Methodological Approaches to Measuring the Effects of Implementation of Health Information Technology (HIT)

Lance Roberts[1], Marcia M. Ward[1], Jane M. Brokel[1,2], Douglas S. Wakefield[3], Donald K. Crandall[2], and Paul Conlon[2]
[1]University of Iowa, [2]Trinity Health, and [3]University of Missouri

Abstract

The research design, evaluation methodology, and statistical analysis of the clinical efficacy of healthcare information technology (HIT) implementation can be a challenging task. Much of the research to date involves weakly designed studies. We discuss some of the more rudimentary experimental designs and analytical approaches that have typically been used. Approaches to strengthen a research design include: adding a matched control unit or hospital; using multiple observations before and after the HIT implementation; making observations across a number of hospitals that are implementing the same HIT application; comparing the changes in these hospitals to a matched set that have not yet implemented the HIT application; and applying statistical approaches that permit changes in trends over time to be examined. Here we report our use of linear piecewise spline mixed effects models and compare our models to other methodological approaches that could be used for this evaluation.

1. Introduction

Like many healthcare organizations, Trinity Health was faced with the need to proactively and strategically address current healthcare delivery issues such as increasing demand, changing demographics, keeping pace with new and changing drugs/devices, cost escalation, and societal expectations. Complicating these issues was the fact that Trinity Health was formed through the merger of two health systems resulting in an organization spanning 44 hospitals and a range of other healthcare facilities in seven states. As a result they inherited a mix of hardware platforms, legacy systems, and vendors that elevated operational complexity. Their strategic vision to transform the way care is delivered, improve patient safety, and simultaneously address revenue/cost issues included an initiative to combine state-of-the-art, commercially-available, standardized healthcare information technology (HIT) systems with best-practice business processes. Trinity Health's system-wide strategy – termed "Project Genesis" – included the replacement of a variety of legacy business and clinical workflow HIT applications from multiple different vendors with an integrated set of over 20 HIT applications, primarily from two widely-used commercial vendors [1].

Given the nature of their substantial investment in HIT tools and management processes, it was important to Trinity Health to measure the effect of this HIT application on clinical processes and/or outcomes. One of the clinical processes we analyzed was the effect of the HIT implementation on the detection of potential adverse drug events (ADEs) in the medication ordering stage including potential drug allergies, drug-drug interactions, and drug problems when the patient's laboratory results are outside the normal range.

Salient studies investigating the nature of adverse events in hospitalized patients found that the most common events are related to drug complications. Early studies found that the incidence of ADEs comprise nearly 20% of all types of adverse events studied [2,3]. Subsequently, the U.S. Institute of Medicine's Committee on Identifying and Preventing Medication Errors was tasked to carry out a study on medication errors. They conservatively estimated that a hospital patient is at risk for at least one medication error per day and that eliminating preventable hospital ADEs would save approximately $3.5 billion (2006 dollars) annually [4-6]. The IOM committee suggested that the level and consequences of medication errors are unacceptable and could be remedied in the hospital setting by the effective integration of such HIT system strategies as computerized provider order entry (CPOE) with clinical decision-support systems (CDSS) [4].

To date, surprisingly, most of the research on CPOE and CDSS has focused on the effectiveness of "homegrown" HIT applications developed in-house by technology champions in academic or large institutional settings [7-9]. Because the findings from existing research may not generalize to most healthcare purchasers, there is considerable need for research focused on the commercial HIT applications that are purchased and used most frequently [7-9].

Trinity Health received a grant from the Agency for Healthcare Research and Quality to examine the

978-0-7695-3450-3/09 $25.00 © 2009 IEEE

effects of implementation of CPOE and CDSS in one of their hospitals – Mercy Medical Center–North Iowa (MMC-NI). This grant presented an exciting opportunity to closely investigate the changes in processes and outcomes resulting from implementation of a set of HIT applications available from current vendors having large market shares. However, conducting meaningful research and analyses on the effects of one hospital's implementation is challenging.

Although the implementation of HIT applications are recognized as valuable tools that may assist healthcare professionals in addressing patient safety issues like ADEs, safety experts caution users to evaluate the implementation of such innovative technologies from a systems viewpoint. They suggest that in an optimal situation, decisions to implement any change to a system should be based upon the review of sound evidence resulting from well-designed randomized controlled trials utilizing patients. Proponents of evidence-based clinical practice point to these types of trials as yielding the highest grade of evidence of the effectiveness of any treatment [10]. However, the ability to conduct high-quality trials that study technically complex tools at a systems level is largely prohibitive from a cost standpoint. Therefore, for many interventions there is a lack of evidence on which to base decision making. When the evidence or the options to conduct high-grade experiments is constrained, experts suggest that for some types of innovations evidence from randomized trials is neither sufficient nor necessary for acceptance of a practice [11].

