**Information Technology and Medical Errors:**
**Evidence from a Randomized Trial**

by

Jonathan C. Javitt
Potomac Institute for Policy Studies, Arlington, VA
and President's Information Technology Advisory Committee,
White House Office of Science and Technology Policy, Washington DC

James B. Rebitzer
Carlton Professor of Economics
Case Western Reserve University,
National Bureau of Economic Research and
The Levy Institute

Lonny Reisman
Chief Executive Officer
Active Health Management, Inc.
New York, NY

2/16/2005

Send Correspondence to
James Rebitzer, Carlton Professor and
Chair, Economics Dept.
Room 274 PBL
Case Western Reserve University
11119 Bellflower Rd.
Cleveland, OH 44106
James.rebitzer@case.edu
216 368 5537

Abstract

This paper analyses the effect of a decision support tool designed to help physicians detect and correct medical "errors". The data comes from a randomized trial of the technology in a population of commercial HMO patients. The key findings of the study are easily summarized. The new information technology enhances the efficiency of medical services by reducing costs while also improving care quality. Average charges were 6% *lower* in the study group than in the control group. These savings were the result of reduced in-patient charges (and associated professional charges) for the most costly patients. The rate at which potential errors were resolved was generally higher in the study group than in the control group—suggesting an improvement in care quality.

Beyond medical costs and care quality, this study also adds to the growing body of research documenting the ways that information technology can improve the economic efficiency of decisions made by highly trained professionals operating in complex environments.

1.  Introduction

In 1987, Nobel Laureate Robert Solow famously remarked, "you can see the computer

age everywhere but in the productivity statistics." ( (Solow, 1987), p. 36).  Solow's aphorism

neatly summarized the state of knowledge in the late 1980's and early 1990s.  Since that time,

however, economists have been able to identify measurable economic effects of the revolution in

information technology (IT).   The emerging consensus from this research is that the effect of IT

varies depending on the design of organizations and the nature of production processes.  IT

*complements* the work of people engaged in non-routine problem solving and communication

while it *substitutes* for lower-skill tasks involving the sorts of explicit rules that are relatively

easy to program into computers.[1]

Studying the effect of IT on work processes involving non-routine problem solving and

communication is hard -- in large part because the inherent complexity of these processes make

it difficult to identify meaningful performance measures that are also directly related to specific

IT innovations.  The search for good performance indicators and cleanly demarcated innovations

has moved economists away from the analysis of aggregate productivity and technology data

towards more narrowly focused studies. [2]  The added institutional knowledge made possible by

the limited scope of these studies also helps analysts address the selection problems created by

the non-random distribution of new innovations across organizations and work places.[3]

---

[1] For discussions of this perspective see (Autor et al., 2003); (Brynjolfsson and Hitt, 2000); and (Levy and Murnane, 2004).

[2] See for example (Athey and Stern, 2002) on IT and the delivery of emergency medical services; (Autor et al., 2002)on banking;  (Bartel et al., 2004) on computer controlled machines in manufacturing and (Hubbard, 2003) on capacity utilization in the trucking industry.

[3] The econometric challenges involved in studying the effect of IT innovations closely parallel the issues involved in the study of innovations in human resource practices. For an illuminating discussion and review see (Ichniowski and Shaw, 2003) and for an application to the health care

In this paper, we also analyze the effects of an IT enabled innovation in a narrowly defined production process characterized by non-routine problem solving and communication. The information technology we study is a decision support tool designed to notify physicians about potential medical "errors" as well as deviations from evidence based clinical practice guidelines. Our approach is closest in spirit to (Athey and Stern, 2002)'s study of emergency medical services. Like Athey and Stern, we focus on the introduction of a discrete innovation that altered the handling of information in a health care setting and we assess the efficacy of the innovation by tracking health related outcomes. Our econometric approach, however, differs from theirs in that we use a randomized control trial to identify the effect of the new technology.[4]

Although we focus on specific production process, the results we report have implications for broad management and economic issues in health care. A large and influential body of research suggests that preventable medical errors have a substantial effect on the cost and quality of medical care.[5] In response to these findings, a number of high-profile public and

---

setting see (Gaynor et al., 2004).

[4]    (Athey and Stern, 2002) identify the effect of the technology in their study by comparing early and late adopters in a differences-in-differences framework. An obvious issue with this approach is that participants who choose to adopt early might be those for whom the benefits of the innovation are especially large. A randomized controlled trial eliminates this source of bias because the participants receiving the treatment are a random sample of the subject pool. Randomized trials have other limitations, however. The subject pools are often small and may not be representative of the underlying population. This can bias estimates of the effect of the intervention on a population. For a practical example of this sort of bias in a health care setting see Duggan (2003).

[5]    Evidence on the incidence of medical errors was recently reviewed by the Institute of Medicine of the National Academy of Sciences ( (Institute of Medicine Committee on Quality of Health Care in America, 2000)). This report concluded that at least 44,000 Americans die each year as a result of medical error during hospitalization and some estimates suggest the number may be as high as 98,000. To date, very little is known about the incidence of errors in outpatient settings, but the incidence may be high ( (Lapetina and Armstrong, 2002)). A recent study of errors in intensive care units is (Landrigan et al., 2004). A discussion of the literature on the use of IT to reduce errors and increase compliance with evidence-based guidelines can be found in (Institute of Medicine Committee on Quality of Health Care in American, 2001).

private initiatives have called for major new investments in information technology and decision support tools to reduce the incidence of errors and increase compliance with evidence-based treatment guidelines ( (President's Information Technology Advisory Committee, 2004), and (Institute of Medicine Committee on Quality of Health Care in American, 2001)).[6] Economists who have examined these issues generally agree that new information technologies and decision support tools – perhaps combined with novel incentive arrangements -will likely have a substantial influence on both errors and efficiency in the delivery of health care, yet economic studies concerning the efficacy these interventions have been scarce ( (Newhouse, 2002); and ).[7]

The data in this study comes from a randomized trial of a physician decision support technology introduced to a population of commercial HMO patients. As we discuss below, the design of the study allows for especially clean performance measures and strong causal inferences. The key findings are easily summarized. The new information technology enhances the economic efficiency of health care services by reducing costs while also improving care quality. Average charges were 6% *lower* in the study group than in the control group. These

---

[6] As an indicator of the high level of public policy interest in these issues it is worth noting that reference to the use of IT to reduce errors appeared in President Bush's *Economic Report of the President* in 2004 ( (The Council of Economic Advisors, 2004)) and in his State of the Union Address. "By computerizing health records, we can avoid dangerous medical mistakes, reduce costs, and improve care." (President's Information Technology Advisory Committee, 2004, 3). Private sector initiatives concerned with preventing medical errors have also been formed. The most prominent of these may be the Leapfrog Group, a coalition of more than 150 large public and private organizations that provide health care benefits. The purpose of this organization is to use employer purchasing power to speed the adoption of processes that improve patient safety ( (The Leap Frog Group for Patient Safety, 2004)).

