar results have been found for patients with chronic obstructive pulmonary disease [148] and hypertension[149]. Also in ordinary clinical care, it is seldom possible to enforce a strict protocol. Providers cannot offer treatment for free, seldom follow up with the rigor that study protocols demand, and will often adjust therapies experimentally to determine combinations that have the optimal individual effects.

When treatment is for one or more chronic disease, which persists indefinitely over time, findings from short trials may be irrelevant. Although trials could be lengthened, cost is a barrier, and trials which require adherence to study protocols established at inception inevitably become pragmatically and ethically corrupt the longer they are underway. Pragmatically, more participants may be unable or unwilling to remain adherent; study physicians may change who have less knowledge and/or commitment to initial protocol. Ethical troubles emerge if protocol adherence conflicts with participants' ability to use newly discovered therapies.

Although these pragmatic and ethical problems can be addressed by relaxing protocols, relaxation eliminates those features that make RCTs successful, and introduces another set of epistemological concerns. Trials that do not require strict adherence, but which are framed as an intervention compared to usual care suffer from erosion of validity over time, as what

constitutes “usual care” will inevitably change. As it changes, it becomes increasingly difficult to know what the comparison to the intervention actually is. (Appendix C describes how these issues afflicted the world-renowned Multiple Risk Factor Intervention Trial).

These issues are widely prevalent when the objective is to learn how best to treat a patient with one or more chronic disease. There is no expectation of cure, thus there is no definitive endpoint that can be used to compare one randomly allocated group to another. Rather, there is the hope that the condition will be managed successfully enough to enable the patient to function normally, to experience a high quality of life, and to prevent or forestall a debilitating or life-threatening event. Pharmacological management may be, and often is, a lifelong experience, which can induce iatrogenic effects. These then must be managed, particularly when other diseases arise, which also call for pharmacological management. Over time, a patient may be taking several different drugs. To meet regulatory requirements, each drug may have been tested for safety and efficacy independently, but the combination used by one particular patient may never have been.

For the clinician caring for a patient with one or more chronic disease, the relevant questions are not so much whether this particular drug works, or works better than another, under a set of controlled circumstances limited in time. Rather, clinicians want to know which therapeutic regimen that may include multiple drugs will, over a perhaps extended period of time, give the patient the highest quality of life, within the patient’s constraints. Those constraints include other medications, treatment preferences, as well as the patient’s physical, cognitive, financial ability to execute the prescribed regimen [136, 139, 144]. RCTs that are designed to test the safety and efficacy of one particular drug or regimen over a circumscribed period of time cannot generate this kind of evidence.

IV.E. USER-CENTRIC CDS FOR CHRONIC DISEASE AND A HEALTHCARE THAT LEARNS

It is presumably for these reasons that a Roundtable on Evidence-Based Medicine convened by the Institute of Medicine questioned whether the RCT should remain endowed as the "gold-standard" for evidence, and also questioned the entire approach towards generating both evidence and guidelines. Contributors to a Roundtable Workshop on EBM stated that as intervention options become more numerous and complex, and as more is learned about heterogeneous responses to therapy, "… the prevailing approach to generating clinical evidence is impractical today and may be irrelevant tomorrow". Roundtable members envisioned a future in which evidence is generated as a by-product of the normal delivery of care, and is fed directly back into the care process through decision support, although they had no tested, pragmatic solutions for how this might be accomplished [150, 151], or even working prototypes. This chapter suggests that in the case of chronic disease, the vision could potentially be manifested through the use of CDS tools that embed evidence into the choices appearing on screen. The same tools retain in their underlying databases both treatment choices and outcomes, which potentially generate data that support the creation of a new and evolving evidence base.

For one condition (asthma) and for one CDS tool, the technical challenges to implementing this vision are minor. Other challenges and barriers, however, exist; some are methodological, others are social. The IOM Roundtable members identified several challenges, but others arose during the HIT Asthma project which Roundtable members missed.

The barrier that draws most attention is the privileged status that RCTs currently enjoy among advocates of evidence-based medicine, and the still nascent methodologies for rigorous alternatives that rely on data generated through natural experiment. Alternatives to

RCTs are increasingly being developed and proposed, which alter criteria for evidence acceptability, taking into account effectiveness and not just efficacy [152], or assessing external validity [153, 154], or balancing internal and external validity [155, 156]. Federal sponsors are inviting proposals to develop methodologically sound processes [157] to analyze clinically rich observational data that are produced as byproducts of delivering care. While methodological challenges and developing appropriate analytic technique will always persist, resources are currently being directed towards improvement.

A deeper and potentially hidden social barrier exists within the informatics community. For more than twenty years, the community has embraced a document-centric vision of CDS, and has dedicated itself to finding ways to automatically translate narrative guidelines into executable code [12, 158, 159]. This is a large and expensive undertaking to which many have committed their careers. The Agency for Health Research and Quality has provided millions of dollars to fund the effort, and abandoning it would be difficult. Yet here it is argued that abandonment is necessary if the Roundtable vision is to manifest. Too many resources in the form of dollars and human talent are currently locked into a venture that the most knowledgeable observers say is unlikely to succeed [7, 104, 160], primarily due to its reliance on heavily engineered technology.

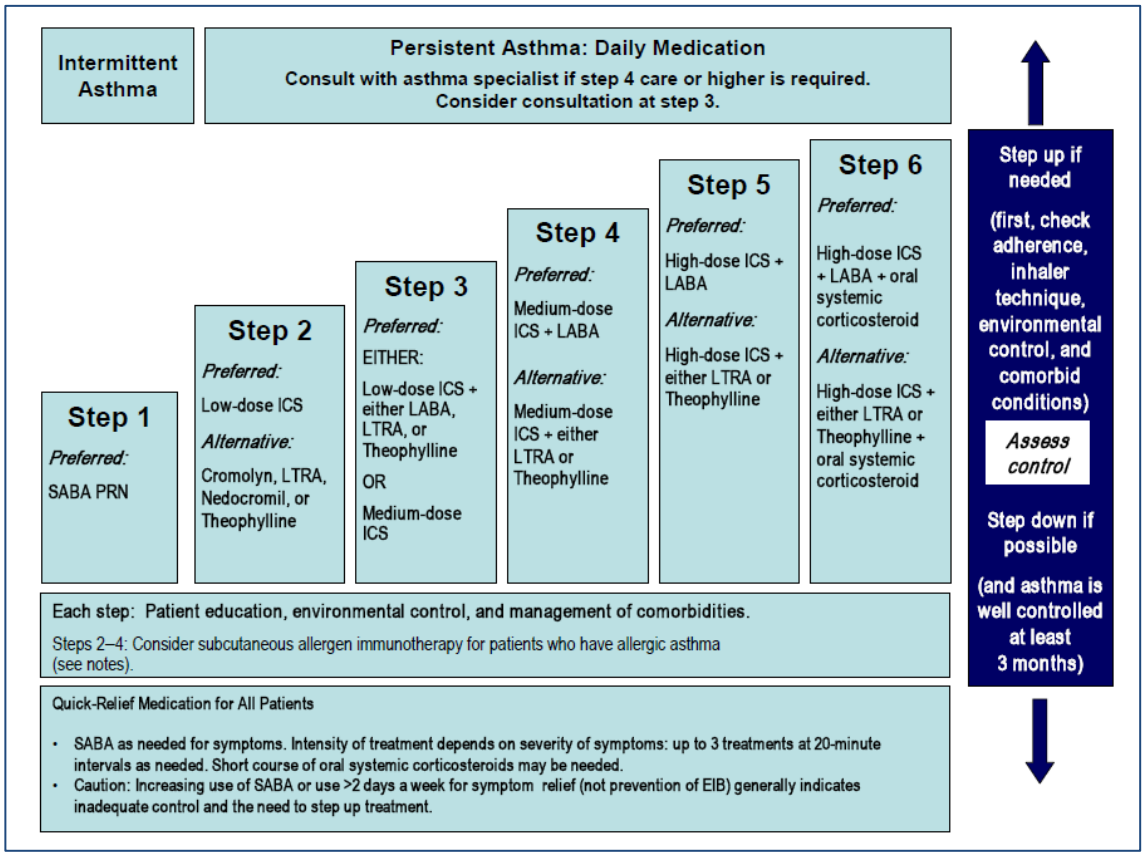
The experience of the HIT Asthma project shows that it is possible to create CDS tools using technologically simple web-based languages in widespread use. Doing so would vastly expand the pool of developers capable of writing the code, unleashing a flood of talent that collectively could “computerize” hundreds of guidelines currently waiting until the automatic conversion languages are perfected enough to be usable. That could be a long wait, and physicians are impatient. They want something that will help them now. Using ubiquitous

web-based coding languages also would shrink the challenge of local implementation to the simple task of installing software on a web server which can be invoked using strategically placed hyperlinks.

An additional barrier concerns the business model to support the proposed enterprise. There must be a mechanism to recover the financial costs associated with maintaining both the content and the functioning of CDS tools that serve as practice guidelines unto themselves. These questions fall outside the scope of this paper. To address them, it is necessary to consider the role of professional societies that currently support guideline development, intellectual property rights for CDS tools in which clinical content appears in publicly available literature, reimbursement models that reward quality and how quality is assessed.

Although the latter issues are formidable, so has been the social and methodological challenge of generating and accepting evidence from sources other than RCTs. Yet with attention and dedication there is the possibility that methodologically sound alternatives to RCTs will emerge, and that with their emergence, social agreement about the acceptability of practice-based evidence will be attained. With similar attention and dedication it may be possible to redirect the informatics enterprise towards more promising avenues of CDS tool development, and also to create fiscally sustainable models of the “health care system that learns”. Such attention is warranted because without it, development of CDS tools will continue to be afflicted with the ethical quandaries that inevitably arise if they only are conceived as translations of guidelines based on evidence derived primarily from RCTs.