However, experts also caution against taking an overly simplistic viewpoint when evaluating technologically and systematically complex systems. For example, Shojania and colleagues point to the lack of studies on commercially available CPOE implementations and the unintended consequences that may easily arise in complex systems [12]. They also point to CPOE studies that demonstrate the importance of carefully examining the impact of technically complex innovations on systems of care and the human factor elements that must be assessed when marrying technology with systems that, by their nature, are comprised of a large amount of human interaction. They provide examples of unanticipated events such as the complete failure of a CPOE implementation, clinicians bypassing the CPOE system by reverting back to handwritten orders, and the human factors-related unanticipated consequences of an increase in potential ADEs attributed to the design of the drug ordering screen for potassium [13-15]. Thus, Shojania and colleagues suggest that "one of the major lessons of the safety literature is that this year's safety measure often contributes to next year's incident" [12].

Thus, the challenge for researchers studying the safety, efficacy, and effectiveness of complex HIT innovations is to carefully choose a research design that is not overly simplistic, too rigorous and costly, yet effectively analyzes the "harmony" of the innovation within its system. The National Academy of Engineering and the Institute of Medicine have stated "the U.S. health care enterprise has devoted relatively little technical talent, material resources, or intellectual effort to improving or optimizing the operations of health care systems…" and suggest that there should be a new partnership between engineering and healthcare to promote the application of "the principles, tools, and research from engineering disciplines associated with the analysis, design, and control of complex systems." In this paper we compare the relatively simplistic analytical approaches that could have been used to analyze Trinity Health's implementation of HIT with the data, designs, and statistical tools we used to effectively analyze the introduction of a technically and systematically complex implementation of information technology in a healthcare system. In essence, the analytical approach we used integrates (1) some key engineering-derived tools – statistical process control run charts designed to allow a visual examination of fluctuations in process over time and help identify if those fluctuations are due to random events or a systematic change; (2) a human factors approach that considers the effect of an innovation implementation upon the human interactions within the system; (3) the capture of robust data that enables stronger analyses of system performance; and (4) appropriate quantitative statistical tools designed to analyze and interpret system models [16].

In particular, we used a number of approaches to strengthen our research design to help make the findings more interpretable [10]. First, we added a single, matched, control hospital. Second, where possible, we examined multiple observations before and after implementation. Third, where common metrics had been collected system-wide, we grouped findings across a number of Trinity Health hospitals that were implementing the same HIT applications at different times. Fourth, we compared the changes in these hospitals to a matched set of hospitals that had not yet implemented the HIT applications. Finally, we applied statistical approaches that let us examine changes in trends over time. Here we report our use of linear piecewise spline mixed effects models and compare our models to other methodological approaches that could have been used for this evaluation. For the analyses presented here, we focus on metrics related to the prevention of ADEs.

2. Background

Trinity Health's comprehensive HIT implementation – Project Genesis – represents the "treatment" intervention analyzed in this study. This treatment included the implementation of new patient management systems and a suite of clinical applications. Trinity Health uses an extensive readiness approach involving a standardized and phased methodology [1].

Phase One of Project Genesis consists of installing the central clinical data repository; interfaces for dictated reports, existing pharmacy system patient drug profiles, and laboratory results; a results viewer (Cerner PowerChart); and an ADE rules package. During this time, Trinity Health also implemented an Enterprise Master Person Index system to establish unique patient identifiers for the enterprise-wide systems and the PeopleSoft Enterprise Resource Planning system [1].

Phase Two of Project Genesis consists of a new patient management system and a suite of clinical applications (Clinical Documentation, CPOE with pre-developed service specific order sets, a new pharmacy system, medical records system, emergency department system, and radiology system), plus a more sophisticated CDSS consisting of a Multum™ (Cerner Corporation) drug database working alongside their original ADE logic modules (an in-house developed rules package), and over 250 evidence-based order sets designed by Zynx Health [1].

Rather than an incremental approach, the Project Genesis plan uses a "Big Bang" approach for Phase Two which consists of bringing down all the old systems, implementing the new system, inputting current patient data into the new system, and converting users to the new system for all clinical areas at the same time, and over the course of a single weekend.

MMC-NI went through Phase One from February 2003 to July 2005. MMC-NI used the Big Bang approach for Phase Two implementation ("Go-Live") on July 8, 2005 [1].