[7] There is a growing literature on disease management programs that often rely on information technology similar to that which we evaluate in this study. Disease management programs typically analyze billing records and other clinical information to identify patients whose care deviates from accepted clinical practice guidelines. Although disease management programs have become a ubiquitous part of health care and there is some evidence that they can be effective in reducing costs and improving quality ( (Shojania and Grimshaw, 2005/1/1), (Gertler and Simcoe, 2004)), many of the studies are poorly designed and few of them use evidence from randomized controlled trials of interventions (Shojania and Grimshaw, 2005/1/1).

savings were the result of reduced in-patient charges (and associated professional charges) for the most costly patients. The rate at which potential errors were resolved was generally higher in the study group than in the control group—suggesting an improvement in care quality. Beyond the specific implications for the delivery of medical care, this study documents how IT enabled decision support tools can improve the economic efficiency of decisions made by highly trained professionals solving non-routine problems in complex environments.

The plan of the paper is as follows. Section two describes the setting of the trial and the decision support technology. Section three presents the data analysis. Section four concludes and discusses new research questions raised by the study.

## 2. The Setting

Physician Mistakes:

Physicians make mistakes – and these mistakes are increasingly believed to have a substantial effect on the cost and quality of medical care. The causes of errors are not entirely clear, but a leading suspect is the volume and complexity of the information that physicians must process about their patients' medical conditions and the rapidly changing state of medical knowledge ( ; ; ; ).

If errors come from keeping abreast of "too much information", then it makes sense to look to information technology to help manage this burden. Ideally one would like to use IT to collect and analyze patient information and to communicate likely missteps to physicians. In order for these messages to be influential it is important that they be delivered in a timely fashion, be targeted to specific patients, and that they reliably inform physicians of overlooked issues or issues about which he or she lacked adequate knowledge. Generating these timely, targeted and informative messages is hard -- especially because most physician practices are not

4

linked by a common IT system.  In the absence of such linkages, it is difficult for most HMOs

and managed care organizations to construct usable electronic medical records for patients

treated within their physician networks. [8] Since HMOs and managed care organizations are the

predominant form of private sector health insurance, the problems posed by balkanized IT

systems can be a significant barrier to bringing computer assisted decision making to medical

care.

The decision support software evaluated in this trial was designed to overcome the

problems posed by fragmented IT systems.[9]  It collects information about patients from billing

records, lab feeds and pharmacies to assemble a virtual electronic medical record.  It then passes

this information through a set of decision rules culled from the medical literature.  When the

software uncovers an issue it produces a message, called a care consideration, that includes the

patient's name, the issue discovered, a suggested course of corrective action and a cite to the

relevant medical literature.  The care considerations (CCs) fall into three distinct, but not entirely

mutually exclusive, categories: stop a drug; do a test; and add a drug.  The CCs are also coded

into three severity levels.  Level one messages (the most severe) include potentially life-

threatening situations: for example that a patient's blood potassium levels are dangerously off.

---

[8]    "…to be effective, CDSS [clinical decision support systems] diagnostic systems require detailed, patient-specific clinical information (history, physical results, medications, laboratory test results), which in most health care settings resides in a variety of paper and automated datasets that cannot easily be integrated. Past efforts to develop automated medical record systems have not been very successful because of the lack of common standards for coding data, the absence of a data network connecting the many health care organizations and clinicians involved in patient care, and a number of other factors." (Institute of Medicine Committee on Quality of Health Care in American, 2001, 154)

[9]    The technology we analyze is the property of Active Health Management, Inc.  It is important to note that two of the authors have a proprietary interest in the company.  Dr. Reisman is the CEO of Active Health and is also a shareholder in the company.  Dr. Javitt is a shareholder and has a consulting relationship with the company.  Rebitzer has no proprietary interest in Active Health and no other financial relationship with it.

Level two (moderate) CCs refer to issues that might have an important effect on clinical outcomes. An example might be that the patient is a good candidate for an ACE inhibitor, a drug that is useful in treating many cardiovascular conditions. Level three (least severe) CCs refer to preventative care issues such as being sure that diabetics have regular eye exams.

Contrary to most other studies of medical errors, the potential errors detected by the software are not limited to events occurring during hospitalizations. This is important because, most of the evidence regarding medical errors comes from studying the treatment of hospitalized patients, but in-patient treatments are a declining share of medical treatment.

Study Design [10]

The participants selected for the study came from an HMO located in a midwestern city. All were all under age 65 and had had some medical charges in the year prior to the experiment. Once selected, the participants were randomly allocated into study and control groups. The software was turned on for patients in the study group. This means that their physicians received CC's during the yearlong course of the experiment. The software was *not* turned on for patients in the control group until the experiment was over, but the billing, pharmacy and lab data of these patients was collected and saved. At the end of the year, the control group's medical data was analyzed to find CC's that *would* have appeared if the software had been running. An important feature of the study design was that randomization occurred at the level of the patient. This means that some physicians had patients in both the study and the control group. Thus lessons that physicians learned from receiving a CC for a study group patient might spillover to

---

[10] More information about the design as well as a preliminary analysis of results can be found in (Javitt et al., 2005).

their control group patients.  These spillovers could therefore have the effect of biasing the estimated impact of the decision support software downwards.

Error Messages:

When the software generated a CC, doctors employed by the software company manually checked it.  If the CC passed this scrutiny and was coded a level one (most severe) the HMO's medical director was called and he, in turn, called the appropriate physician.  If the CC was at level two (moderate) or level three (least severe), a nurse employed by the HMO received the error message and decided whether to fax the CC on to the enrollee's physician.  The HMO's nurse passed on most (but not all) level two CC's and some of the level three CC's.  Unfortunately there are no records documenting the nurse's decisions concerning which messages to pass on.  From informal discussions, however, it appears that some types of CCs were not sent because they duplicated advice in disease management programs already in place at the HMO.  These were mostly level three CC's focused on preventative care.  In addition, the nurse decided that some other CCs didn't make clinical sense.

Outcomes:

By comparing the study and the control groups, we can identify the effect of the error detecting software on "remediable" medical errors.  Remediable errors occur when physicians are not fully aware of all the available information concerning their patients and /or the state of relevant medical knowledge.  These errors are important because they might not have occurred if the physician had had timely access to the appropriate information.

The data collected in this study yields two natural performance measures: the rate at which problems identified by CC's are resolved; and the average costs of medical care.  We discuss each of these in turn.

*Resolution Rates:* Physicians often have better information about their patients than does the error detecting software. Actions that look like a misstep to the computer may in fact be the result of informed physician choice and/or patient non-compliance. For this reason, the HMO and the software company viewed CCs as recommendations that physicians were free to ignore if they disagreed.[11] It is thus reasonable to assume that when physicians took steps to resolve the issues identified in a CC, they were doing so in the belief that these actions will be beneficial (or at least not detrimental) for the patient. Rates of resolution of CCs are thus a reasonable indicator of whether the decision support software is improving care quality. Of course, some issues identified by the software would have been resolved even if no messages had been sent to physicians. If we observe higher rates of resolution in the study group than the control group, however, it is reasonable to infer that at least some of the recommendations issued had a positive effect on care quality. Differences in resolution rates between study and control groups can thus be interpreted as an ordinal measure of care quality. A positive differential tells us that quality has improved in the study group, but does not tell us by how much quality has increased.