FIGURE IV.1. EXAMPLE OF A FIGURE IN THE EPR-3 SUMMARIZING RECOMMENDATIONS FOR PHARMACOLOGICAL MANAGEMENT OF ASTHMA



CHAPTER V. DISCUSSION.

V.A. OVERVIEW

The HIT Asthma project was a user-driven effort to provide physicians with the computerized support they require to perform administrative tasks and make clinical decisions to deliver high quality asthma care. The effort yielded a computerized decision support tool with clinical content derived from the most recently released and widely cited guidelines for asthma in the US. Because medical informaticists in the US and Europe have been creating "computerized guidelines" for chronic disease for more than twenty years, with respect to using technology to deliver guidelines at the point of care, there was nothing novel about the project's objectives. Nor was there anything novel or controversial about its overall goals, which was to improve the quality of asthma care by providing physicians with just-in-time information about the most recently developed evidence-based recommendations available.

The novelty was in the approach. In the HIT Asthma project, the technology's requirements were specified by the clinicians who would be using the tool, and working clinicians were involved at every step of development. By contrast, informaticists who create guideline-based computerized decision support for chronic disease typically take a "guideline-centric" approach, focusing more on the guidelines to be computerized than the users who are the intended beneficiaries. Research has shown that implementation of complex guided computerized support for chronic disease has had only a limited effect on process measures for quality of care or on patient outcomes, and deeper probing reveals that most failures to detect effects can be attributed to users' rejection of the tools. Users have complained that they are awkward or otherwise unpleasant to use, and don't offer support they find valuable.

A reasonable hypothesis, supported by theory in cognitive ergonomics, is that if a CDS tool

were designed to explicitly meet user-expressed needs, barriers towards adoption would fall, and that guideline-adherence objectives would be achieved. The results shared in Chapter III sustain this hypothesis. At the site where clinical users led the development of the tool, it was adopted quickly by many physicians. This Dissertation is being prepared approximately 18 months after the tool became available, and in that time, it has been implemented at two different healthcare delivery sites with numbers of invocations rising each month. Negotiations with a third site are underway, and interest has been expressed by more than 500 clinicians from approximately 100 healthcare delivery sites across the US.

Besides the improved user acceptance of the technology, other insights emerged from the user-centric approach. These insights are aligned with the vision articulated by the Institute of Medicine (IOM) Roundtable on Evidence-Based Medicine, which stated that the nation needs "A healthcare system that learns". A healthcare system that learns is one in which evidence is conceived as "system experience", and delivery is conceived as collective systemic behavior that adjusts as a result. According to IOM Roundtable members, to achieve this vision, evidence must be both developed and applied as a natural product of care delivery [14].

The HIT Asthma project suggests that user-centric, technologically simple, computerized decision support tools to manage chronic disease can become prototypes for the learning system that IOM Roundtable participants imagined, but currently does not exist.

V.B. NOVEL INSIGHTS AND ILLUSTRATIONS OF A "LEARNING HEALTHCARE SYSTEM"

V.B.1. There is an ethical imperative to reconsider the privileged position of RCTs

The HIT Asthma project found gaps between the clinical scenarios typically faced in ambulatory primary care, and the evidence upon which guideline recommendations are based. A central theme from the IOM Roundtable Workshop on the Learning Healthcare System is

that such gaps are common, and emerge from the cultural gulf between the health research and practice communities.

In the HIT Asthma project, these gaps generated conflicts for the tool's physician-developers. Their respect for the authority of the guidelines, and those who developed them, compelled them to present all guideline recommendations on the screen of the CDS tool, and to only present guideline recommendations on the screen. But as described in Chapter IV, because evidence was absent or equivocal, instances arose in which such faithful adherence had the potential to harm some patients and deprive others of benefit. The clinician/developers thus faced a conflict between their respect for and adherence to the authority of the guidelines, and their ethical obligations to patients.

In the argument developed in Chapter IV, this conflict could have been resolved by using the tool not only as a means to disseminate evidence obtained primarily from RCTs, but also as a means to generate evidence, obtained from records of treatment decisions and patient outcomes in the tool's underlying database. The solution was ethically defensible, technically feasible, and directly addressed a concern repeatedly emphasized by the IOM Roundtable: "Because RCTs are not practical, ethical, feasible, or appropriate to all circumstances, large data systems with built-in study design and feedback loops allow for investigations that have real rigor, utility and reliability in large populations. These systems will also provide a way to close the loop by identifying relevant order sets, tracking order set utilization, and routinely feeding this performance data back into order set development."[161]

The Workshop participants were speaking of order sets employed in large, enterprise-wide EHR systems, but the concept is the same. If hosted by an organization with the necessary resources and clinical authority, new evidence gleaned from analyzing results collected in CDS

tools' databases could automatically be incorporated into new versions of the tools, which could be distributed back to users with the same technical mechanisms that Java and Firefox use when they announce that "updates are ready to be installed on your computer".

However, in the HIT Asthma project, this solution was rejected on the grounds that such evidence would not have been generated by randomized clinical trials (RCTs), despite the fact that it was nearly exclusive reliance on RCTs that caused the conflicts to emerge. This study design occupies the most privileged position in the hierarchy of evidence, although it is not well-suited to addressing the most difficult and important questions about chronic disease care. Its privileged position, however, means that most studies cited by guideline developers employ RCTs. It could be argued that privileging the RCT is thus ethically problematic because such a practice discourages the development of evidence using other means, even when other means are more defensible on both ethical and epistemological grounds.

V.B.2. Interrupted time series analysis is a powerful tool for chronic disease research

One of the valid criticisms applied to analysis of observational data is that because allocation to treatment is seldom randomized, it is not possible to use conventional statistical methods to analyze the results. In the workshop convened by the Institute of Medicine, Soumerai argues that interrupted time series offers feasible and rigorous alternatives [151]. The method addresses many threats to validity, such as competing sources of effects and auto-correlation of results over time. Graphs of the results often convey an intuitive understanding of intervention effects.

Although Soumerai restricted his remarks to using interrupted time series to understand the effect of policy decisions, insights from the HIT Asthma project suggest that the design is also a feasible and rigorous method to evaluate population-level effects of clinical

interventions introduced at multiple sites.

In this study, interrupted time series met physician needs. The clinicians who directed the HIT Asthma project had a normative agenda. They wanted patients to have AAPs. Once the tool was in use, they also wanted the patients to have AAPs generated electronically. There is an established methodology for pursuing quality objectives in health care, disseminated by the internationally-known Institute for Healthcare Improvement (IHI). One of the central components of that methodology is statistical process control, in which outcomes are measured and monitored continuously over time[162]. To generate graphs of results over time that would be intuitively meaningful to the clinicians, we plotted the weekly percentage of asthma patients with current asthma action plans, and marked the plot at the date the tool was introduced. For patients treated at some clinics in some age groups, the impact of the tool was visually unmistakable. For patients treated at other clinics in other age groups, the graphs were less informative, and analysis using the Box-Jenkins ARIMA models for interrupted time series resolved the uncertainty.

ARIMA modeling and interrupted time series analysis has been applied only sporadically in health outcomes [163, 164], health services [165], and epidemiological research [166, 167], but when applied, the results are generally illuminating, offering insights that time-static methods are unable to provide. Results are intuitively meaningful for clinicians, presentations are consistent with recommendations from IHI about measuring outcomes over time, and the approach is among those endorsed by the IOM Roundtable on Evidence Based Medicine.

V.B.3. The CDS enterprise in medical informatics requires a revision of goals

The informatics literature contains abundant documentation that practice guidelines are vague, ambiguous, logically incomplete and internally contradictory. In reports of CDS

endeavors, informatics authors often mention these guideline properties as "usual" [7, 15, 22, 107, 168]. While some informaticists refer to these guideline features as "deficiencies", and call for their elimination, others are more thoughtfully circumspect, and express awareness that vagueness is an intentional use of natural language* to deflect attention away from gaps in the evidence [7, 15]. An experienced informaticist writes: "…Guidelines intended to be based on evidence will necessarily run into absence of evidence, particularly for all but the theoretically simplest cases. Yet it is exactly the non-simple cases that physicians would find guidelines most helpful"[15].

In other words, thoughtful observers within the informatics community are aware that practice guidelines are often based on scant and unhelpful evidence. Despite this awareness, the informatics approach towards developing CDS tools is predominantly focused on methods to convert these same guidelines into executable code. This focus limits the role of CDS tools to disseminating weak and incomplete evidence, and is blind to the ethically troublesome issues that arise as a consequence. The approach inhibits the use of health information technology to manifest the vision of a healthcare system that learns.

Moreover, informed, knowledgeable observers of and participants in the informatics community long have commented that the current, heavily engineered approach towards "computerizing guidelines" poses enormous and perhaps insurmountable challenges to implementation at local healthcare delivery sites [7, 23, 34, 47, 104, 168-171]. The code is complex for both the "computerized guideline" and the sites' enterprise electronic health record systems. Interoperable standards and languages for large complex systems do not exist,

* The Introduction of the EPR-3 states that the equivocalness of the vocabulary used in a recommendation reflects Panel commitment to it, and commitment reflects the strength of the evidence on which it is based.

thus these systems cannot exchange information or functioning [104]. It is not surprising then, that in 2006, after more than twenty years of development, Sonnenberg and Haggerty reported that there was no computerized guideline system in use outside of the institution in which it had been developed[7]. By contrast, less than two years after the technologically "light" product developed in the HIT Asthma project went live, it was in use in two institutions (one where it had been developed, and one not), and negotiations were underway to implement it at a third.

In industrialized countries, the greatest population health burdens are created through chronic disease. Yet guidelines for managing chronic disease are arguably in epistemic infancy because RCTs have a limited ability to produce practically relevant evidence to manage chronic disease over time. Informatics talent and resources currently are directed towards developing heavily engineered computerized decision support tools to push this meager "evidence-based medicine" to the point of care, even though local sites cannot accept or digest the technology.