The implementation of these HIT tools substantially changes the process by which potential ADEs are identified and processed during the medication ordering stage. To monitor the medication ordering process and how potential ADE alerts are handled, Trinity Health has been collecting data on a number of ADE alert process measures for several years. The first 4 metrics reflect specific steps in the process used by physicians and pharmacists to detect and address potential ADEs – 1) *Number of Potential ADE Alerts Printed* (the number of potential ADEs identified by the system and printed for pharmacist review), 2) *Number of Potential True Positive Alerts* (the pharmacist reviewed the printed output and determined that an ADE was potential if the medication order was activated and that the physician needed to be consulted), 3) *Number of Times Physicians Were Contacted* (the pharmacist contacted the physician to discuss the potential ADE alert), 4) *Number of Times Physician Agreed with Recommendation* (the physician agreed with the pharmacist's recommendation to change a particular medication therapy). The remaining two metrics reflect summary measures. In particular, the 5) *Ratio of Orders Changed to Physician Contacts* (Number of Times Physician Agreed with Recommendation / Number of Times Physicians Were Contacted x 100) was calculated to measure the relative yield of the ADE identification and verification process. Finally, the 6) *Number of Potential True Positive Alerts per 1000 Admissions* (Number of Potential True Positive Alerts / Average monthly hospital admissions x 1000) was created to adjust for fluctuations in inpatient census.

Because Trinity Health had been collecting information in a standardized manner on each of these metrics for a number of years preceding and during Phase 1 of Project Genesis at all of their hospitals, these data provided us with an excellent opportunity to examine clinical processes and outcomes affected by the HIT implementation.

3. Evaluative Methods

3.1. Single Group Differences over Time

A simple method one might use to analyze the effectiveness of the Project Genesis implementation at MMC-NI would be to measure the dependent variable before and after key changes, such as the implementation of the Phase Two clinical systems. This is essentially a one-group pretest-posttest design utilizing a Paired T-test statistical approach to evaluate the difference in the hospital's process scores over time - before and after the Project Genesis Phase Two implementation. The results of analyses for each of the six ADE process metrics are shown in Table 1. We observe statistically significant differences before and after the HIT implementation in the mean difference for one of the metrics, but not for the other five metrics. However, a weakness in this one-group pretest-posttest design and analysis is that we cannot control by design the effects of history or maturation on the treatment hospital [17]. Thus, the mean difference in each

ADE process measure may not be attributable to the HIT implementation.

Table 1. Single group results: Paired T-test

	Mean Difference	p
Number of Potential ADE Alerts Printed	335.80	< 0.001
Number of Potential True Positive Alerts	-0.25	0.97
Number of Times Physicians Were Contacted	-3.99	0.56
Number of Times Physician Agreed with Recommendation	-1.66	0.80
Ratio of Orders Changed to Physician Contacts	3.71	0.18
Number of True Positives per 1000 Admissions	-2.40	0.73

It is clear from these simple analyses that the metric with the most robust changes was the *Number of Potential ADE Alerts Printed*. Thus, we use that metric to demonstrate additional approaches.

3.2. Adding a Matched Control Hospital

The weaknesses in this approach are greatly reduced when the changes over time in the treatment unit are compared with a control group that did not undergo an HIT implementation, but was similar to the treatment unit in other ways. In our grant proposal, we had identified Mercy Medical Center at Clinton Iowa (MMC-Clinton) as our matched control hospital. It is also part of Trinity Health, but it was not scheduled to undergo the HIT implementation until 18 months after the intervention hospital – MMC-NI. It is also a rural referral hospital located in the same state, and thus was well matched to the intervention hospital in a number of important ways.

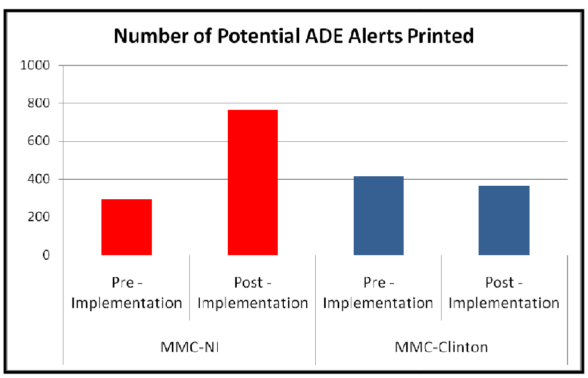

Figure 1. Mean number of potential ADE alerts printed before and after implementation in the intervention hospital and during a comparable period in the control hospital

To compare the two hospitals, we collected data from MMC-Clinton during the same time period as for MMC-NI. Thus, the pre-implementation and post-implementation time periods for MMC-Clinton identified here were not its own (which occurred much later), but the time periods that matched the Phase Two implementation schedule for MMC-NI. As can be seen in Figure 1, the two hospitals were statistically different, on the number of potential ADE alerts printed during MMC-NI's pre-implementation period. We can also observe that the *Number of Potential ADE Alerts Printed* during the post-implementation period at MMC-NI showed a dramatic increase, but remained quite constant at MMC-Clinton.

