Modeling the Diffusion of Sentinel Lymph Node Biopsy in Breast Cancer Treatment

John Bainbridge

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics.

Chapel Hill
2016

Approved by:

Pranab Sen

Chirayath Suchindran

Lloyd Edwards

Shelley Hwang

Kandace McGuire

John Preisser Jr.

ABSTRACT

John Bainbridge: Modeling the Diffusion of Sentinel Lymph Node Biopsy in Breast Cancer Treatment.
(Under the direction of Pranab Sen and Chirayath Suchindran)

Use of a generalized linear mixed model with a binary outcome and logit link function is proposed to generate trajectories of the probability of use of novel medical procedures. It is hypothesized that the shape of these innovation adoption trajectories vary by institution and region and are influenced by patient, institutional, and geographic factors. The example of the adoption of sentinel lymph node biopsy in the treatment of early stage breast cancer is used to demonstrate the model's utility and improvement over those typically used in registry and claims-based research. Surveillance, Epidemiology, and End Results (SEER)-Medicare data from 1999 to 2007 was used as the basis for these model-based trajectories. Fixed effects included patient, institution, and regional variables including a cubic polynomial of time for each region. Random effects were at the institution level and included a cubic polynomial of time. Results indicated a better fit of the multilevel model with a polynomial of time in comparison to standard models and that patient, institutional, and geographic factors influence the shape of the adoption trajectory of this novel medical procedure.

Additionally, an evidence-based medical implementation index (EMII) was developed and tested using sentinel node biopsy adoption trends. Data were analyzed in aggregate and at the institution level. A single summary metric, based upon the area under the curve, was derived to quantify the pattern of adoption ranging from 0-100, with higher scores reflecting earlier adoption. The EMII was compared between SEER regions and between institutions. Differences in adoption patterns were found for SEER regions and institutions ($p < .001$ for each effect). For SEER regions (n=15) the SLNB EMII range was 33 (New Mexico) to 66 (Seattle). For all institutions: n = 720, range = 4 - 87, mean = 46, S.D. = 20, bell-shaped distribution.

Finally, four estimation techniques for the random effects parameters were compared to maximum likelihood using quadrature based estimates, two types of pseudo-likelihood (PL), and jackknifed estimates based on these. The estimates were compared via D-, A-, and E-efficiency. Results indicated that even with jackknifing, PL estimates of the variance and thier confidence intervals were biased.

ACKNOWLEDGMENTS

Thanks to all who helped me through this process including my committee members, for their useful feedback, patience, and insights and my family and friends for their support and encouragement. I'd also like to thank my places of employment during this process that either directly or indirectly encouraged my completion of the DrPH program. At the Cecil G. Shep's Center for Health Services Research I was given the opportunity to learn about SEER-Medicare data and leverage that opportunity in this work. At the Duke Human Vacccine Instititue my colleagues have been supportive and encouraging the completion of this work.

This work used the linked SEER-Medicare database. The interpretation and reporting of these data are the sole responsibility of the author. The author acknowledges the efforts of the Applied Research Program, NCI; the Office of Research, Development and Information, CMS; Information Management Services (IMS), Inc.; and the Surveillance, Epidemiology, and End Results (SEER) Program tumor registries in the creation of the SEER-Medicare database. The collection of the California cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; NCI contracts N01-PC 35136, N01-PC-35139, and N02-PC-15105; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement U55/CCR921930-02.

This work was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, through Grant Award Number 1UL1TR001111. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 LITERATURE REVIEW

1.1 Introduction

1.1.1 Measuring Quality of Cancer Care

Quality of care in the treatment of cancer is an important issue and awareness has been heightened ever since the Institute of Medicine released its recommendations in **Ensuring Quality Cancer Care** (39). One of the IOM's recommendations of primary importance to this work is to "Measure and monitor the quality of care using a core set of quality measures." Other recommendations included; "Services for the un- and underinsured should be enhanced to ensure entry to, and equitable treatment with, the cancer care system", "Studies are needed to find out why specific segments of the population ... do not receive appropriate cancer care", and "Cancer care quality measures should be used to hold providers ... accountable for demonstrating that they provide and improve quality of care." While the first and last call explicitly for a set of metrics, the other two implicitly require metrics in order to be achieved. The focus of this work is the research used to generate the metrics that are used to measure, and ultimately improve, quality of medical care for breast cancer patients.

One indication that quality care is being provided is the use of new evidence-based procedures. One such procedure used in early stage breast cancer lymph node staging is Sentinel Lymph Node Biopsy (SLNB), originally called sentinel lymphadenectomy (20). It is being used in place of Axillary Lymph Node Dissection (ALND) in some cases when appropriate (20; 8; 15; 19), although seemingly with disparities for at risk groups (22; 35; 11; 45; 9; 44). In section 1.1.2 we describe the procedures.

1.1.2 Lymph Node Staging Procedures for Breast Cancer Treatment

Axillary Lymph Node Dissection (ALND)

ALND involves the removal of all level one and two ipsilateral axillary lymph nodes. It is an extensive procedure and associated morbidities include, lymphedema, nerve paresthesias, axillary seromas, and infections. Its primary purpose is to provide pathological nodal staging information. It also serves the function of removing any metastatic tissue that is involved with the axillary lymph nodes.