It seems reasonable to suggest that informatics talent and resources might be better deployed to building simpler, user-centric CDS tools that can also generate practice-based evidence. Such a refocusing would more quickly and assuredly achieve the goals identified in a 2008 health policy conference convened by the American Medical Informatics Association, in which participants emphasized the contributions that health information technology could make to advance the IOM Roundtable workshop on evidence based medicine. [172]

V.C. SUMMARY AND CONCLUSIONS

The user-centricity of CDS tool development led to a number of unexpected findings and insights, which were directly germane to the IOM Roundtable vision of a learning healthcare system. The Roundtable, comprised of leaders in healthcare delivery sectors across the US,

was convened in 2006 by the Institute of Medicine, which charged it to consider new ways to generate and disseminate evidence, because “The prevailing approach to generating evidence is inadequate today and may be irrelevant tomorrow, given the pace and complexity of change.” The Roundtable offered a futuristic vision, but could only hint at the technological developments that would be required to attain it.

The contemporary informatics community does not offer an immediately viable method for manifestation, because its efforts have been directed towards developing automatic methods to electronically disseminate guideline recommendations that are based on conventionally generated evidence. The technical solutions for guided decision support that have emerged from these efforts cannot easily be integrated into the commercially available electronic health record systems which Roundtable members viewed as potential generators of evidence. To support the complex decision making necessary for managing chronic disease, there is a gulf between technologies capable of generating evidence (large, commercially available EHR systems), and technologies designed to disseminate it (artificial languages that automatically convert narrative guidelines into executable code). The current informatics agenda offers no immediately practical ways to bring them together.

The HIT Asthma project demonstrated that a user-centric and low-tech approach for developing computerized decision support tools for chronic disease does offer that possibility. It is a practical prototype for a health care system that learns.

BIBLIOGRAPHY

1. *Expert panel report 3. Guidelines for the diagnosis and management of asthma. Full report 2007.* 2007, Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program.: Bethesda, MD. p. NIH Publication No. 07-4051.

2. Bracha, Y., G.M. Brottman, and A. Carlson, *Physicians, Guidelines, and Cognitive Tasks.* Evaluation & the Health Professions, 2010.

3. Heller, B., et al., eds. *Computer-Based Support for Clinical Guidelines and Protocols*. Studies in Health Technology and Informatics. Vol. 83. 2001, IOS Press: Amsterdam. 200.

4. Lenz, R., et al., *IT support for clinical pathways--Lessons learned.* International Journal of Medical Informatics, 2007. **76**(Supplement 3: Ubiquity: Technologies for Better Health in Aging Societies - MIE 2006): p. S397-S402.

5. Duff, L. and A. Casey, *Implementing Clinical Guidelines. How Can Informatics Help?* J Am Med Inform Assoc, 1998. **5**(3): p. 225-226.

6. Blaser, R., et al., *Improving pathway compliance and clinician performance by using information technology.* MIE 2003, 2007. **76**(2-3): p. 151-156.

7. Sonnenberg, F. and C. Hagerty, *Computer-Interpretable Clinical Practice Guidelines. Where are we and where are we going?*, in *IMIA Yearbook of Medical Informatics 2006*, R. Haux and C. Kulikowski, Editors. 2006, International Medical Informatics Association. p. S145-58.

8. Grimshaw, J. and I. Russell, *Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations.* Lancet, 1993. **342**(8883): p. 1317-22.

9. McGlynn, E.A., et al., *The quality of health care delivered to adults in the United States.* New England Journal of Medicine, 2003. **348**(26): p. 2635-45.

10. Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*. 2001, Washington, DC: National Academies Press.

11. Dawes, M. and U. Sampson, *Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians.* International Journal of Medical Informatics, 2003. **71**(1): p. 9-15.

12. Shiffman, R.N., et al., *Bridging the guideline implementation gap: a systematic, document-centered approach to guideline implementation.* J Am Med Inform Assoc, 2004. **11**(5): p. 418-426.

13. Gugiu, P.C. and M. Ristei Gugiu, *A Critical Appraisal of Standard Guidelines for Grading Levels of Evidence.* Evaluation & the Health Professions, 2010. **33**(3): p. 233-255.

14. IOM Roundtable on Evidence-Based Medicine, ed. *The Learning Healthcare System*. ed. L.A. Olsen, D. Aisner, and J.M. McGinnis. 2007, Institute of Medicine, National Academies Press: Washington, DC.

15. Seroussia, B., J. Bouauda, and G. Chatellier, *Guideline-based modeling of therapeutic strategies in the special case of chronic diseases.* MIE 2003, 2005. **74**(2-4): p. 89-99.

16. Bonacum, D., *Improving patient safety*, in *Connected for Health. Using Electronic Health Records to Transform Care Delivery*, L.L. Liang, Editor. 2010, Jossey-Boss: San Francisco. p. 157-178.

17. Weisenthal, A.M., *Physician leadership and engagement*, in *Connected for Health. Using Electronic Health Records to Transform Care Delivery*, L.L. Liang, Editor. 2010, Jossey-Boss: San Francisco. p. 57-68.

18. Fiks, A.G., et al., *Impact of Clinical Alerts Within an Electronic Health Record on Routine Childhood Immunization in an Urban Pediatric Population.* Pediatrics, 2007. **120**(4): p. 707-714.

19. Johnston, M.E., et al., *Effects of Computer-based Clinical Decision Support Systems on Clinician Performance and Patient Outcome: A Critical Appraisal of Research.* Annals of Internal Medicine, 1994. **120**(2): p. 135-142.

20. Shea, S., W. DuMouchel, and L. Bahamonde, *A Meta-analysis of 16 Randomized Controlled Trials to Evaluate Computer-Based Clinical Reminder Systems for Preventive Care in the Ambulatory Setting.* Journal of the American Medical Informatics Association, 1996. **3**(6): p. 399-409.

21. Chaudhry, B., et al., *Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care.* Annals of Internal Medicine, 2006. **144**(10): p. E12-W18.

22. Maviglia, S.M., et al., *Automating Complex Guidelines for Chronic Disease: Lessons Learned.* J Am Med Inform Assoc, 2003. **10**(2): p. 154-165.

23. De Clercq, P.A., K. Kaiser, and A. Hasman, *Computer-Based Interpretable Guideline Formalisms*, in *Computer-based medical guidelines and protocols: a primer and current trends: A Primer and Current Trends*, A. ten Teije, S. Miksch, and P. Lucas, Editors. 2008, IOS Press: Amsterdam. p. 22-43.

24. Peter Szolovits (ed). *Artificial Intelligence in Medicine, 1982.* 1982 January 2000 [cited 2009 January]; Overview of the major research efforts in the United States in the application of artificial intelligence techniques to medical decision making.]. Available from: http://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch0.html#authors.

25. Heselmans, A., et al., *Effectiveness of electronic guideline-based implementation systems in ambulatory care settings - a systematic review.* Implementation Science, 2009. **4**(1): p. 82.

26. Garg, A.X., et al., *Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review.* JAMA, 2005. **293**(10): p. 1223-1238.

27. Rousseau, N., et al., *Practice based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care.* BMJ, 2003. **326**(7384): p. 314.

28. Kawamoto, K., et al., *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.* BMJ, 2005. **330**(7494): p. 765-.

29. Damiani, G., et al., *The effectiveness of computerized clinical guidelines in the process of care: a systematic review.* BMC Health Services Research, 2010. **10**(1): p. 2.

30. Barahona, P., et al., *Computerising a guideline for the management of diabetes.* International Journal of Medical Informatics, 2001. **64**(2-3): p. 275-284.

31. Quaglini, S., et al., *Smooth Integration of Decision Support into an Existing Electronic Patient Record*, in *Artificial Intelligence in Medicine*, S. Miksch, J. Hunter, and E. Keravnou, Editors. 2005, Springer: Berlin/Heildelberg. p. 89-93.

32. Bates, D.W., et al., *Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality.* Journal of the American Medical Informatics Association, 2003. **10**(6): p. 523-530.

33. Fox, J., et al., *From Guidelines to Careflows: Modelling and Supporting Complex Clinical Processes*, in *Computer-based medical guidelines and protocols: a primer and current trends: A Primer and Current Trends*, A. ten Teije, S. Miksch, and P. Lucas, Editors. 2008, IOS Press: Amsterdam. p. 44-62.

34. Schadow, G., D.C. Russler, and C.J. McDonald, *Conceptual alignment of electronic health record data with guideline and workflow knowledge.* MIE 2003, 2001. **64**(2-3): p. 259-274.

35. Dadam, P., M. Reichert, and K. Kuhn, *Clinical Workflows - The Killer Application for Process-oriented Information Systems?*, in *BIS 2000 - Proc. of the 4th Int'l Conference on Business Information Systems, Poznan Poland, April 2000*, W. Abramowicz and M.E. Orlowska, Editors. 2001, Springer-Verlag. p. 36-59.

36. Bisantz, A.M. and J.J. Ockerman, *Lessons from a focus on artefacts and implicit theories: Case studies in analysis and design*, in *Handbook of Cognitive Task Design*, E. Hollnagel, Editor. 2003, Lawrence Erlbaum Associates: Mahwah, NJ. p. 633-652.

37. Woods, D.D., *Designs are hypotheses about how artifacts shape cognition and collaboration.* Ergonomics, 1998. **41**(2): p. 168-173.

38. Woods, D.D., E.M. Roth, and K.B. Bennett, *Explorations in Joint Human-Machine Cognitive Systems*, in *Cognition, Computing and Cooperation*, S.P. Robertson, W. Zachary, and J.B. Black, Editors. 1990, Ablex Publishing Company: Norwood, NJ. p. 123-158.