*Average Costs:* The CC's issued by the system recommended roughly three types of actions: "add a drug", "stop a drug" and "do a test". The first and the third of these entail a direct increase in the utilization of medical resources. If, however, these actions prevent subsequent costly complications, the net effect might be to reduce charges relative to the control group.

---

[11]    Informal conversations with physicians who attended a discussion group about the software indicated that they also viewed the care considerations as suggestions that they could disregard.

Administrative data from the HMO was collected on average charges per member per month in the year prior to the study and also during the year of the study.[12] By comparing average costs in the study and control groups over the year, we can estimate the costs of "remediable" medical errors. These cost estimates cover only a fraction of the total costs of errors. They do not include: the costs of lost work time and productivity; disability costs or the costs of personal care not paid for by insurers; or the costs of litigation resulting from errors. Clearly the charges tracked in this study understate the full social cost of medical mistakes.

It is also worth noting that the study is designed to assess only the short-run cost effects of the intervention. Many of the benefits of avoiding missteps likely appear years after the error occurred. Given the reportedly high rate of turnover among HMO patients, however, much of the financial incentive to act to prevent mistakes may be captured by the short-run savings identified in this study.

### 3. Data and Results

Descriptive Statistics

Table 1 presents descriptive statistics. The analysis does not include enrollees younger than age 11 because the decision support software had very few pediatric treatment guidelines in place at the time of the study. In Panel A, there were 19693 individuals in the study group and 19775 in the control group older than age 12 in the year 2000. The numbers in the study and control groups older than age 12 in the year 2001 were 19716 and 19792 respectively. There is some attrition from the study in the year 2001, mostly because of the change of insurers that

---

[12]    Charges are best understood as the "retail" price of health care….a price that no one actually pays. Charges are the standard measure of cost used in health care studies, but their relationship to economic costs is never very clear. In our setting, we will be looking at the *difference* in charges between study and control group members, and this may reduce some of the noise introduced by these very imperfect cost measures.

takes place at the beginning of each calendar/contract year.  In both the study and control groups, roughly 72% of respondents stayed in the sample for all 12 months.

Panel B of Table 1 focuses on care considerations issued for the study group and detected, ex-post, in the control group.   In the study group 1299 CCs were issued compared to 1519 in the control group.  Five percent of study group members received CCs compared to 6% in the control group, and some of these participants received multiple CCs.  The mean number of CCs received by members who received any CCs was 1.3 in the study group and 1.25 in the control group.  In contrast the median for both the study and control groups was one.  Most of the CCs produced were of moderate severity, i.e. level two.  The most serious CCs were fortunately quite rare – only 0.1% of study group and 0.2% of control group patients received level-one CCs.

Panel C of Table 1 presents average costs in the study and the control group in the year prior to the experiment, i.e. the year 2000, and the experiment year, 2001.  These charges are discussed in more detail below, but it is worth noting that patients who received CCs accounted for a disproportionate share of medical expenditures.  In the study group, only 5% of patients received CCs, but these patients accounted for 17% of the costs.  Similarly in the control group we observe that the 6 percent of patients with CCs accounted for 21% of the total costs.

<u>Who Gets CCs?</u>

Table 2 examines who gets CCs, how many they get, and of what severity.  Column (1) is a probit with a dependent variable equal to one if the member received any CC's at all.  It is estimated for members of the study and control group because CC's were generated for both groups.  The first set of right hand side variables code for age.  The omitted group consists of those between age 12 and 20 and the coefficients are expressed as derivatives.  Thus a participant

aged 20 to 30 is 3.4 percentage points more likely to receive a CC than a participant in the omitted group. The likelihood of generating a CC increases with each subsequent decade and peaks for the oldest group. HMO members aged 60-65 are 35 percentage points more likely to generate a CC than those in the omitted group. Women are 0.5 percentage points less likely to generate a CC than are men, a change of roughly 10% of the mean. This gender differential probably is due to two factors: first, the software program did not have codes for many obstetric and gynecological issues and secondly, cardiac and other issues that were well represented in the software often manifest themselves a decade later in life for woman than for men. Since this study focused on a commercial insurance population less than 65 years old, this decade delay in onset would reduce the number of CCs generated for women. Finally, observe that participants with higher levels of charges in the year 2000 are more likely to generate CCs in the year 2001. The effect, while statistically significant, is also small. Moving from 0 charges in 2000 to the mean level of $280 increases the odds of generating a CC by 0.0028, an increase of 5.6% above the mean.

Taken together, the results in column (1) of Table 2 indicate that older, male patients with high medical charges in the previous year are more likely to have errors. The findings are consistent with the notion that care complexity is an important determinant of physician missteps. As bodies age, more things are likely to go wrong -- leading to more treatment and also more opportunities for lapses. Similarly, the more charges a patient generates, the greater the medical activity undertaken on their behalf. Managing these activities creates additional opportunities for errors.

The models estimated in columns (2) and (3) of Table 2 redo the analysis focusing respectively on the number of CCs received (a negative binomial model) and the severity of the

11

most severe CC received (an ordered probit model). In both cases we find that older patients, male patients and patients with more charges prior to the experiment are likely to generate more, and more severe, CCs. These results are both statistically and large in magnitude. They underscore the likely role that complexity of care plays in generating errors.

Cost Differentials

Table 3 analyzes the effect of the intervention on average charges per member per month. We adopt an "intention to treat" approach and compare the average charges in the study group and the treatment group. There are many possible treatment mechanisms in this study and it is hard to appropriately identify them all. As discussed above, we observe CCs generated by the software and approved by MD's working for the software company, but the HMOs nurse passed only a subset of these along to the treating physician. Similarly it is hard to know if the effect of the intervention was due to the information content of the particular CC or simply the fact that a physician received a CC at all. The "intention to treat" approach allows us to be agnostic about the mechanisms of action.[13]

Column (1) in Table 3 analyzes the average differences in total charges (pmpm) between the study group and the control group. The variable *Study* is a dummy variable equal to one if the participant is in the study group. The coefficient of –2.944 (*t = 0.35)* means that the average costs in the study group were $2.94 lower in the study group. This difference is both small and statistically insignificant. The variable *Year = 2001* is an indicator variable that is equal to one in 2001 and 0 in the year 2000. The coefficient 68.92 means that average costs rose from the pre-study year to the study year by $68.92 pmpm. The key variable that identifies the average

_____

[13]     An alternative measurement approach would be to compare the average costs of members in the
        study group who received CCs with members of the control group who *would* have received CCs.
        This approach is discussed in more detail later in the paper.