Table 2. Mean number of potential ADE alerts printed before and after implementation in the intervention hospital and during a comparable period in the control hospital – Cross-sectional results for Mann-Whitney U Test

	MMC-NI Mean	MMC-Clinton Mean	Δ	p-value
Pre-Implementation	293.0	415.2	-122.2	< 0.001
Post-Implementation	765.1	364.1	401.0	< 0.001

The use of a control group in this design improves our ability to discern what might have happened to the treatment group in the absence of the treatment. As shown in Table 2, we used the Independent T-test, and its alternative the Mann Whitney U, statistical techniques to analyze the difference in process scores in the post-implementation period. We also used this same analysis in the pre-implementation period to establish a measure of selection bias – how different are the two groups before the implementation takes place? The differences between the two hospitals in the post-implementation period are highly significant, but interestingly, the differences during the pre-implementation period are also statistically significant. This finding raises the question that the two hospitals are not ideally matched on this metric.

3.3. Examining Multiple Observations

An approach to strengthen the research design and power of the statistical analyses that can often be used in HIT evaluation involves examining multiple observations. Usually HIT implementations greatly facilitate data collection, permitting data points to be collected and compiled at frequent intervals. Examining monthly means in data is often meaningful as a way to plot changes over time.

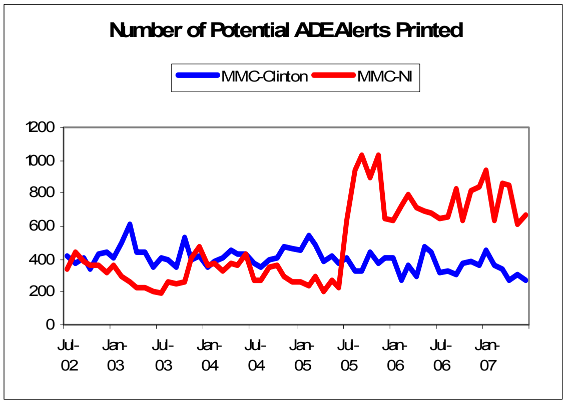

Figure 2. Mean number of potential ADE alerts printed monthly in the intervention hospital and during a comparable period in the control hospital

As can be seen in Figure 2, when we plot the monthly means of *Number of Potential ADE Alerts Printed* in our intervention hospital (MMC-NI) and our matched control hospital (MMC-Clinton) we see an obvious change at MMC-NI in July, 2005. July 8, 2005 was the “Go-Live” date for Phase Two HIT implementation for MMC-NI. The change is all the more obvious because the monthly means at MMC-Clinton show similar variability, but no dramatic change in level.

3.4. Combining Multiple Hospitals

In many research studies of HIT applications, this is as far as the research design and analyses go in terms of approaches to reduce threats to internal validity [10,17]. We had the advantage of studying HIT applications in a healthcare system. And fortunately, Trinity Health was implementing similar HIT applications throughout numerous hospitals using a unified approach [1].

One way to ameliorate the threat of selection bias in the event that hospitals, or any other unit of analysis, cannot be randomly chosen, or assigned to control or treatment groups, is to match hospitals based on a relevant set of matching variables [10]. Benchmarking against a similar set of hospitals is often a useful way to compare a particular facility with like-kind facilities. In this study, a total of 9 Trinity Health hospitals implemented Project Genesis Phase Two technologies and processes before March 2006. The data from these 9 intervention hospitals that had implemented Phase Two of Project Genesis (treatment group) were matched with 9 Trinity Health hospitals from the same Midwestern region (Michigan, Iowa, Ohio, and Indiana) that continued to use the Phase One tools and processes (control group) but had not yet implemented Phase Two. Each treatment group hospital was matched with a control group hospital on the basis of ranked bed size, using the total number of facility beds set up and staffed as the matching variable.

The same ADE Alert process measures described above were available at all treatment and control hospitals for 14 months of pre-Project Genesis Phase Two “Go-Live” data, and 14 months of post-Project Genesis Phase Two “Go-Live” data.

3.5. Longitudinal Design Using Untreated Matched Controls with Multiple Pre-test and Post-test Observations

An untreated matched control with multiple pre-test and post-test design improves upon the aforementioned designs by addressing many threats to experimental internal validity simultaneously – maturation, history, selection bias, and regression to the mean [10]. Each control group hospital’s 28 data points were paired in time with their matched treatment group hospital’s “Go-Live” date. Because all the treatment hospitals implemented Project Genesis on different dates we simply matched each control hospital’s data with its paired treatment hospital data so that the matched pair’s data spanned the same dates. We then centered each hospital’s 28 monthly data points on the treatment hospital’s “Go-Live” implementation date.