39. Hollnagel, E., *Prolegomenon to Cognitive Task Design*, E. Hollnagel, Editor. 2003, Lawrence Erlbaum Associates: Handbook of Cognitive Task Design. p. 3-16.

40. Chipman, S., J. Schraagen, and V. Shalin, *Introduction to cognitive task analysis*, in *Cognitive Task Analysis*, J. Schraagen, S. Chipman, and V. Shalin, Editors. 2000, Lawrence Erlbaum Associates: Mahwah, NJ. p. 3-23.

41. Crystal, A. and B. Ellington. *Task analysis and human-computer interaction: approaches, techniques, and levels of analysis*. in *Tenth Americas Conference on Information Systems*. 2004. New York, NY: Association for Information Systems.

42. Barnard, P. and J. May, *Towards a theory-based form of cognitive task analysis of broad scope and applicability*, in *Cognitive Task Analysis*, J. Schraagen, S. Chipman, and V. Shalin, Editors. 2000, Lawrence Erlbaum: Mahwah, NJ. p. 147-164.

43. Kushniruk, A.W., *Analysis of Complex Decision-Making Processes in Health Care: Cognitive Approaches to Health Informatics.* Journal of Biomedical Informatics, 2001. **34**(5): p. 365-376.

44. Jaspers, M.W., et al., *The use of cognitive methods in analyzing clinicians' task behavior.* Studies in Health Technology & Informatics, 2002. **93**: p. 25-31.

45. Baxter, G.D., et al., *Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit.* Artificial Intelligence in Medicine, 2005. **35**(3): p. 243-257.

46. Beuscart-Zephir, M.C., F. Anceaux, and J.M. Renard, *Integrating users' activity analysis in the design and assessment of medical software applications: the example of anesthesia.* Studies in Health Technology & Informatics, 2000. **77**: p. 234-8.

47. Lenz, R. and M. Reichert, *IT Support for Healthcare Processes*, in *Business Process Management*, W.M.P. van der Aalst, et al., Editors. 2005, Springer-Verlag: Berlin; Heidelberg. p. 354-363.

48. Lenz, R. and M. Reichert, *IT support for healthcare processes – premises, challenges, perspectives.* Data & Knowledge Engineering, 2007. **61**(1): p. 39-58.

49. Ram, F.S.F., *Primary care management of asthma using written management plans and asthma clinics: evidence of effectiveness from recent Cochrane systematic reviews.* Primary Care Respiratory Journal, 2003. **12**(1): p. 21-24.

50. Gibson, P.G., et al., *Self-management education and regular practitioner review for adults with asthma.* Cochrane Database of Systematic Reviews, 2002(3): p. Art. No.: CD001117.

51. Clark, N.M., et al., *Long-term effects of asthma education for physicians on patient satisfaction and use of health services.* European Respiratory Journal, 2000. **16**(1): p. 15-21.

52. Star, S.L. and J.R. Griesemer, *Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39.* Social Studies of Science, 1989. **19**(3): p. 387-420.

53. Glaser, R. and M.T.H. Chi, *Overview*, in *The Nature of Expertise*, M.T.H. Chi, R. Glaser, and M.J. Farr, Editors. 1988, Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ.

54. Johnson, E.J., *Expertise and Decision under Uncertainty: Performance and Process*, in *The Nature of Expertise*, M.T.H. Chi, R. Glaser, and M.J. Farr, Editors. 1988, Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ. p. 209-228.

55. Dubois, D. and V. Shalin, *Describing job expertise using cognitively oriented task analysis (COTA)*, in *Cognitive Task Analysis*, J. Schraagen, S. Chipman, and V. Shalin, Editors. 2000, Lawrence Erlbaum: Mahwah, NJ. p. 41-56.

56. Estabrooks, C.A., et al., *A guide to knowledge translation theory.* Journal of Continuing Education in the HealthProfessions, 2006. **26**(1): p. 25-36.

57. Straus, S.E., J. Tetroe, and I. Graham, *Defining knowledge translation.* Canadian Medical Association Journal, 2009. **181**(3-4): p. 165-168.

58. Choi, B.C.K., *Understanding the basic principles of knowledge translation.* Journal of Epidemiology and Community Health, 2005. **59**(2): p. 93-93.

59. Graham, I.D., et al., *Lost in knowledge translation: Time for a map?* Journal of Continuing Education in the Health Professions, 2006. **26**(1): p. 13-24.

60. Genuis, S.K. and S.J. Genuis, *Exploring the continuum: medical information to effective clinical practice. Paper I: the translation of knowledge into clinical practice.* Journal of Evaluation in Clinical Practice, 2006. **12**(1): p. 49-62.

61. Bhattacharyya, O., S. Reeves, and M. Zwarenstein, *What Is Implementation Research?: Rationale, Concepts, and Practices.* Research on Social Work Practice, 2009. **19**(5): p. 491-502.

62. Ferlie, E., L. Fitzgerald, and M. Wood, *Getting evidence into clinical practice: an organisation behaviour perspective.* Journal of Health Services Research and Policy, 2000. **5**(2): p. 96-102.

63. Ferlie, E., et al., *Evidence-Based Medicine and Organisational Change: An Overview of Some Recent Qualitative Research*, in *Organisational Behaviour and Organisational Stuides in Health Care: Reflections on the Future*, L. Ashburner, Editor. 2001, Palgrave: Basingstoke. p. 18-42.

64. Mittman, B., X. Tonesk, and P. Jacobson, *Implementing clinical practice guidelines: social influence strategies and practitioner behavior change.* Quality Review Bulletin, 1992. **18**(12): p. 413-422.

65. Taylor, V.M., et al., *Modifying community practice styles: The back pain outcome assessment team information dissemination effort.* Journal of Continuing Education in the Health Professions, 1996. **16**(4): p. 203-214.

66. Locock, L., et al., *Understanding the role of opinion leaders in improving clinical effectiveness.* Social Science & Medicine, 2001. **53**(6): p. 745-57.

67. Doumit, G., et al., *Local opinion leaders: effects on professional practice and health care outcomes (Review)*. 2007, Cochrane Database of Systematic Reviews p. DOI: 10.1002/14651858.

68. Gabbay, J. and A. le May, *Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care.* British Medical Journal, 2004. **329**(7473): p. 1013-1017.

69. Lathlean, J. and A. Le May, *Communities of practice: an opportunity for interagency working.* Journal of Clinical Nursing, 2002. **11**(3): p. 394-398.

70. Grimshaw, J.M., et al., *Changing Provider Behavior: An Overview of Systematic Reviews of Interventions.* Medical Care, 2001. **39**(8): p. II2-II45.

71. Grimshaw, J.M., et al., *Effectiveness and efficiency of guideline dissemination and implementation strategies.* International Journal of Technology Assessment in Health Care, 2005. **21**(01): p. 149-149.

72. Dopson, S., et al., *No Magic Targets! Changing Clinical Practice To Become More Evidence Based.* Health Care Management Review, 2002. **27**(3): p. 35.

73. NHLBI Communications Office. *National Asthma Guidelines Updated. New Approaches for Monitoring Asthma Control, Expanded Recommendations for Children.* NIH News 2007 August

29, 2007 [cited 2009 November 17]; Available from: http://public.nhlbi.nih.gov/newsroom/home/GetPressRelease.aspx?id=2442.

74. COGS: The Conference on Guideline Standardization. *COGS Statement*. 2003 August 28 [cited 2009 January]; Available from: http://gem.med.yale.edu/cogs/statement.do.

75. Jackson, R. and G. Feder, *Guidelines for Clinical Guidelines.* BMJ, 1998. **317**(15August1998): p. 427-8.

76. Shaneyfelt, T.M., M.F. Mayo-Smith, and J. Rothwangl, *Are Guidelines Following Guidelines?: The Methodological Quality of Clinical Practice Guidelines in the Peer-Reviewed Medical Literature.* JAMA, 1999. **281**(20): p. 1900-1905.

77. The AGREE Collaboration. *Appraisal of Guidelines for Research and Evaluation*. 2001 September 2001 [cited 2009 January]; Available from: http://www.agreecollaboration.org/pdf/agreeinstrumentfinal.pdf.

78. Nonino, F. and A. Liberati, *Essential requirements for practice guidelines at national and local levels.* Neurological Sciences, 2004. **25**(1): p. 2-7.

79. Rosenfeld, R.M. and R.N. Shiffman, *Clinical practice guideline development manual: A quality-driven approach for translating evidence into action.* Otolaryngology-Head and Neck Surgery, 2009. **140**(6): p. S1-S43.

80. Newton, J., D. Knight, and G. Woolhead, *General practitioners and clinical guidelines: a survey of knowledge, use and beliefs.* British Journal of General Practice, 1996. **46**(410): p. 513-7.

81. Hardern, R.D. and S. Hampshaw, *What do accident and emergency medical staff think of practice guidelines?* European Journal of Emergency Medicine, 1997. **4**(2): p. 68-71.

82. Flores, G., et al., *Pediatricians' Attitudes, Beliefs, and Practices Regarding Clinical Practice Guidelines: A National Survey.* Pediatrics, 2000. **105**(3): p. 496-501.

83. Cabana, M.D., et al., *Reasons for Pediatrician Nonadherence to Asthma Guidelines.* Archives of Pediatrics and Adolescent Medicine, 2001. **155**(9): p. 1057-1062.

84. Annett, J. and K. Duncan, *Task Analysis and Training Design.* Occupational Psychology, 1967. **41**: p. 211-221.

85. Annet, J. and N. Stanton, eds. *Task Analysis*. 2000, Taylor & Francis: London.

86. Shepherd, A., *Hierarchical Task Analysis*. 2001, New York: Taylor & Francis.

87. Klein, G. and L. Militello, *Some guidelines for conducting a cognitive task analysis*, in *Advances in Human Performance and Cognitive Engineering Research*. 2001, Emerald Group Publishing. p. 163 - 199.