12

treatment effect is *Study*Year = 2001*. The coefficient reported implies that the increase in average costs from the pre-study year to the study year was $21.83 less in the study group than in the control group. Thus the intervention reduced the average of total charges in the study group by 6.1% of the average $352 pmpm control group charges. This difference is both economically significant and statistically significant at the 5% level.[14]

Columns (2)-(5) of Table 3 present estimates of the average treatment effect for the components of total charges. These are in-patient charges (charges incurred during hospitalization); out-patient charges, prescription (or Rx ) charges, and professional charges (charges resulting from professional services such as radiology). Focusing attention on the key variable, *Study*Year= 2001*, in-patient charges are reduced by $12.70 ($t = 1.78$) in the study group relative to the control. This accounts for 58% of the total cost differential. In contrast, out-patient and Rx cost differentials are quite small and statistically insignificant, -$1.84 ($t = 0.60$) and $0.70 ($t=0.90$) respectively. Professional charges, however, are $7.98 ($t= 2.21$) smaller in the study group.[15] The final column of Table (3) estimates a linear probability model of the determinants of hospitalization. The dependent variable is equal to one if the participant had ever been hospitalized and is 0 otherwise. The coefficient on *Study*Year = 2001* is very small and statistically insignificant, -0003 ($t=0.09$). This suggests that the reduction in in-patient costs observed in column (2) is likely due to reduced resource use when hospitalized or due to a

---

[14]    (Gertler and Simcoe, 2004) report that a disease management program for diabetes reduced costs by about 8 percent in the first twelve months and larger cost savings were realized in subsequent quarters.

[15]    As a check that the randomization "worked", we also re-estimated equations (1)-(5) using individual fixed effects. While the individual fixed effects had a statistically significant effect on costs, including them in the Table 3 estimates had virtually no effect on the estimates of the average treatment effect.

reduction in the number of hospitalizations in the year.  This result makes sense, as the likely effect of some of the CCs is to prevent re-hospitalization.[16]

The results in Table 3 imply that in-patient and professional charges account for 95% of the total cost differential between study and control groups.  In-patient charges arise from the use of hospital resources, but professional charges include services that can be delivered in either an inpatient or outpatient setting.   Our findings suggest that the experiment did not reduce all professional charges, but only those associated with hospitalization.  Taking the coefficient on *Study\*Year =2001* in column (2) and dividing it by the analogous coefficient in column (5) we get 1.6.  Thus every dollar decrease in professional charges that is due to being in the study group is associated with a $1.66 reduction in in-patient charges.  Similar calculations using results from column (3) of Table 3 suggest that a dollar reduction in professional charges resulting from the experimental intervention is associated with a drop in outpatient charges of only $0.23.

The results in Tables 3 indicate that the reduction in total charges is largely driven by inpatient costs and associated professional charges.  This suggests that the cost-savings generated in the experiment are the result of reduced costs for high-cost participants.  Table 4 examines this hypothesis through the use of quantile regressions.  Column (1) of Table 4 is a median regression.  The variable of interest is once again *Study\*Year=2001*.  The coefficient of 0.56 (*t=0.20*) suggests that the median participants in the study and control groups had virtually identical total charges (pmpm).   The corresponding coefficients in Column (2) suggest cost differences are $26.51 (*t* = 1.06) at the 90th percentile and $1658.61 (*t=1.85*) at the 99th

---

[16]    Examples of issues reported as CCs that were likely to prevent re-hospitalization include inadequate use of ACE inhibitors or beta-blockers for patients with myocardial infarctions or congestive heart failure.

percentile. Clearly the intervention is having its effect at the far right tail of the distribution of costs.

The experiment was concluded at the end of December 2001, but the system was kept in place for study group members until the end of February 2002. At that time the entire software system was turned off. In June 2002 the software was started up again and CCs were sent to all HMO enrollees, including those in the original study and control groups. The general rollout of the system makes possible an additional test of the system's effects on costs. If the reduction in charges observed in the study group was indeed the result of the intervention, one should expect charges in the two groups to converge when the controls began receiving CCs.

Table 5 compares charges in the study and control groups in the two years following the end of the experiment. Panel A of Table 5 analyzes cost data from calendar 2002. The coefficient on *Study* in column (1) indicates that average total charges in the study group were about $8.58 lower in the study than the control group. This difference is about 40% of that observed in the year of the experiment and it is imprecisely measured ($t = 0.78$) and not statistically distinct from 0. The corresponding coefficients for inpatient, outpatient, prescriptions and professional charges are presented in columns (2)-(4) respectively. They are similarly small, imprecisely measured, and not different from 0 at conventional significance levels. Column (6) is a probit where the dependent variable is 1 if a patient was ever hospitalized in the year and 0 otherwise. The probability of any hospitalization in 2002 was 0.5 percentage points lower in the study group than the control group ($z=2.27$). Column (7) is a quantile regression comparing the study and control groups at the 99th percentile. The coefficient on *Study* is –238.91, slightly more than a third of the analogous coefficient in Table 3 and imprecisely measured ($t=1.07$). Panel B compares the remaining members of the study and

control groups in the year 2003, two full years after the experiment. We find no statistically significant difference in charges between the two groups in any component of costs. Taken together, the absence of cost differentials in years when the intervention was rolled out to both treatment and control groups supports the conclusion that the cost differentials observed during the study year were the result of the intervention. These findings also suggest that the effect that the intervention had on average charges was both fast-acting (appearing in the first year of the study) and also quickly dissipated.

Resolution Rates:

Assuming that physicians will not respond to computer generated recommendations unless they believe them to be in the interest of patients, the rates of at which issues raised in CCs are resolved is a reasonable indicator of the effect of the software system on the quality of care.

The recommendations issued by the software fell into three, not-quite mutually exclusive, categories: "add a drug", "do a test", and "stop a drug". To identify whether an "add a drug" recommendation was complied with, the computer scanned pharmacy records following the recommendation. If a prescription for the indicated drug was filled, the issue was declared resolved. Similarly, billing records were scanned following a "do a test" recommendation. If bills for the suggested test were sent, the recommendation was also declared to be resolved. An important limitation of the "add a drug" resolution measure is that the database does not record whether or not the patient in question actually took the drug after buying it. Calculating resolution rates for the "stop a drug" recommendations was more problematic than for the other two categories of suggestions. Individuals might have months-long supplies of the drug at home and the records only tell us that no new prescriptions for the drugs were filled. To identify

16

compliance with "stop a drug" recommendations, pharmacy records were scanned for 60-150 days after the CC was transmitted, and the issue was declared resolved if no new scripts for the indicated medication were filled in that time.

We know from casual conversations with physicians that the software sometimes generated inappropriate recommendations that were ignored by the physician. In addition some problems identified by the software would have been discovered and resolved by caregivers in the absence of computer-generated messages. For both these reasons an adequate assessment of the clinical efficacy of the decision support tool requires a comparison of the resolution rates in the treatment and the control groups. Table 6 offers just such a comparison.