This longitudinal experimental design has an advantage over a cross-sectional or single repeated measures design in that it allows us to analyze trends in ADE process performance over time – both before and after the Project Genesis “Go-Live” month – while still allowing us to analyze performance both within and between treatment and control groups [17]. A simple analysis approach would be to collapse hospitals’ repeated monthly performance measures into one single observation per hospital per time period and use a repeated measures design. However, by collapsing repeated measures into one observation we are not able to review patterns of change over time. The strength of a longitudinal design with multiple pre-test and post-test observations is that we can get a more precise estimate of changes over time by appropriately modeling the correlation in repeated measures for each hospital [18].

3.6. Linear Mixed Model with Piecewise Linear Spline

The first approach we used to examine the longitudinal pattern in the data was a linear mixed effects model with a piecewise linear spline. The mixed effects model allows us to more appropriately model the natural heterogeneity among hospitals such as hospital-specific random effects. The mixed model entails the use of both fixed and random effects. Most importantly, in our model the random effects characteristic allows each hospital to have a unique starting point, or intercept, and a unique slope for both the pre-implementation period and post-implementation period.

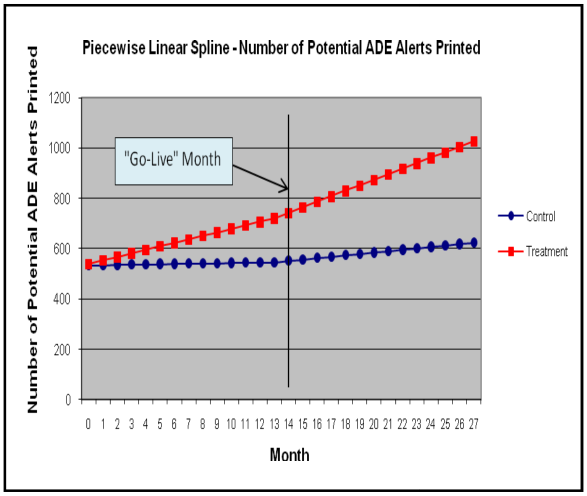

Figure 3. Monthly potential ADE alerts printed shown with a piecewise linear spline

The piecewise linear spline is another important feature to this longitudinal design as it allows an inflection point at which the trend can change direction in a non-linear graph of trends over time. Thus, each hospital group has a unique slope for both the pre-implementation and post-implementation period as shown in Figure 3. The inflection point where each groups' linear segments join, otherwise known as the common "knot", is the "Go-Live" month (Month = 14) for Project Genesis Phase Two in this model. The resultant piecewise linear curve is termed a spline [18]. The statistical result of this model is shown in Table 3.

Table 3. Potential ADE alerts printed - Linear mixed model with piecewise linear spline

	Estimate	SE	p-value
Intercept	533.37	136.44	< 0.001
Group	7.45	192.41	0.97
Pre-Month	0.94	4.30	< 0.05
Post-Month	4.64	7.98	0.26
Group*Pre-Month	12.94	5.97	< 0.05
Group*Post-Month	3.25	11.22	0.77

Comparing Figure 3 combining data across the treatment group and the control group with the pattern of changes observed at a single treatment hospital (MMC-NI) as shown in Figure 2, suggests that this model may not adequately capture the nature of the underlying data. Thus, we used a second approach which involved modeling a change in intercept at the point of "Go Live".

3.7. Linear Mixed Model with Piecewise Linear Spline and "Jump" at Knot

The implementation of an integrated package of computerized HIT resulted in a significant "jump" in the *Number of Potential ADE Alerts Printed* for the healthcare team to review. Thus, a modification to the piecewise linear spline model discussed above was added to capture this "jump" in performance at the Project Genesis "Go-Live" date. This was accomplished by introducing a Period variable – a dichotomous variable representing the pre and post implementation 14-month periods – and by recoding

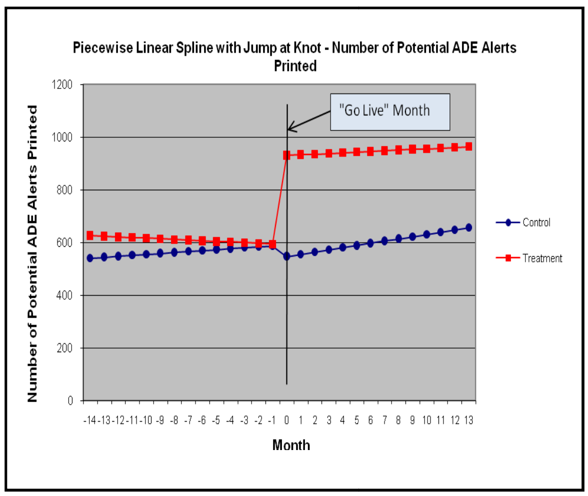

Figure 4. Monthly potential ADE alerts printed shown with a piecewise linear spline and "jump" at knot

the time (Month) variable so that Month=0 corresponded to the implementation date. Figure 4 shows the mean response profile of the hospital groups using this model. As shown in Figure 3, the piecewise linear spline model smooths out the variation in performance at "Go-Live"; however the piecewise linear spline with a jump at the knot model, as shown in Figure 4, seems to better capture the "Big-Bang" effect of the overnight HIT implementation. It also better captures the treatment group's trend after the implementation. As shown in Figure 4, the trend is actually not as steep as it appears in Figure 3, rather the implementation resulted in a substantial increase in the *Number of Potential ADE Alerts Printed* at the point of "go-Live" with a gradually increasing trend thereafter.