88. Patel, V.L., J.F. Arocha, and D.R. Kaufman, *A Primer on Aspects of Cognition for Medical Informatics.* J Am Med Inform Assoc, 2001. **8**(4): p. 324-343.

89. Kirlik, A., *Requirements for psychological models to support design: Toward ecological task analysis.*, in *Global perspectives on the ecology of human-machine systems*, J.M. Flach, et al., Editors. 1995, Lawrence Erlbaum Associates, Inc.: Hillsdale, NJ. p. 68-120.

90. Norman, D., *The Design of Everyday Things*. 2000, New York: MIT Press.

91. Connelly, D., et al., *Knowledge resource preferences of family physicians.* Journal of Family Practice, 1990. **30**(3): p. 353-359.

92. McDonald, C.J. and J.M. Overhage, *Guidelines you can follow and trust. An ideal and an example.* JAMA, 1994. **271**(11): p. 872-3.

93. Patel, V.L. and G. Groen, *Knowledge Based Solution Strategies in Medical Reasoning.* Cognitive Science, 1986. **10**: p. 91-116.

94. Patel, V.L. and G. Groen, *The general and specific nature of medical expertise: A critical look*, in *Towards a General Theory of Expertise: Prospects and Limits*, K.A. Ericsson and J. Smith, Editors. 1991, Cambridge University Press: Cambridge. New York. Melbourne. p. 93-125.

95. Ericsson, K.A. and J. Smith, *Prospects and limits of the empirical study of expertise: An Introduction*, in *Towards a General Theory of Expertise: Prospects and Limits*, K.A. Ericsson and J. Smith, Editors. 1991, Cambridge University Press: Cambridge. New York. Melbourne. p. 1-38.

96. Patel, V., G. Groen, and J. Arocha, *Medical expertise as a function of task difficulty.* Memory and Cognition, 1990. **18**(4): p. 394-406.

97. Naylor, C.D., *Putting evidence into practice.* The American Journal of Medicine, 2002. **113**(2): p. 161-163.

98. Naylor, C.D., *The Complex World of Prescribing Behavior.* JAMA, 2004. **291**(1): p. 104-106.

99. Cabana, M.D., et al., *Why Don't Physicians Follow Clinical Practice Guidelines?: A Framework for Improvement.* JAMA, 1999. **282**(15): p. 1458-1465.

100. Lugtenberg, M., et al., *Why don't physicians adhere to guideline recommendations in practice? An analysis of barriers among Dutch general practitioners*. 2009. p. 54.

101. Grol, R., et al., *Attributes of clinical guidelines that influence use of guidelines in general practice: observational study.* BMJ, 1998. **317**(7162): p. 858-861.

102. Shekelle, P.G. and R.L. Kravitz, *Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making.* Health Services Research, 2000. **34**(7): p. 1429-1448.

103. Michie, S. and M. Johnston, *Changing clinical behaviour by making guidelines specific.* BMJ, 2004. **328**(7435): p. 343-345.

104. Isern, D. and A. Moreno, *Computer-based execution of clinical guidelines: A review.* J Am Med Inform Assoc, 2008. **77**(12): p. 787-808.

105. Margolis, A., et al., *Computerized practice guidelines for heart failure management: the HeartMan system.* Proc Annu Symp Comput Appl Med Care., 1995: p. 228-32.

106. Tierney, W.M., et al., *Computerizing guidelines to improve care and patient outcomes: the example of heart failure.* J Am Med Inform Assoc, 1995. **2**(5): p. 316-322.

107. Sun, Y., et al., *The challenges of automating a real-time clinical practice guideline.* Clinical performance and quality health care., 1999. **7**(1): p. 28-35.

108. Lobach, D., *A model for adapting clinical guidelines for electronic implementation in primary care.* Proc Annu Symp Comput Appl Med Care., 1995: p. 581-585.

109. Seroussia, B., J. Bouauda, and G. Chatellier, *Guideline-based modeling of therapeutic strategies in the special case of chronic diseases.* International Journal of Medical Informatics, 2005. **74**(2-4): p. 89-99.

110. Scott, I.A., et al., *Towards more effective use of decision support in clinical practice: what the guidelines for guidelines don't tell you.* Internal Medicine Journal, 2004. **34**(8): p. 492-500.

111. Gibson, P.G. and H. Powell, *Written action plans for asthma: an evidence-based review of the key components.* Thorax, 2004. **59**(2): p. 94-9.

112. Global Initiative for Asthma, *Pocket Guide for Asthma Management and Prevention. A Pocket Guide for Physicians and Nurses. Based on the Global Strategy for Asthma Management and Prevention*. 2010.

113. British Thoracic Society and Scottish Intercollegiate Guidelines Network (SIGN), *British Guideline on the Management of Asthma.* . 2008 (revised 2009).

114. Barton, C., et al., *Management of asthma in Australian general practice: care is still not in line with clinical practice guidelines.* Primary Care Respiratory Journal, 2009. **18**(2): p. 100-5.

115. Hoskins, G., et al., *Results of a national asthma campaign survey of primary care in Scotland.* International Journal for Quality in Health Care, 2005. **17**(3): p. 209-215.

116. Barton, C.A., et al., *Caregivers' use of, and attitudes towards, written asthma action plans in North-West Melbourne.* Primary Care Respiratory Journal, 2005. **14**(3): p. 143-146.

117. Wiener-Ogilvie, S., et al., *Do practices comply with key recommendations of the British Asthma Guideline? If not, why not?* Primary Care Respiratory Journal, 2007. **16**(6): p. 369-377.

118. Sulaiman, N.D., et al., *Factors associated with ownership and use of written asthma action plans in North-West Melbourne.* Primary Care Respiratory Journal, 2004. **13**(4): p. 211-217.

119. Cabana, M.D., et al., *Barriers pediatricians face when using asthma practice guidelines.* Archives of Pediatrics and Adolescent Medicine, 2000. **54**(7): p. 685 - 93.

120. Zwar, N.A., et al., *General practitioner views on barriers and facilitators to implementation of the Asthma 3+ Visit Plan.* Medical Journal of Australia, 2005. **183**(2): p. 64-7.

121. Glasgow, N.J., et al., *Proactive asthma care in childhood: general practice based randomised controlled trial.* BMJ, 2003. **327**(7416): p. 659.

122. Cabana, M.D., et al., *Reasons for pediatrician nonadherence to asthma guidelines.* Archives of Pediatrics and Adolescent Medicine, 2001. **155**(9): p. 1057 - 62.

123. Box, G.E.P. and G.C. Tiao, *Intervention Analysis with Applications to Economic and Environmental Problems.* Journal of the American Statistical Association, 1975. **70**(349): p. 70-79.

124. Box, G., G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*. 3 ed. 1994, Englewood Cliffs, NJ: Prentice Hall.

125. Liu, L.-M., *Time Series Analysis and Forcasting*. 2005, River Forest, ILL: Scientific Computing Associates Corp.

126. Douglass, J., et al., *A qualitative study of action plans for asthma.* BMJ, 2002. **324**(7344): p. 1003-5.

127. Toelle, B. and F.S.F. Ram, *Written individualised management plans for asthma in children and adults.* Cochrane Database of Systematic Reviews, 2004(1).

128. Eccles, M., et al., *Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial.* BMJ, 2002. **325**(7370): p. 941.

129. Homer, C.J., et al., *Impact of a Quality Improvement Program on Care and Outcomes for Children With Asthma.* Arch Pediatr Adolesc Med, 2005. **159**(5): p. 464-469.

130. Bell, L.M., et al., *Electronic Health Record-Based Decision Support to Improve Asthma Care: A Cluster-Randomized Trial.* Pediatrics, 2010. **125**(4): p. e770-777.

131. Raine, R., C. Sanderson, and N. Black, *Developing clinical guidelines: a challenge to current methods.* British Medical Journal, 2005. **331**(7517): p. 631-633.

132. The AGREE Collaboration, *Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project.* Quality and Safety in Health Care, 2003. **12**(1): p. 18-23.

133. Jadad, A.R., et al., *Systematic reviews and meta-analyses on treatment of asthma: critical evaluation.[Erratum appears in BMJ 2000 Jul 29;321(7256):275].[Erratum appears in BMJ 2000 Apr 8;320(7240):984].* BMJ, 2000. **320**(7234): p. 537-40.

134. Tunis, S.R. and S.D. Pearson, *Coverage Options For Promising Technologies: Medicare's 'Coverage With Evidence Development'.* Health Affairs, 2006. **25**(5): p. 1218-1230.

135. Bach, P.B. *Coverage with evidence development*. in *Roundtable on Evidence-Based Medicine Workshop. The Learning Healthcare System.* 2006. Washington, DC: Institute of Medicine.

136. Worrall, J., *Evidence: philosophy of science meets medicine.* Journal of Evaluation in Clinical Practice, 2010. **16**(2): p. 356-362.

137. Rawlins, M., *De testimonio: on the evidence for decisions about the use of therapeutic interventions.* LANCET, 2008. **372**(9656): p. 2152-2161.

138. Ashcroft, R.E., *Current epistemological problems in evidence based medicine.* Journal of Medical Ethics, 2004. **30**(2): p. 131-135.

139. Bluhm, R., *The epistemology and ethics of chronic disease research: Further lessons from ECMO.* Theoretical Medicine and Bioethics, 2010. **31**(2): p. 107-122.

140. Goodman, K.W., *Ethics, evidence, and public policy.(evidence-based medicine).* Perspectives in Biology and Medicine, 2005. **48**(4): p. 548-554.

141. Slowther, A., S. Ford, and T. Schofield, *Ethics of evidence based medicine in the primary care setting.* Journal of Medical Ethics, 2004. **30**(2): p. 151-155.