Column (1) of Table 7 compares the resolution rates for participants in the study and control groups who received an *"add a drug"* CC. The resolution rates were 8.6 percentage points higher in the study group ($z$=2.51) than the control. This is a 48 percent improvement over the control group's resolution rate of 0.18. Column (2) of Table 6 focuses on the "do a test" CCs. Here the study group's resolution rate is 5.8 percentage points higher than the control group ($z$= 2.24): an increase of 19% over the control group's resolution rates. Column (3) presents resolution rates for "stop a drug" CCs. The resolution rate in the study group appears to be *lower* than in the control group, but this differential is imprecisely measured ($z$=1.53) and one cannot reject the hypothesis that the true differential is zero.[17]

We do not, at present, know why there appears to be no effect of CCs on "stop a drug" resolution rates and substantial effects on "add a drug" and "take a test" resolution rates. One

---

[17]     In evaluating these resolution rate improvements, it is important to observe that it is hardly automatic that physicians respond to interventions in a positive way. (Shojania and Grimshaw, 2005/1/1) report in their assessment of quality improvement studies that interventions targeting provider behavior typically produce only modest improvements in compliance with care guidelines and the variation in results across studies is often large.

possible explanation is that many of the drug-drug inter-actions that trigger "stop a drug" CCs are also caught by pharmaceutical databases used by major pharmacies.

Why Were Extra CC's Triggered in the Control Group?

The treatment effect presented in columns (1) – (3) of Table 6 compares resolution rates for participants who triggered CC's in the study group with those who *would have triggered* CC's if they had been in the study group. This approach requires that the process for identifying CCs was the same in both groups and that the randomization "worked" in the sense that the clinical conditions of the study and control group participants were ex-ante similar. Under these conditions one would expect to find that, ex-post, CCs were equally prevalent in both groups – but the results in column (4) of Table 6 reject this hypothesis. Indeed, it appears that a patient was 1.1 percent less likely to receive a CC in the study group than in the control. If the "extra CCs" in the control group had different inherent resolution rates than those triggered in the study group, the estimates in Table 6 may be biased, and this bias might be either positive or negative.

There are at least four potential causes for excess CC's in the control group. First, it is possible that the randomization didn't work in the sense that a larger proportion of individuals with characteristics that trigger CCs ended up in the control group than the study group. From Table 2, we know that CCs were more likely for older males with a history of high charges and these were likely to have more complicated, and hence more expensive, conditions. If more individuals of this type were indeed sorted into the control group one would expect to see evidence of this in the cost differentials between groups. In Table 3, however, we report that the average cost differential between study and control groups is almost completely unchanged by the addition of fixed, participant effects. For this reason, it is unlikely that excess CCs are due to a flawed randomization process.

A second possible cause of excess CCs in the control group might be that the decision support tool was having the effect of reducing potential CC triggers in the study group. One can imagine that a physician receiving a CC might initiate actions that prevent additional CCs from being triggered. The evidence is, however, inconsistent with this hypothesis. From Table 1 it is clear that 80.96 percent of the control group CCs were issued to participants with only one CC compared to 76.87 percent in the study group. If extra scrutiny of a patient's file was indeed preventing additional CCs from being triggered, one should expect *a lower* fraction of participants with a single CC in the control group.

A third explanation for excess CCs arises from the way that the CCs were identified in the control group after the end of the experiment. The software company saved all the data generated by the controls and ran this data set through their software in early 2002. After CCs were triggered a group of physicians would look them over and send along those that made sense to them. The three physicians who comprised this group joined the software company in July of the year 2000 and played an increasingly important role in shaping the rules that triggered CCs. The clinical thinking of this group likely evolved over the year 2001, thus the decisions they made at the end of the experiment might have been different than the ones they made when evaluating the control group data in early 2002. If the thinking governing the issuance of CCs by this group became more expansive over time, and if these physicians applied their more recent understanding of the software rules to data generated for the control group early in the year, this would create excess CCs in the control group. This explanation for excess CCs, if correct, would not introduce a bias in our "intention to treat" estimate of cost savings from the experiment. It might, however, bias our estimates of the effect of the software on resolution rates. If, for example, the clinical conditions associated with excess CCs in the control group were such that

these were intrinsically easier to spot and resolve, our estimates in Table 6 would *understate* the improvements in resolution rates due to the experiment.

The fourth explanation for the excess CCs in the control group is random chance. Even if randomization "worked" in the sense that the patients were ex-ante identical in the study and control groups, there is still a 5 percent chance that random events would cause the observed differential in the number of CCs.

Cost Differentials Revisited:

If random chance is the cause of the excess CCs, then it is possible that our "intention to treat" estimates overstate the cost reductions attributable to the intervention. We know from Table 1, that the 5% of study members who triggered CCs accounted for 17 percent of total costs. Since members with CCs were more expensive than typical participants, it is possible that some of the cost savings were due to the presence of more of the expensive, CC-triggering participants in the control group.

In order to assess the potential magnitude of this bias, we decomposed the cost differential between the study and control groups as follows. We begin by noting that the average costs in the study and the control groups can be expressed by the following equation:

$$(1) \qquad Avg.\ Charges(pmpm)_g = \frac{\sum_{i=1}^{N_g} \hat{\beta}_g X_{ig}}{N_g} = \hat{\beta}_g \bar{X}_g$$

where *Avg. Charges* (*pmpm*)$_g$ is average charges (pmpm) in the study group (*g=study*) or the control group (*g=control*); and $\hat{\beta}_g$ is the vector of estimated coefficients from an OLS regression of charges on a vector of explanatory variables $X_{ig}$ for individual *i* in group *g*. $\bar{X}_g$ are the mean values of these variables averaged over the $N_g$ individuals in each group.

It is straight-forward to estimate (1) for the study and control group using as $X_g$ the patient characteristics in Table 2 (age, gender, baseline year charges) together with a vector of dummy variables indicating which care considerations were issued to that individual. Using the resulting estimates of $\hat{\beta}_g$ together with $\bar{X}_g$, the difference in charges between the study and control group can be decomposed as follows:

$$(2) \qquad Avg.\ Charges(pmpm)_s - Avg.\ Charges(pmpm)_c = (\hat{\beta}_s - \hat{\beta}_c)\bar{X}_s + \beta_c(\bar{X}_s - \bar{X}_c)$$

where subscripts $s$ and $c$ refer to study and control groups respectively. If the same conditions triggered CCs in both the study group and control group, the first term on the right hand side of equation (2) is that part of the average treatment effect attributable to cross-group differences in the way charges are linked to CCs (and other observable patient characteristics).[18] The second term is that part of the average treatment effect due to the groups having different patterns of CCs (and other observable patient characteristics).

To better understand equation (2), consider a hypothetical care consideration that recommends use of a drug that can prevent a re-hospitalization. Assume that this care consideration, called $z$ for convenience, is associated with a $50.00 increase in costs (pmpm) in the study group because the medicine is moderately expensive. In the control group, care consideration $z$ is associated with a $1000.00 increase in average costs (pmpm) because the hospital stay resulting from a failure to take the recommended medicine is even more expensive than the medication. If 1% of the members of the study group received consideration $z$, then

---

[18]  Interpretation is more difficult if the excess CCs in the control group were the result of subtle cross-group differences in the criteria that triggered CCs. Imagine, for example, that the rules triggering CCs in the control group caused less expensive patients to initiate CCs. In this case, term one of equation (2) would under-estimate that part of the average treatment effect attributable to cross-group differences in the way charges are linked to CCs (and other observable patient characteristics).

according to the first term in equation (2), care consideration $z$ would produce a $9.50 reduction in average costs in the study relative to the control groups (-$950*0.01).