Table 4. Number of potential ADE alerts printed - Linear mixed model with piecewise linear spline and "jump" at knot

	Estimate	SE	p-value
Intercept	547.32	157.02	< 0.01
Period	44.46	29.97	< 0.001
Group	383.04	222.08	0.39
Period*Group	-383.09	42.30	< 0.001
Month	8.34	4.11	0.25
Period*Month	-4.74	3.82	0.07
Group*Month	-5.82	5.82	0.26
Period*Group*Month	-0.25	5.31	0.96

The statistical results of this model are shown in Table 4. In addition, the 3-way interaction of the Group, Period, and Month variables portrays changes in the dependent variable across time, periods, and between groups.

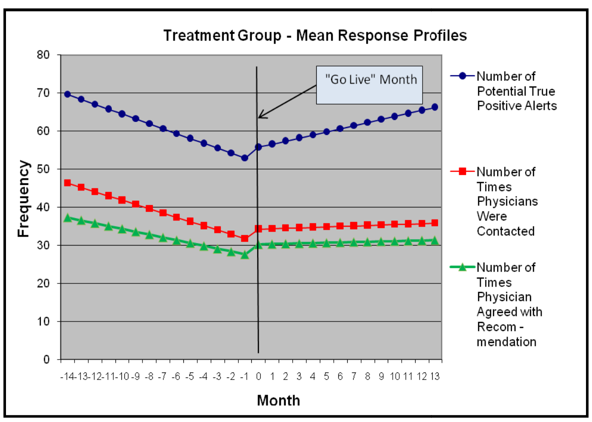

Figure 5. Piecewise linear spline and "jump" at knot showing pattern for ADE alert metrics involving interaction between pharmacist and physician

The comparison of various design and analyses approaches presented above was demonstrated using only one of the ADE alert metrics. We used this final approach to examine the pattern of monthly changes for the other ADE alert metrics.

As shown in Figure 5, three of the metrics – the *Number of Potential True Positive Alerts*, the *Number of Times Physicians Were Contacted*, and *Number of Time Physicians Agreed with Recommendations* – all showed a similar pattern over time. In contrast to the pattern observed for the *Number of Potential ADE Alerts Printed*, these metrics showed a consistent decline before implementation, a slight jump at "Go-Live", and then a level (for two metrics) or less robust increase (for one metric) after implementation.

As shown in Figure 6, a different pattern of changes over time is observed for the remaining two metrics. In contrast to the other metrics, for the *Ratio of Orders Changed to Physician Contacts* and for the *Number of Potential True Positive Alerts per 100 Admissions* there does not appear to be any "jump" at the knot.

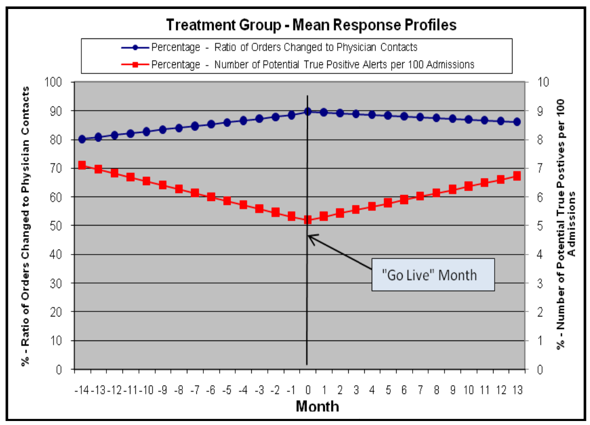

Figure 6. Piecewise linear spline and "jump" at knot showing pattern for composite ADE alert metrics

An overview of the findings observed using this type of statistical modeling suggests a pattern of changes in the process of reviewing ADE alerts at the Trinity Health hospitals that implemented Phase Two. Most obviously, the implementation of an integrated package of computerized HIT resulted in a significant "jump" in the *Number of Potential ADE Alerts Printed* for the healthcare team to review. The trends also suggest that pharmacists' workload was increasing as the slope for both the *Number of Potential ADE Alerts Printed* and the *Number of Potential True Positive ADE Alerts* identified increased at rates faster than the *Number of Times*

Physicians are Contacted. The findings suggest that post-implementation, pharmacists play an important role in the process of ADE alert review, effectively "weeding out" a very high percentage of the potential ADE alerts printed, in efforts to prevent a substantial increase in the number of times physicians are contacted for review of potentially "false positive" ADE alerts.