142. Grapow, M.T.R., et al., *Randomized controlled trials do not reflect reality: Real-world analyses are critical for treatment guidelines!* The Journal of Thoracic and Cardiovascular Surgery, 2006. **132**(1): p. 5-7.

143. Dreyfuss, D., *Beyond randomized, controlled trials.* Current Opinion in Critical Care, 2004. **10**(6): p. 574-8.

144. Vliet Vlieland, T.P.M., *Managing Chronic Disease: Evidence-Based Medicine or Patient Centred Medicine?* Health Care Analysis, 2002. **10**(3): p. 289-298.

145. Van Spall, H.G.C., et al., *Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals.* JAMA: The Journal of the American Medical Association, 2007. **297**(11): p. 1233-1240.

146. Godwin, M., et al., *Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity.* BMC Medical Research Methodology, 2003. **3**: p. 28.

147. Solberg, L.I., et al., *How Can Primary Care Cross the Quality Chasm?* Annals of Family Medicine, 2009. **7**(2): p. 164-169.

148. Travers, J., et al., *External validity of randomised controlled trials in asthma: to whom do the results of the trials apply?* Thorax, 2007. **62**(3): p. 219-223.

149. Uijen, A.A., et al., *Hypertension patients participating in trials differ in many aspects from patients treated in general practices.* Journal of Clinical Epidemiology, 2007. **60**(4): p. 330-335.

150. Clancy, C. *Hints of a Different Way - Case Studies in Practice-Based Evidence, Opening Remarks*. in *Roundtable on Evidence-Based Medicine Workshop. The Learning Healthcare System.* 2006. Washington, DC: Institute of Medicine.

151. Soumerai, S. *Hints of a Different Way - Case Studies in Practice-Based Evidence. Potential of Quasi-Experimental Trial Designs for Evaluating Helath Policy.* in *Roundtable on Evidence-Based Medicine Workshop. The Learning Healthcare System.* 2006. Washington, DC: Institute of Medicine.

152. Gartlehner, G., et al., *A simple and valid tool distinguished efficacy from effectiveness studies.* Journal of Clinical Epidemiology, 2006. **59**(10): p. 1040-8.

153. Rothwell, P.M., *Commentary: External validity of results of randomized trials: disentangling a complex concept.* International Journal of Epidemiology, 2010. **39**(1): p. 94-96.

154. Dekkers, O.M., et al., *How to assess the external validity of therapeutic trials: a conceptual approach.* International Journal of Epidemiology, 2010. **39**(1): p. 89-94.

155. Tunis, S.R., D.B. Stryer, and C.M. Clancy, *Practical Clinical Trials.* JAMA: The Journal of the American Medical Association, 2003. **290**(12): p. 1624-1632.

156. Glasgow, R.E., et al., *Practical Clinical Trials for Translating Research to Practice: Design and Measurement Recommendations.* Medical Care, 2005. **43**(6): p. 551-557.

157. Yang, W., et al., *Observational studies: going beyond the boundaries of randomized controlled trials.* Diabetes Research and Clinical Practice, 2010. **88**(Supplement 1): p. S3-S9.

158. Shankar, R.D., et al., *Integration of textual guideline documents with formal guideline knowledge bases.* Proc AMIA Symp, 2001: p. 617-621.

159. Wang, D., et al., *Representation of Clinical Practice Guidelines for Computer-Based Implementations* in *Computer-Based Support for Clinical Guidelines and Protocols*, B. Heller, et al., Editors. 2001, IOS Press. p. 285-289.

160. ten Teije, A., S. Miksch, and P. Lucas, eds. *Computer-based medical guidelines and protocols: A Primer and Current Trends*. Studies in Health Technology and Informatics, ed. J. Christiansen, et al. Vol. 139. 2008, IOS Press: Amsterdam.

161. IOM Roundtable on Evidence-Based Medicine, *Summary*, in *The Learning Healthcare System. Workshop Summary*, L.A. Olsen, D. Aisner, and J.M. McGinnis, Editors. 2007, Institute of Medicine, National Academies Press: Washington, DC. p. Page 29.

162. Pronovost, P.J., et al., *How can clinicians measure safety and quality in acute care?* The Lancet, 2004. **363**(9414): p. 1061-1067.

163. Crabtree, B.F., et al., *The individual over time: time series applications in health care research.* Journal of Clinical Epidemiology, 1990. **43**(3): p. 241-60.

164. Imhoff, M. and M. Bauer, *Time series analysis in critical care monitoring.* New Horizons, 1996. **4**(4): p. 519-31.

165. Wargon, M., et al., *A systematic review of models for forecasting the number of emergency department visits.* Emergency Medicine Journal, 2009. **26**(6): p. 395-399.

166. Allard, R., *Use of time-series analysis in infectious disease surveillance.* Bulletin of the World Health Organization, 1998. **76**(4): p. 327-33.

167. Xirasagar, S., H.-C. Lin, and T.-C. Liu, *Seasonality in pediatric asthma admissions: the role of climate and environmental factors.* European Journal of Pediatrics, 2006. **165**(11): p. 747-52.

168. Goldstein, M.K., et al., *Translating research into practice: organizational issues in implementing automated decision support for hypertension in three medical centers.* Journal of the American Medical Informatics Association, 2004. **11**(5): p. 368-376.

169. Rosenbrand, K., J. van Croonenborg, and J. Wittenberg, *Guideline Development*, in *Computer-based medical guidelines and protocols: a primer and current trends: A Primer and Current Trends*, A. ten Teije, S. Miksch, and P. Lucas, Editors. 2008, IOS Press: Amsterdam. p. 3-21.

170. Jenders, R.A. and R.M. Sailors, *Convergence on a Standard for Representing Clinical Guidelines: Work in Health Level Seven*, in *Medinfo 2004 : proceedings of the 11th World Congress on MedicalInformatics*, M. Fieschi, et al., Editors. 2004, IOS Press: Amsterdam; Washington, D.C. p. 130-134.

171. de Clercq, P.A., et al., *Approaches for creating computer-interpretable guidelines that facilitate decision support.* Artificial Intelligence in Medicine, 2004. **31**(1): p. 1-27.

172. Bloomrosen, M. and D.E. Detmer, *Informatics, evidence-based care, and research; implications for national policy: a report of an American Medical Informatics Association health policy conference.* Journal of the American Medical Informatics Association, 2010. **17**(2): p. 115-23.

173. US Department of Health and Human Services, National Institutes of Health, and National Heart Lung and Blood Institute. *Expert Panel Report 3 (EPR3): Guidelines for the Diagnosis and Management of Asthma*. 2007 August 28, 2007 [cited 2009 October 18]; Available from: http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm.

174. Yaffee, R. and M. Monnie, *Introduction to Time Series Analysis and Forecasting with Applications of SAS and SPSS*. 2000, San Diego, CA: Elsevier Academic Press.

175. Dickey, D.A. and W.A. Fuller, *Distribution of the Estimators for Autoregressive Time Series With a Unit Root.* Journal of the American Statistical Association, 1979. **74**(366): p. 427-431.

176. Hosking, J.R.M., *Equivalent Forms of the Multivariate Portmanteau Statistic.* Journal of the Royal Statistical Society. Series B (Methodological), 1981. **43**(2): p. 261-262.

177. Li, W.K. and A.I. McLeod, *Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models.* Journal of the Royal Statistical Society. Series B (Methodological), 1981. **43**(2): p. 231-239.

178. Akaike, H., *A new look at the statistical model identification.* Automatic Control, IEEE Transactions on, 1974. **19**(6): p. 716-723.

179. Schwarz, G., *Estimating the Dimension of a Model.* The Annals of Statistics, 1978. **6**(2): p. 461-464.

180. McDowall, D., et al., *Interrupted Time Series Analysis*. Quantitative Applications in the Social Sciences, ed. J.L. Sullivan. 1980, Beverly Hills; London: Sage Publications.

181. Multiple Risk Factor Intervention Trial Research Group, *Multiple Risk Factor Intervention Trial.* JAMA: The Journal of the American Medical Association, 1982. **248**(12): p. 1465-1477.

182. Davey Smith, G., et al., *Incidence of Type 2 Diabetes in the Randomized Multiple Risk Factor Intervention Trial.* Annals of Internal Medicine, 2005. **142**(5): p. 313-322.

183. Braude, H., *Clinical intuition versus statistics: different modes of tacit knowledge in clinical epidemiology and evidence-based medicine.* Theoretical Medicine and Bioethics, 2009. **30**(3): p. 181-198.

184. Hill, A.B., *I.--THE AIM OF THE STATISTICAL METHOD.* The Lancet, 1937. **229**(5914): p. 41-43.

185. Meldrum, M.L., *A brief history of the randomized controlled trial. From oranges and lemons to the gold standard.* Hematology/Oncology Clinics of North America, 2000. **14**(4): p. 745-760.

186. Armitage, P., *Fisher, Bradford Hill, and randomization.* International Journal of Epidemiology, 2003. **32**(6): p. 925-928.

APPENDIX A. RECOMMENDATIONS TO AHRQ

This appendix contains the recommendations the *HIT Asthma* project team made to AHRQ in a report that documented gaps between clinician requirements for knowledge and workflow support and the asthma guidelines.

The analysis performed in Chapter II enhances empirical findings which were documented in a 2009 report to the Agency for Health Research and Quality (AHRQ), the sponsor of the *HIT Asthma Project.* The project's goal was to improve the quality of asthma care through a computerized decision support tool that would guide a clinician through the typical tasks and decisions required during an office visit for asthma. While obtaining user requirements [28, 32] for the tool, and comparing them to the structure and content of the most recent national guidelines for asthma[173], there were numerous instances in which the guidelines did not meet users' needs. The agency requested a summary of these instances, along with recommendations for improvement, which it eventually shared with the office at the National Heart, Lung and Blood Institute (NHLBI) that had sponsored guideline development. The purpose was not to cite deficiencies in this particular set of guidelines, but to call NHLBI attention to the gaps between the support that working physicians required, and the narratives through which guidelines most frequently are expressed.