The second term on the right-hand side of equation (2) is the difference in charges due to cross-group differences in patient characteristics and care considerations (evaluated using the control groups charges equation). Returning to our hypothetical CC in the previous paragraph, if the incidence of $z$ is 0.1% higher in the control group than the study group, this higher incidence would reduce average pmpm costs in the study group by $1.00 (0.001*$1000) relative to the control group.[19]

We implemented the decomposition described in equation (2) using estimates of $\hat{\beta}_g$ based on the characteristics of patients (age, gender, and baseline charges) and the set of CCs generated. The coefficients on most of the specific care consideration dummies were statistically insignificant, in part because the numbers receiving a specific type of CC were often small. Nevertheless, of the $21.83 pmpm differential between the study and the control groups, roughly 80% is due to systematic cross-group differences in the way charges are determined, i.e. to the first term on the right-hand side of equation (2).

Figure 1 presents the first term in equation (2) for members of the study group. Each dot on the graph represents an individual in the study group who triggered a CC. The horizontal axis is the predicted charges associated with that individual's characteristics and care considerations in the control group. The vertical axis is the average charges that the same characteristics and care considerations produce in the study group. If charges related to characteristics and care considerations in the same way in both the study and control groups, all the dots would lie along

---

[19]     In this example $(\bar{X}_s - \bar{X}_c) = -0.001$ and $\hat{\beta}_c = \$1,000$, so $(\bar{X}_s - \bar{X}_c)\,\hat{\beta}_c = -\$1.00$

the 45-degree line that passes through the origin. Indeed, most dots do lie along this line. As the level of control group charges increase, however, there is an increasing mass of points lying below the 45-degree line. This visual impression is reinforced when we fit a regression line to the data. As is clear from the Figure, the intercept of the regression line is greater than zero and the slope is less than one.[20] Thus, consistent with our quantile regression results, the decision support tool is saving the most money for the most expensive patients – and this difference is due in part to differences in the costs associated with specific CCs in the study and the control group.

## 4. Conclusions

This paper analyses the effect of a decision support tool designed to help physicians detect and correct medical "errors". Research suggests that physician errors have a substantial effect on the cost and quality of medical care, and a number of high-profile public and private initiatives are premised on the notion that new information technologies can reduce the incidence of errors. Economic studies of the efficacy of these technological fixes have, however, been scarce.

The data in this study comes from a *randomized* trial of the new technology in a population of commercial HMO patients. The key findings of the study are as follows. The new information technology enhances the efficiency of care by reducing costs while also improving care quality. Average charges were 6% *lower* in the study group than in the control group. These savings were the result of reduced in-patient charges (and associated professional charges) for the most costly patients. The rate at which potential errors were resolved was

---

[20]   The slope and intercept for the regression line in Figure 1 are  0.595 ($\sigma$=0.152) and 369.720 ($\sigma$= 160.134) respectively.  Because the regression line is the result of regressing one estimate of costs against another, conventional standard errors are inappropriate and we rely instead on bootstrapped standard errors.

generally higher in the study group than in the control group—suggesting an improvement in care quality.

The conclusion that error-spotting information technology reduces average costs and increases the rate at which potential problems are resolved is noteworthy, but the limitations of this study leave a number of important issues unresolved. First among these is the short time horizon of the trial. To the extent that some of the benefits of spotting errors spill over into future years, this analysis understates the cost saving potential of the physician decision support system.[21]

A second limitation of this study is that it was conducted on a commercial insurance population where everyone was less than sixty-five years old. Since the likelihood of errors increases dramatically with age, much of the impact of this new technology will be found in older age groups not included in this study. For this reason, trials similar to the one reported here are currently being conducted for Medicare populations aged sixty-five and older.

A final issue left unresolved by this study is the mechanism by which care considerations influence outcomes. Specifically, the analysis does not identify the lessons physicians learned from the messages they received. It is possible that physicians learned only that the patient named in the care consideration required additional attention. Alternatively, it might be that the specific clinical content of the care considerations sometimes imparted new and useful information to the physician. This latter possibility suggests that widespread implementation of computer based decision support systems may increase the dynamic efficiency of the health care system by increasing the rate at which new knowledge is diffused. Understanding the effect of

---

[21]    (Gertler and Simcoe, 2004) find that cost improvements in a diabetes disease management program became apparent only after the first six months of the intervention and that most cost savings occurred beyond the first year.

IT based decision support tools on the diffusion of new medical knowledge will be the subject of future investigations.

Looking beyond medical costs and care quality, this study also adds to the growing body of research documenting the multifarious ways that the information technology revolution complements work involving non-routine problem solving and communication. Specifically IT enabled support tools can improve the economic efficiency of decisions made by highly trained professionals making complex decisions in challenging environments.

## Table 1
## Descriptive Statistics

| | Study | Control |
|---|---|---|
| **Panel A** | | |
| Number of Members in 2000 > 12 years old | 19693 | 19775 |
| Number of Members in 2001 > 12 years old | 19716 | 19792 |
| Fraction of Members in Study for all 12 Months 2001 | 0.729661 | 0.723727 |
| **Panel B** | | |
| Number of CC's Issued | 1299 | 1519 |
| Fraction of Members with at least 1 CC in 2001 | 0.050 | 0.061 |
| Distribution of CC's Among Members who Have Any CCs | | |
| Percent of Members having any CC who have 1 CC | 76.87 | 80.96 |
| Percent of Members having any CC who have 2 CCs | 16.46 | 14.56 |
| Percent of Members having any CC who have 3 CCs | 4.55 | 3.17 |
| Percent of Members having any CC who have 4 CCs | 2.12 | 1.14 |
| Percent of Members having any CC who have 5 CCs | 0 | 0.16 |
| Mean CC's for Members with CC's | 1.321 | 1.253 |
| Number of Distinct Types of CC's Issued | 90 | 83 |
| Severity of CCs | | |
| Fraction of Members with at least 1 Level 1 CC in 2001 (most serious) | 0.001 | 0.002 |
| Fraction of Members with at least 1 Level 2 CC in 2001 (less serious) | 0.036 | 0.042 |
| Fraction of Members with at least 1 Level 3 CC in 2001 (least serious) | 0.019 | 0.024 |
| **Panel C** | | |
| Average Charges in 2001 (pmpm) | | |
| Total Charges | 327.54 | 352.31 |
| In-patient Charges | 58.15 | 72.06 |
| Out-patient Charges | 71.69 | 74.11 |
| Rx Charges | 65.21 | 65.27 |
| Professional Charges | 132.48 | 140.87 |
| Fraction of All Charges in 2001 due to patients with CCs in 2001 | 0.17 | 0.21 |
| Average Charges in 2000 (pmpm) | | |
| Total Charges | 280.45 | 283.39 |
| In-patient Charges | 58.23 | 59.43 |
| Out-patient Charges | 57.39 | 57.97 |
| Rx Charges | 47.13 | 47.88 |
| Professional Charges | 117.71 | 118.11 |