The results also suggest that pharmacists' engagement with the new HIT tools and medication order review process was effective. Treatment group estimates in the post-implementation period suggest that physicians agreed with pharmacists' recommendations a high percentage of the time. And, the parallel trends suggest that pharmacists' decisions to contact physicians regarding potential ADEs are reliably stable and accurate over time despite the increasing frequency of potential ADE alerts being printed and potential true positive ADE alerts identified.

Overall, the data suggest that the capture of potential and highly preventable ADEs was effectively cascaded down through the process of ADE review – through pharmacist review, pharmacist identification of potential true positive ADE alerts, physician contact, and physician agreements with pharmacists' recommendations – resulting in a significantly increased trend in the number of potential true positive ADE alerts identified per 100 admissions and a near-significant increased trend in the number of times physicians agree with pharmacists recommendations to alter a medication order.

4. Discussion

The stronger experimental designs and longitudinal data analysis techniques described here have a number of advantages. As discussed above, the use of a matched control group and multiple pre-test and post-test observations within the experimental design will bolster the internal validity of any assessment of an HIT implementation.

For example, by integrating these features into the design, the researcher may address the threats of history – was there another event that could have produced the same results over time? Revisiting Figures 2 and 4, we saw a significant "jump" of our process measure in the treatment group in a singular month that is not seen in any of the other 27 months of measurement. In addition, this "jump" is not seen in the control group indicating that it is not likely there was a historical event that caused a jump in performance across all hospitals. Without a suitable control group, the threats of statistical regression to the mean, selection bias, and/or selection-maturation interaction issues could not be addressed [10]. For example, we would not know whether the treatment group was biased by being composed of higher or lower performing hospitals, of differing ability, or of other organizational characteristics within the treatment hospitals that would cause us to incorrectly attribute changes in performance to the HIT implementation.

The use of multiple pre-test and post-test measurements increases the power of our tests, in other words, it increases our ability to identify that there is a treatment effect of some importance when the effect truly exists. This was especially important in this study, and likely many others involving resource-intensive HIT implementations that involve a limited number of units of analysis (e.g., hospitals, clinics, departments), because we only have a sample size of 9 treatment hospitals. However, experimental power is a function of both sample size and the number of equally-spaced repeated measures. Although doubling the sample size results in a greater increase in power than doubling the number of repeated measures, the power of this experiment was increased by the addition of multiple pre-test and post-test observations, essentially increasing the number of usable data points from 18 in a Paired or Independent t-test model (9 hospitals with 2 repeated measures each) to 504 when we consider that we have 2 groups of 9 hospitals with 28 monthly repeated measures for each hospital.

The longitudinal approach utilized in this study has some unique advantages. With repeated measures techniques we can assess within-hospital differences, whereas in cross-sectional techniques we can only capture between-hospital differences. The *raison d'etre* of a longitudinal study is to characterize within-individual changes in a response variable over time [17]. The linear mixed effects piecewise spline model as built for this study captures both within and between-hospital changes over time.

Another appealing aspect of linear mixed effects modeling is its flexibility in handling imbalance in the number of responses [17,19]. Hospitals that miss a measure in a given month have a different number of monthly measures compared to a hospital that reports all 28 monthly data points. A missing data point within a data set necessarily implies an imbalanced data set. Most longitudinal health sciences studies are missing data on at least one measurement, and this study was no exception to this general rule, however many of today's statistical packages can handle missing data [17,19].

Additionally, researchers should be aware that a linear mean model may not adequately represent the underlying trend in the data. It may be that a curvilinear model is more appropriate in situations in which the means show a monotonically increasing or decreasing trend. A model utilizing higher order polynomials (e.g. – quadratic or “squared” terms) may be tested to determine their “fit” in contrast to linear models. In this study we used a piecewise linear spline approach which is a simple way to model a non-linear trend that cannot be handled parsimoniously by simple polynomials [17,20,21].

Finally, the robustness of today’s statistical software has its advantages in empowering relatively novice empirical researchers, but as Judith D. Singer warns “statistical software does not a statistician make” [22]. There are model-building decisions that should not be ignored by researchers when using linear mixed effects models in longitudinal designs. We provide a short description of the types of issues confronted here and provide resources in the literature that could be used to guide model building efforts.