The report to AHRQ concluded with a set of recommendations directed towards AHRQ, which in turn shared with the NHLBI. Recommendations were primarily editorial. A differentiation was made between the narrative, called the "foundational reference text", and all other artifacts that may be used to convey the developers' recommendations. The foundational reference text is the artifact that most closely resembles the artifact that currently conveys guideline recommendations. The recommendations called for three

additional artifacts: 1) Summary documents (equivalent to pocket guides and/or single web pages), 2) A coder's manual, and 3) An educators' guide.

This appendix summarizes editorial recommendations about the foundational reference text, and presents recommendations for the coder's manual.

THE FOUNDATIONAL REFERENCE TEXT

The foundational reference text is typically the only document that developers currently prepare. It is a lengthy natural language document written in the format of a comprehensive review article intended for publication in peer-reviewed clinical journals. Like the EPR-3, these documents summarize research findings, elucidate the reasoning used to draw conclusions from the findings, and communicate collective judgments about the best actions to take under certain circumstances. Although the intended readership is presumably practicing clinicians, the actual readership may more likely be payers, policy makers and quality evaluators who look towards the text to determine how health care quality should be assessed.

Editorial clarity in the reference text

Guideline developers should more clearly communicate instances in which they believe that individual clinical judgment is most appropriate. For example, the EPR-3 contains the following statement on page 249, "Consider stepping up one or two steps, or starting a course of oral corticosteroids, or both". Developers should communicate the same sentiment using a statement such as: "Research and experience have shown that all of the following actions can be beneficial, but the evidence is not strong enough to support a declarative recommendation: …" and then list the beneficial actions the clinician could take.

Guideline developers should speak more directly to policy and payer audiences by referring not just to the strength of the evidence on which recommendations are based, but by clearly using the terminology that the Dartmouth Atlas Project (DAP) uses in attempting to explain geographic variation in per capita health care costs. The DAP differentiates between three types of medical care:

- Necessary care (care for which there is consensus about effectiveness and necessity)
- Preference-sensitive care (care that requires a consideration of tradeoffs between risks and benefits)

- Uncertain care (the DAP actually calls this “supply-sensitive care” because they have found that the most robust predictor of how much is provided is the geographically available supply of it. It is care for which there is no clinical consensus on effectiveness, and may not be addressed as part of medical education).

By clearly articulating recommendations for care considered necessary, preference-sensitive, and uncertain, guideline developers will help ensure that those who use the guidelines to develop quality measures will more likely reflect what the guideline developers know and believe.

Error-checking the reference text

As recommended by the medical informatics community, guideline developers should check the reference text with an informatics language developed to automatically convert narration into executable code. Called a “markup” language, running a narrative through it identifies instances of internal inconsistency, logical incompleteness, and lack of clarity.

Some in the medical informatics community argue that all instances of internal inconsistency, logical incompleteness, and/or lack of clarity should be resolved. We argue that resolution is appropriate if the identified inconsistency or ambiguity arose through editorial mishap: developers may have intended to issue a definitive, clearly-worded prescriptive statement, but wrote passages that didn’t work as intended. Such instances should be corrected because the developers’ original intent is preserved.

However, guidelines contain numerous instances of intentional ambiguity, in which authors use natural language to communicate not just information, but social messages about that information. There may have been insufficient evidence to support a clear, prescriptive statement, or there may have been disagreement among the guideline developers about the best ways to interpret the evidence that did exist, or the guideline developers may have disagreed about the value of recommended actions, relative to the anticipated resources such action might require. Intentional ambiguity should be communicated explicitly, by stating that guideline developers believe that multiple actions are reasonable, and that they are not endorsing any particular one. Such language will clearly signal the policy and payer audience that these are <u>not</u> evidence-based best practices to be used to evaluate that quality of medical

care, and it will clearly signal software developers that multiple options should be made available from which the user may choose.

Guideline developers should articulate responses to all logically possible clinical circumstances, even if there is no evidence on which to base a preference for one recommendation over another. Theory and evidence suggests that if clinicians take the trouble to consult guidelines only to find that their questions are not even addressed, they will be much less likely to take such trouble in the future. While it is pragmatically infeasible to list such responses in the primary reference text, it is feasible and necessary to identify such responses in the coder's manual, discussed below.

THE CODER'S MANUAL AS AN ARTIFACT

Guideline developers should prepare a "coder's manual", intended for use by software developers who are creating decision support software that is executable at the point of care. A useful manual will have the following properties:

- Sections will be differentiated according to users' workflows. When considering a single clinical condition, such as asthma, the workflow is dictated in part by the setting where the encounter with the patient takes place (inpatient hospital, emergency room, outpatient clinic, school), in part by the patient's previous encounter history (first encounter, or follow-up encounter), and in part by the patient's level of asthma control. Workflows arising in all circumstances will be clearly described, in terms of the sequence of actions clinicians must take, and the knowledge they require to perform each action appropriately.
- There will be a glossary containing clear operational definitions for all terms.
- There will be a section that covers circumstances in which data are missing,
- There will be two indices, one indexing subjects within the manual, and the other indexing the same subjects in the reference text.
- Every statement in the coder's manual must be consistent with other statements in the manual; statements must cover all logically possible circumstances; statements can be interpreted only one way.

APPENDIX B. METHODS TO ASSESS THE EFFECT OF THE CDS TOOL

This appendix contains detailed descriptions of the methods used to assess the effect of introducing the computerized decision support tool for asthma at Hennepin County Medical Center (HCMC).

APPENDIX B.1. METHODS TO IDENTIFY ASTHMA PATIENTS

"Patients with asthma" were identified using the following criteria applied to records from the electronic health record (EHR) system at HCMC:

- The patient had at least one visit in which the encounter level diagnostic code (ICD-9) in the first or second position began with 493, OR
- The patient had at least two visits in which at least one encounter-level diagnostic code in any position began with 493, OR
- The patient had at least one visit in which there was at least one encounter-level diagnostic code in any position beginning with 493 (any position), and in addition, asthma was listed as active on the problem list.

A "visit" was defined as a patient-date-site combination; all encounters taking place with that patient on that date at that site were collapsed into the same visit. Multiple encounters per visit could arise if, for example, a patient was scheduled for an office encounter and an encounter with a diagnostic device on the same day at the same site. It was possible for the patient to have more than one visit on a single day, if the encounters on that day took place at multiple sites. Because analyses were performed for each site individually and for all sites overall, patients with visits at different sites were included within each site's analysis.

The date the patient was "known to have asthma" was the earliest date in which any one of the three criteria above were met. Using these criteria, 9287 patients were "known to have asthma".

APPENDIX B.2. RATIONALE FOR THE SAMPLING STRATEGY

Resources were available to perform up to 1000 reviews necessary to identify patients with paper-generated plans scanned into their electronic records. From the 9287 patients known to have asthma, the sample was selected so that results would yield the most useful information for quality improvement purposes, as well as the information most likely to generalize for research purposes.

A quality goal at HCMC was for all asthma patients to have current AAPs generated electronically using the CDS tool[*]. If asthma patients were not receiving AAPs generated by the tool, it was necessary to know why in order to design the most effective quality improvement strategy. Two primary hypotheses were entertained: 1) Providers were not generating AAPs at all, and 2) Providers were generating paper plans but were resisting the use of the CDS tool. If providers were not generating any plans at all, the strategy would focus on the need for them. If providers were generating plans using paper, but not using the CDS tool, the strategy would focus on better dissemination of the tool.

To determine the reasons that plans were not being generated by the tool, it was necessary to review records from patients with variable baseline rates for paper AAPs, and for whom there was no a-priori reason to believe that their providers would universally accept or reject the tool. This called for sampling patient records from multiple clinics, because it was known that clinic cultures differed in ways that were likely to affect the use of the tool. Sampling patient records from multiple clinics also would help address the research question of whether introduction of the tool achieved comparable or different results across different

[*] Use of the electronic tool was a goal because the tool provided decision support that brought evidence-based guidelines to the point of care, and also because it was easier to identify patients with completed plans electronically than through chart review.

types of practice sites. While it was possible to take a random sample of all asthma patients regardless of where they received care, this would have reduced the ability to detect differences due to the effect within clinics, because of the necessary reduction in within-clinic sample size. The alternative was to select targeted samples of asthma patient records from a subset of clinics available.

The subset was chosen using the following reasoning:

- Providers who practiced at the two adult general medicine clinics on the main campus had expressed a strong resistance to using the CDS tool, arguing that because their patients had so many complex chronic conditions, they could not afford to devote a single 15-minute visit only to asthma. Previously performed chart reviews confirmed this sentiment: Less than 1% of the adult asthma patients who received care at these clinics had paper AAPs. Because it was anticipated that uptake of the tool within these clinics was likely to be low, these clinics were rejected as observation sites, because the information value gleaned from the review effort would also likely be low.

- Among pediatric patients, the age group with the highest rates of electronic AAPs was 5-11 years. This age segment was the target of reviews so that results would not be distorted by age effects. Records for children in this age group were selected if they received care at: 1) one of the sub-clinics within the largest family medicine clinic in the community (FM1), 2) the smallest family medicine clinic in the community (FM2), 3) the smaller of the two hospital-based pediatrics clinic (Peds1), and 4) one of the sub-clinics within the larger hospital-based pediatrics clinic (Peds2).

After choosing pediatric patients, resources remained to perform additional chart reviews. These resources were used to review the charts of patients ages 21 and older who received care at the same neighborhood-based family medicine clinics who cared for the children whose records were reviewed. The purpose was to determine if, within the same clinic, effects differed between pediatric and adult patients.