## Table 2
## Who Recieved Care Considerations?

| | Probit Receive any CC? | Negative Binomial Number CC's Received | Ordered Probit Severity of CCs |
|---|---|---|---|
| 20<AGE≤ 30 | 0.034 | 2.553 | -0.323 |
| | (4.8) | (4.95) | -(4.61) |
| 30 ≤AGE < 40 | 0.060 | 4.589 | -0.543 |
| | (8.73) | (8.77) | -(8.56) |
| 40 ≤ AGE < 50 | 0.111 | 9.599 | -0.880 |
| | (14.7) | (13.52) | -(14.44) |
| 50 ≤ AGE < 60 | 0.206 | 18.824 | -1.194 |
| | (20.11) | (17.57) | -(19.46) |
| 60 ≤ AGE | 0.352 | 30.934 | -1.520 |
| | (23.96) | (20.17) | -(23.33) |
| Female | -0.005 | 0.865 | 0.0561 |
| | (-2.83) | (-3.19) | (2.56) |
| Charges (pmpm) in 2000 | 0.00001 | 1.00041 | -0.0001 |
| | (8.18) | (13.4) | -(8.94) |
| | | | |
| Observations | 39468 | 39468 | 39468 |
| Log pseudo-likelihood | -7425.479 | -8882.608 | -9061.594 |

Column (1) is a probit with coefficients expressed as derivatives. For dummy variables this is discrete change from 01.
For continuous variables the derivative is evaluated at the mean.
Column (2) is a negative binomial count model of the number of CCs received. Parameter $\alpha = 3.3486$.
  the coefficients are expressed as incident rate ratios so that the number of cc's for those 20-30 is 2.55 times that of the omitted age group.
Colum (3) is an ordered probit of an indicator of cc severity. CC's were ranked from least (3) to most (1) dangerous. Those
 ' with no CCs were given a 4. Members were assigned the level of the most dangerous CC they recieved.
The omitted age category is teenagers between 12 and 20.

Numbers in ( ) are z scores

Table 3
The Effect of Exposure to the Intervention on Average Costs (pmpm)

| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS | (6) OLS |
|---|---|---|---|---|---|---|
| | Total Charges (pmpm) | In-patient Charges (pmpm) | Out-patient Charges (pmpm) | Rx Charges (pmpm) | Professional Charges (pmpm) | In Hospital |
| Study | -2.944 | -1.201 | -0.578 | -0.758 | -0.407 | -0.003 |
| | (0.35) | (0.25) | (0.28) | (0.80) | (0.13) | (1.40) |
| Year = 2001 | 68.921 | 12.63 | 16.143 | 17.387 | 22.762 | 0.002 |
| | (8.10)** | (2.18)* | (6.97)** | (30.93)** | (8.60)** | (0.81) |
| Study*Year = 2001 | -21.833 | -12.705 | -1.845 | 0.701 | -7.983 | -0.0003 |
| | (1.98)* | (1.78) # | (0.60) | (0.90) | (2.21)* | (0.09) |
| Constant | 283.392 | 59.43 | 57.97 | 47.88 | 118.112 | 0.052 |
| | (48.37)** | (16.92)** | (39.74)** | (67.56)** | (53.90)** | ( 32.88 )** |
| Observations | 78976 | 78976 | 78976 | 78976 | 78976 | 78976 |
| Individuals | 39508 | 39508 | 39508 | 39508 | 39508 | 39508 |
| R-squared | 0.001100 | 0.0001 | 0.0009 | 0.0062 | 0.0008 | 0.0001 |

Absolute Value of robust t-statistics in parentheses. t-statistics are adjusted for clustering by individual.
# significant at 10% * significant at 5%; ** significant at 1%

Re-estimating columns (1) - (5) with individual fixed effects produces very similar results.
  The coefficients (t-statistics) for  Study*Year  for the fixed effects version of columns (1)-(5)  are respectively:
 -21.92 (-1.99);  -12.83 (-1.80); -1.82 ( -0.60); .700 ( 0.90) and -7.96 (-2.20)

## Table 4
### Exposure to Technology Reduces Costs at the Far-Right Tail of the Distribution of Costs

| | (1) | (2) | (13) |
|---|---|---|---|
| | | Quantile regressions | |
| | Median Total Charges (pmpm) | 90th %tile Total Charges (pmpm) | 99th %tile Total Charges (pmpm) |
| *Study* | -0.541 | -20.942 | 158.436 |
| | (0.36) | (1.31) | (0.74) |
| *Year = 2001* | 13.910 | 166.308 | 962.548 |
| | (7.16)** | (9.07)** | (3.49)** |
| *Study*Year = 2001* | 0.561 | -26.512 | -658.612 |
| | (0.20) | (1.06) | (1.85)# |
| Constant | 86.47083 | 636.9142 | 3189.613 |
| | (77.92) | (54.00) | (19.50) |
| Observations | 78976 | 78976 | 78976 |
| Individuals | 39508 | 39508 | 39508 |
| R-squared | 0.0005 | 0.0032 | 0.0037 |

Quantile regressions estimated for the median, 90th and 99th percentiles respectively.
  The standard errors for these regressions were bootstrapped with 1000 repetitions
 The R-squared for the quantile regressions are pseudo R-squared.

Table 5
Charges in the Study and Control Groups After the Care Engine was Rolled Out to Both Groups

Charges In 2002

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) Quantile Regression 99th percentile |
|---|---|---|---|---|---|---|---|
| | OLS | OLS | OLS Out-patient Charges (pmpm) | OLS | OLS Profession al Charges (pmpm) | Probit | percentile |
| | Total Charges (pmpm) | In-patient Charges (pmpm) | | Rx Charges (pmpm) | | In Hospital | Total Charges (pmpm) |
| Study | -8.528 | -5.796 | 2.416 | -1.783 | -3.365 | -0.005 | -238.91 |
| | (0.78) | (0.86) | (0.89) | (1.29) | (0.90) | (2.27)* | (1.07) |
| Constant | 338.544 | 70.051 | 65.342 | 65.258 | 137.891 | | 3948.585 |
| | (41.94)** | (13.89)** | (33.61)** | (62.98)** | (49.50)** | | (25.23)** |
| Observations | 38056 | 38056 | 38056 | 38056 | 38056 | 38056 | 38056 |
| $R^2$ | 0 | 0 | 0 | 0 | 0 | 0.0004 | |

Charges in 2003

| | (1) | (2) | (3) | (4) | (5) | (6) | Quantile Regression 99th percentile |
|---|---|---|---|---|---|---|---|
| | OLS | OLS | OLS Out-patient Charges (pmpm) | OLS | OLS Profession al Charges (pmpm) | Probit | percentile |
| | Total Charges (pmpm) | In-patient Charges (pmpm) | | Rx Charges (pmpm) | | In Hospital | Total Charges (pmpm) |
| Study | 15.574 | 14.072 | 0.589 | 2.273 | -1.36 | 0.001 | -184.170 |
| | (0.94) | (1.45) | (0.12) | (1.17) | (0.24) | (0.22) | (0.39) |
| Constant | 379.225 | 78.716 | 88.1 | 50.632 | 161.776 | | 5214.387 |
| | (37.25)** | (14.55)** | (25.85)** | (39.57)** | (43.77)** | | (16.03)** |
| Observations | 27288 | 27288 | 27288 | 27288 | 27288 | 27288 | 27288 |
| $R^2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Robust t statistics in parentheses in columns (1)-(5) and (7). Column (6) presents z statstics.
The standard errors for the quantil regressions were calculated by bootstrapping.
The $R^2$ statistics in columns (6) and (7) are pseudo $R^2$.
* significant at 5%; ** significant at 1%


The experiment concluded at the end of December 2001, but the "care engine" remained on for two months
At the end of February, the entire system was "turned off" and in June the system was rolled out to all members.
Panel A compares charges in the study and control group over the year 2002 and Panel B compares
these groups in 2003.