5. Conclusion

The evaluation of HIT systems involves complex challenges. The “gold-standard” for research design, randomized controlled trials, can rarely be conducted in HIT because of cost constraints. Thus, alternative, observational designs must be used and strengthened as much as possible to assure the highest value for the evaluation. An obvious asset to exploit in evaluations of HIT systems is the wealth of data produced by most clinical applications that can be used to enable stronger analyses of system performance. As demonstrated in this paper, appropriate quantitative statistical tools can enhance the analysis and interpretation of system models. Borrowing from engineering, analytical approaches can integrate statistical process control run charts which examine fluctuations in process over time and help identify if those fluctuations are due to random events or a systematic change. Human factors approaches can be used to consider the effect of an innovation implementation upon the human interactions within the system. Understanding the effects of HIT implementation, including both intended and unintended process and outcome changes, will benefit from adoption of the many research design, methodological, and analysis tools available.

Acknowledgments

The research team acknowledges the support of the Agency for Healthcare Research and Quality through Grant #HS015196, Trinity Health of Novi, Michigan, The University of Iowa, and the University of Missouri Center for Health Care Quality.

References

[1] Crandall DK, Brokel JM, Schwichtenberg T, et al. Redesigning Care Delivery through Health IT Implementation: Exploring Trinity's IT Model. *The Journal of Healthcare Information Management.* Fall 2007;21(4):41-48.

[2] Leape LL, Brennan TA, Laird N, et al. The Nature of Adverse Events in Hospitalized Patients. Results of the Harvard Medical Practice Study II. *New England Journal of Medicine.* February 7, 1991;324(6):377-384.

[3] Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and Types of Adverse Events and Negligent Care in Utah and Colorado. *Medical Care.* March 2000;38(3):261-271.

[4] Aspden P, Wolcott JA, Bootman JL, Cronenwett LR. *(Eds.), Preventing Medication Errors: Quality Chasm Series*. Washington D.C.: The National Academies Press; 2007.

[5] Bates DW, Boyle DL, Vander Vliet MB, Schneider J, Leape LL. Relationship between Medication Errors and Adverse Drug Events *Journal of General Internal Medicine.* April, 1995;10(4):199-205.

[6] Kaushal R, Bates DW, Landrigan C, et al. Medication Errors and Adverse Drug Events in Pediatric Inpatients. *JAMA.* April 25, 2001;285(16):2114-2120.

[7] Chaudhry B, Jerome W, Shinyi W, et al. Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine.* 2006;144(10):742-752.

[8] Kaushal R, Shojania KG, Bates DW. Effects of Computerized Physician Order Entry and Clinical Decision Support Systems on Medication Safety: A Systematic Review. *Archives of Internal Medicine.* June 23, 2003;163(12):1409-1416.

[9] Shekelle PG, Morton SC, Keeler EB. Costs and Benefits of Health Information Technology. Evidence Report/Technology Assessment No. 132. (Prepared by the Southern California Evidence-based Practice Center under Contract No. 290-02-0003.) AHRQ Publication No. 06-E006. Rockville, MD. *Agency for Healthcare Research and Quality.* April 2006.

[10] Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs*. Boston, MA: Houghton Mifflin Company; 2002.

[11] Leape LL, Berwick DM, Bates DW. What Practices Will Most Improve Safety?: Evidence-Based Medicine Meets Patient Safety. *JAMA.* July 24, 2002;288(4):501-507.

[12] Shojania KG, Duncan BW, McDonald KM, Wachter RM. Safe but Sound: Patient Safety Meets Evidence-Based Medicine. *JAMA.* July 24, 2002;288(4):508-513.

[13] Ash J, Gorman P, Hersh W. Physician Order Entry in U.S. Hospitals. *Proceedings American Medical Informatics Association Annual Symposium.* 1998:235-239.

[14] Bates DW, Teich JM, Lee J, et al. The Impact of Computerized Physician Order Entry on Medication Error Prevention. *J Am Med Inform Assoc.* July 1, 1999;6(4):313-321.

[15] Massaro T. Introducing Physician Order Entry at a Major Academic Medical Center, II. Impact on Medical Education. *Acad Med.* January, 1993;68(1):25-30.

[16] Reid PP, Compton DW, Grossman JH, Fanjiang G. *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington D.C.: The National Academies Press; 2005.

[17] Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, New Jersey: John Wiley & Sons; 2004.

[18] Isaac S, Michael WB. *Handbook in Research and Evaluation*. 3rd ed. San Diego, CA: Educational and Industrial Testing Services; 1997.

[19] Cnaan A, Laird NM, Slasor P. Using the General Linear Mixed Model to Analyze Unbalanced Repeated Measures and Longitudinal Data. *Statistics in Medicine.* 1997;16(20):2349-2380.

[20] Littell RC, Henry PR, Ammerman CB. Statistical Analysis of Repeated Measures Data Using SAS Procedures. *Journal of Animal Science.* April 1, 1998;76(4):1216-1231.

[21] Littell RC, Pendergast J, Natarajan R. Modeling Covariance Structure in the Analysis of Repeated Measures Data. *Statistics in Medicine.* 2000;19(13):1793-1819

[22] Singer JD. Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics.* Winter 1998;23(4):323-355.