APPENDIX B.3. METHODS TO ESTIMATE PARAMETERS IN ARIMA MODELS

The Box-Tiao strategy for assessing the effect of an intervention given a series of observations taken over time begins with the general representation of each observation *y* at time *t* as:

$$y_t = C + \frac{\omega_{ls}}{\delta_{la}} I_{t\text{-}b} + N_t$$

In this general expression,

- *C* is the overall average of the series;
- $I_{t\text{-}b}$ is a dummy indicator with a value of 1 if the intervention was in effect at time *t* and 0 otherwise; *b* represents the delay until the intervention manifests its effect;
- $\frac{\omega_{ls}}{\delta_{la}}$ is the general expression for the effect of the intervention *I* at time *t-b*, where
  - ω_{ls} represents the initial effect of the intervention, and the subscript refers to the number of lags required before the effect manifests. Absolute values of ω_{ls} closest to 1 indicate sharply accelerating effects, where positive values represent increases and negative values indicate decreases.
  - δ_{la} represents the attenuation of the initial effect, or how quickly the slope levels off. The subscript represents the number of lags over which the attenuating effect is in place [174]. Interpretation of positive values of δ_{la} that approach 1 is no attenuation of the effect over time. Interpretation of positive values of δ_{la} that approach 0 is rapid attenuation of the effect over time. Interpretation of negative values of δ_{la} is troublesome, because the negative value represents effects that "wobble" over time. As values approach -1, this wobbling is interpreted to persist; as negative values approach 0, this wobbling is interpreted to settle down rapidly.
- N_t represents the "noise component" affecting the value of y_t; it is estimated by fitting an autoregressive integrated moving average (ARIMA) model, expressed as (p,d,q), where:
  - *p* represents the autoregressive (AR) component of the model, interpreted as the way in which observations at one moment in time affect observations *p* time periods or lags later;

- *d* represents the number of times successive observations must be subtracted from each other to remove drifts and/or trends and thus achieve stationarity;

- *q* represents the moving average (MA) component of the model, interpreted as the way in which shocks to the system at one moment in time affect observations *q* time periods or lags later;

The best-fitting ARIMA model generates residuals that most closely approximate white noise. Recommended methods[125, 174] call for first identifying the best-fitting noise model for the series, and then adding terms to the model to estimate the effect, if any, of interruptions to the series. In this study, there were six series of over-time observations, given the two patient age groups (pediatric and adults) and the four clinics where the patients had visits (two family medicine clinics in the community that were visited by both pediatric and adult patients, and two pediatric clinics on the main hospital campus that were visited only by pediatric patients).

Interrupted ARIMA models were fit separately to each series. According to Dickey-Fuller tests [175], each age-clinic series achieved stationarity after differencing at lags 1 (one week) and additionally at 52 (one year). The lag at 52 was unsurprising, given the known seasonality of asthma symptoms [167] likely to appear in the series as seasonality in patient visit patterns and provision of AAPs. Another source of annual seasonality in children was the onset of the school year, when children were both more susceptible to viral respiratory infections transmitted by other students (known to exacerbate asthma symptoms), and when school nurses requested AAPs from medical sites.

After differencing, the model for each series was fit with the necessary AR and/or MA terms, using the modified Portmanteau test [176, 177] to compare residuals to white noise.

Selection among competing models was achieved by using the smallest values of the Akaike Information Criterion (AIC) [178], and the Schwarz Bayesian Criterion (SBC)[179], both of which provide estimates of model fit, with a penalty for the number of parameters estimated.

Once the most parsimonious ARIMA model generating the "whitest" noise was found, additional terms representing "interruptions" to the series were fit. For the series of observations from patients (both pediatric and adult) visiting the two family medicine clinics in the community, there was one "interruption", which was the intervention that included education about the importance of asthma action plans and introduction to the computerized tool for generating them.

For the series of observations from patients visiting the two pediatrics clinics on the main hospital campus, there was an additional "interruption", representing the effect of the pilot performed at those two clinics for three weeks prior to introducing the tool across the campus overall. It was necessary to explicitly model the potential effect of the pilot because a flaw was discovered in the tool's underlying database while the pilot was underway. Correction of the flaw required a complete overhaul of the database, which meant that all data about patients receiving electronically generated plans during the pilot were lost. The lost data had the potential effect of continuously undercounting the overall rates of current asthma action plans for patients in these clinics, because if, for patients seen during the pilot, providers abandoned paper plans in favor of electronic ones, the results would only show the abandonment of paper AAPs; they would not show the creation of electronic ones.

To model the effect of the intervention, a common tool-introduction date of July 10, 2009 was set for all six age-clinic series. It was assumed that the effect was delayed until appearances of electronically generated plans within the series occurred.

It was assumed that the onset of the effect of the intervention – if any – would be gradual. Representation of this assumption called for setting the value of $I_{t\text{-}b}$ to be 1 from July 10, 2009 onwards, and to 0 prior to July 10, 2009.

No a-priori assumptions were made about the attenuation of the effect. To model potential attenuation, different δ parameters were introduced at lags that matched the graphs of the series over time, and also matched lag-specific spikes in the autocorrelation functions of the residuals[180].

Parameters representing initial (ω_{ls})and attenuating (δ_{la}) effects were retained in the model if they achieved conventional levels of statistical significance ($p<0.05$). In some instances, terms were retained at less stringent levels of statistical significant if they were required to eliminate autocorrelation among the residuals, or if their inclusion substantially reduced both the AIC and the SBC.

In cases of competing models, the one selected was the one that was easiest to interpret with reference to the series' plot over time.

APPENDIX C. INSTRUCTIVE ANECDOTES FROM RCTS

This appendix contains historical information about uses of randomized controlled trials (RCTs).

APPENDIX C.1. EPISTEMOLOGICAL CHALLENGES IN THE MULTIPLE RISK FACTOR INTERVENTION TRIAL (MRFIT)

The Multiple Risk Factor Intervention Trial enrolled 12,866 participants across 22 sites, randomly allocated them to a Special Intervention (SI) or Usual Care (UC) and followed them for 6-7 years. Due to the time it took to enroll all participants, the trial lasted eight years. The primary question addressed by the trial was whether the special intervention that included smoking cessation counseling, and reducing blood pressure (BP) and cholesterol would reduce the risk of death from coronary heart disease (CHD) among middle aged men, compared to usual care. Eligibility was based on a composite risk score that included smoking status, BP and cholesterol; those enrolled had high scores, but had never experienced a coronary event.

During the 8 years of recruitment, treatment and followup, usual care in the community came to resemble the special intervention. At trial conclusion, there was no statistically significant SI-UC difference in the 3-year post-trial CHD mortality rates, largely attributed secular changes in usual care[181]. Differences also were difficult to detect because overall, mortality rates were lower than expected, attributed to the clinical peculiarity of trial participants, who were enrolled because they were at high risk for a coronary event, but had never experienced one. Twenty years after active intervention and followup ceased, but post-trial mortality rates were still being compared, mortality rates among participants still were lower than expected, prompting trial investigators to joke that the participants were immortal.

Clinical peculiarities also arose because of the composite nature of the risk score: Among

trial participants, those who smoked had lower average BP than those who did not, primarily because if they smoked, they could be enrolled with a low BP. These clinical peculiarities frustrated efforts to use the large database to address other questions. For example, the MRFIT database presumably contained information that would help determine if the special intervention reduced the incidence of new-onset Type II diabetes. Diabetics were excluded from the trial, but the risk factors for CHD were similar to those for diabetes. Analysis showed that the relative risk of developing diabetes over six years was lower in SI compared to UC, but only among the non-smokers. Among the smokers, the effect was reversed; SI participants were much more likely to become diabetic than their UC counterparts [182]. Publication of this finding was delayed by 3-4 years because investigators could not explain it. An untestable explanation was that the finding emerged from the peculiar clinical characteristics of the selected participants.

APPENDIX C.2. HISTORY OF THE RCT IN MEDICINE

The randomized controlled trial as a research design was perfected in the 1920s by RA Fisher, who used the method to address questions arising in large scale agriculture. Fisher needed to know which combination of inputs (seed, fertilizer, water) would, on average, produce the greatest yield, regardless of the soils on which they were applied[183]. In 1937, Dr. Bradford Hill embraced and lauded the design, and argued for its use in medicine[184]. Some of the pressing medical questions at that time concerned the most promising preventative or curative treatments for vector-borne diseases afflicting public health such as smallpox, polio, tuberculosis, syphilis. If a large-scale public investment was to be made, it was imperative to know which investment would yield the greatest overall benefit. In other words, the questions were similar to the ones that Fisher confronted: Which treatment would

generate the highest average yield overall?

The RCT is well-suited to questions of this type. Randomization protects physicians from their own biases (and also provides an ethically acceptable way to allocate limited supplies of treatments believed to have benefit). Control reduces within-treatment variability, thus reducing the size of the sample necessary to detect a statistically significant effect.

Both features also were useful for protecting the public from toxic or useless drugs that were promoted by their manufacturers [185], and in the late 1960s, the RCT acquired legal status in the US for precisely this purpose. The Food and Drug Administration (FDA) had ordered removal from the market drugs that medical experts had deemed ineffective. The drugs' manufacturers challenged the orders in court on the grounds that expert opinion was biased. The FDA responded by naming the randomized, double-blinded, controlled clinical trial as the criteria by which "evidence" of efficacy would be acceptable. Although manufacturers challenged this as well, the decision was upheld by the Sixth District Court of Appeals[185], thus endowing this particular study design with a legal and regulatory status that no other study design can claim.

Dr. Hill, the original proponent of employing the RCT in medicine, never asserted that the design would yield information about the best way to treat individual patients, and stated that an RCT demonstrates what "…can be accomplished with a medicine under careful observation and certain restricted conditions," adding that "… the same results will not invariably or necessarily be observed when the medicine passes into general use" [186].