Table 6

The Effect of the Experiment on Resolution Rates and the Probability of Recieving any Care Consideration
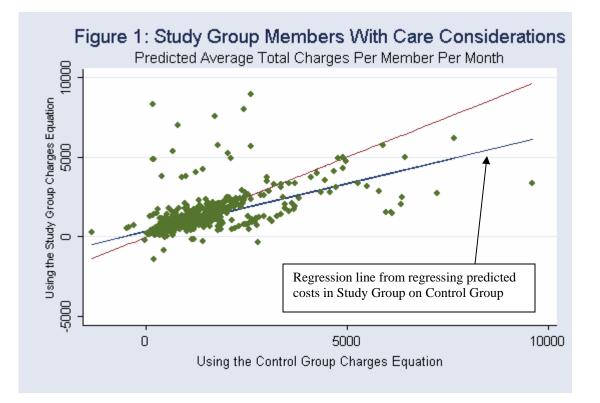
| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Probit | Probit | Probit | Probit |
| | *Successful Resolution Any "Add A Drug" CC* | *Successful Resolution Any "Do a Test" CC* | *Successful Resolution Any "Stop a Drug" CC* | *Any CCs* |
| | [0.18] | [0.31] | | [ 0.061] |
| *Study* | 0.086 | 0.058 | -0.060 | -0.011 |
| | (2.51)* | (2.24)* | (1.53) | (4.94)** |
| Observations (only for 2001) | 601 | 1354 | 592 | 39508 |

Robust z statistics in parentheses.  [ ] is mean of dep. var. in the control group in 2001
* significant at 5%; ** significant at 1%

Probits in column (1) are estimated for all members who received at least one "add a drug" CC in 2001
Probits in columns (2) and (3) are for "Do a Test" CCs and "Receive Any CC" respectively.

Coefficients are expressed as "derivatives.    Thus 0.18 of the control group who received an
  "add a drug" CC in 2001, had a successful resolution of an 'Add a Drug' CC.  The resolution
 rate in the study group was 8.6 percentage points higher or 0.266.

The lower incidence of *Receive Any CCs"* in the study group are nearly identical if we reestimate (4)
 as a linear probability model with member or with physician fixed effects.

Figure 1: Study Group Members With Care Considerations

Predicted Average Total Charges Per Member Per Month

Regression line from regressing predicted costs in Study Group on Control Group

## References

ATHEY, S. AND S. STERN. "The Impact of Information Technology on Emergency Health Care Outcomes," Autumn 2002, *RAND Journal of Economics, Autumn* 33 (3), 2002, pp. 399-432.

AUTOR, D. H., F. LEVY AND R. J. MURNANE. "Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank," April 2002, *Industrial and Labor Relations Review* 55 (3), April, 2002, pp. 432-47.

AUTOR, D. H., F. LEVY AND R. J. MURNANE. "The Skill Content of Recent Technological Change: An Empirical Exploration," November 2003, *Quarterly Journal of Economics* 118 (4), November, 2003, pp. 1279-1333.

BARTEL, A., C. ICHNIOWSKI AND K. SHAW. "The Strategic Investment in Information Technologies and New Human Resource." NBER Summer Institute: Working Paper, July, 2004.

BOHMER, R. "Complexity and Error in Medicine," *Harvard Business School Teaching Note* (9-699-024), November 6, 1998, p. 16.

BRYNJOLFSSON, E. AND L. M. HITT. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," Fall 2000, *Journal of Economic Perspectives, Fall* 14 (4), 2000, pp. 23-48.

CUTLER, D. M. *Your Money or Your Life: Strong Medicine for America's Health Care System*. Oxford, England: Oxford University Press, 2004.

GAYNOR, M., J. B. REBITZER AND L. J. TAYLOR. "Physician Incentives in HMOs," *Journal of Political Economy* 112 (4), August, 2004.

GERTLER, P. AND T. SIMCOE. "Disease Management: Using Standards and Information Technology to Improve Medical Care Productivity," *Working Paper*, March 2004.

HUBBARD, T. N. "Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking," September 2003, *American Economic Review* 93 (4), September, 2003, pp. 1328-53.

ICHNIOWSKI, C. AND K. SHAW. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *Journal of Economic Perspectives* 17 (1), Winter, 2003, p. 155. American Economic Association.

INSTITUTE OF MEDICINE COMMITTEE ON QUALITY OF HEALTH CARE IN AMERICA. *To Err is Human: Building a Safer Health System* (L. T. Kohn, J. M. Corrigan and M. S. Donaldson, eds.). Washington DC: National Academy Press, 2000.

INSTITUTE OF MEDICINE COMMITTEE ON QUALITY OF HEALTH CARE IN AMERICAN. *Crossing the Quality Chasm: A*. Washington D.C,: National Academy Press, 2001.

JAVITT, J. ET AL. "Use of a Claims Data-Based Sentinel System to Improve Compliance with Clinical Guidelines: Results of a Randomized Prospective Study," *Am J Managed Care* (11), 2005, pp. 21-31.

LANDRIGAN, C. P. ET AL. "Effect of Reducing Interns' Work Hours on Serious Medical Errors in Intensive Care Units," *The New England journal of medicine* 351 (18), 2004, p. 1838.

LAPETINA, E. M. AND E. M. ARMSTRONG. "Preventing Errors In The Outpatient Setting: A Tale Of Three States," *Health Affairs* 21 (4), 2002, p. 26.

LEVY, F. AND R. J. MURNANE. *The New Division of Labor*. Princeton, New Jersey: Princeton University Press, 2004.

NEWHOUSE, J. P. "Why Is There A Quality Chasm," *Health Affairs* 21 (4), 2002, p. 13.

PRESIDENT'S INFORMATION TECHNOLOGY ADVISORY COMMITTEE. *National Coordination Office for Information*, June 30 2004*Revolutionizing Health Care Through Information Technology*. Arlington, VA.

SHOJANIA, K. G. AND J. M. GRIMSHAW. "Evidence-Based Quality Improvement: The State Of The Science," *Health Affairs* 24 (1), 2005/1/1, p. 138.

SOLOW, R. "We'd Better Watch Out," *New York Times*, July, 12 1987, p. 36. Book Review.

THE COUNCIL OF ECONOMIC ADVISORS. *Economic Report of the President*. Washington, DC: United States Government Printing Office, 2004.

THE LEAP FROG GROUP FOR PATIENT SAFETY. "Fact Sheet," Http://Www.Leapfroggroup.Org/FactSheets/LF_FactSheet.Pdf, August, 2004.