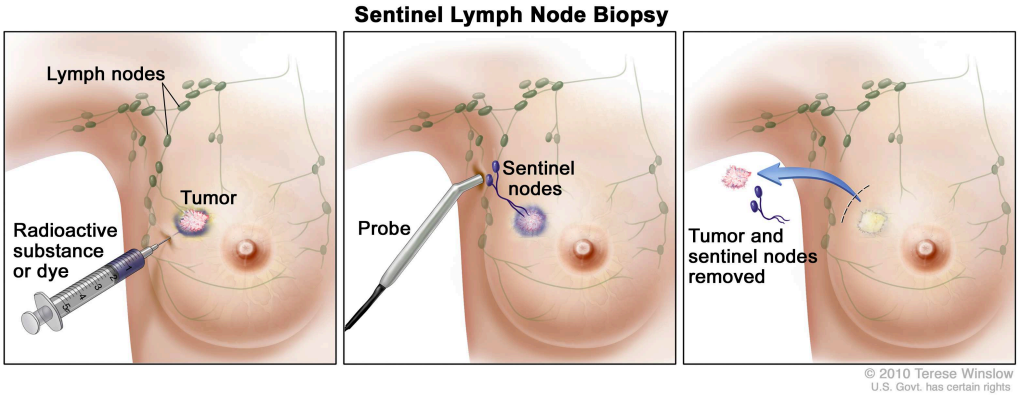

Figure 1.1: Sentinel lymph node biopsy procedure.

Sentinel Lymph Node Biopsy (SLNB)

SLNB is a less invasive procedure than ALND, where a few lymph nodes are removed, typically 1 to 3, and interoperatively assessed for metastatic involvement (8; 24). A standard version of the procedure during the time period of this study started with injecting a blue dye and a radioactive tracer near the site of the tumor (36). After some time has passed both visual inspection and a Geiger counter are used to determine which lymph node(s) the lymph near the tumor drains to. These are excised and assessed for metastatic involvement. If it is determined that there is involvement then the next node in the lymph chain is excised and examined. Using this information it can then be determined whether to perform a completion ALND.

SLNB as Standard of Care for Early Stage Breast Cancer Treatment

Up through the late 1990's the standard practice had been to perform ALND for lymph node staging in cases of early stage breast cancer. With the advent of SLNB in the mid 1990's and a growing body of evidence that it was equivalent to ALND in outcomes for pathologically node negative cases it became the standard of care (62; 13; 52; 14; 41; 54; 55; 61; 56). Medicare started reimbursing its use in 1999 while it was still in clinical trials since early evidence suggested it was a preferred treatment option in many cases. With the conclusion of the clinical trials SLNB became the documented standard of care for lymphatic staging of early stage breast cancer (26).

In the following section the data source for modeling the use of SLNB, SEER-Medicare cancer registry and claims data, is described and reviewed.

1.1.3 SEER-Medicare: Combined Cancer Registry and Medicare Claims data

SEER-Medicare data is a rich data source for treatment and outcomes of cancer in the United States but is rather complex. It is a joining of cancer registry data and the associated Medicare data for individuals that have Medicare coverage. It has been described extensively elsewhere, particularly in a supplement to Medical Care in 2002 (59; 4; 25; 48). What follows in a basic description of each of the components starting with the cancer registry data and how it's collected.

SEER Cancer Registry Program

According to information provided during the National Cancer Institute's SEER-Medicare data training workshop in March 2010 (60), the Surveillance, Epidemiology, and End Results (SEER) program contracts with state health programs and universities to operate incident cancer registries in geographic regions around the United States of America. Currently the registries cover about 26% of the US population. The 17 registries are each headed by a different investigator and each registry works differently. They do however use a standardized reporting system for data transmission to NCI.

It should be noted that the program is for the most part facility based and thus information for services provided in a physician's office might be missing or incomplete. It should also be noted

that for many variables there is a hierarchical structure, that is, some values supersede other values even though both are true e.g. the maximum value is reported. The types of data included are, diagnostic, staging, treatment, and limited demographics.

Registry data can come from a variety of sources but a case is first identified via a report from a service provider where a cancer diagnosis was made, an autopsy report, or from a death certificate submitted to the cancer registry for the geographical region where the entity is located. In SEER-Medicaid research it is a standard practice to exclude those cases that were only identified via autopsy or death certificate. Thus the reporting sources include at least one of; a Hospital Inpatient/Outpatient or Clinic, a Laboratory (Hospital or Private), a Physician's Office/Private Medical Practitioner, or a Nursing/Convalescent Home/Hospice.

There are a variety of missing data issues for SEER data. Registries don't capture 100% of all incident cases, although the expectation is at least 98%. SEER missing data is more likely to have come from patients who were primarily treated at physician's offices (rather than at a hospital or clinic). Pathological staging data 'overwrites' clinical staging data and we don't know if this has happened or not for breast cancer (prostate cancer has separate clinical staging variables). Only 94% of the cases aged 65 or older link to Medicare data and it is unknown if this is differential, but assumed not. Unlinked cases are not included in SEER-Medicare data. Changes in coding systems over time have led to some staging and treatment information being collected during some time frames but not others (or defined somewhat differently). Race data categories depend on time period, SEER race data however is considered better than Medicare race data. Missing/unknown categories for some SEER variables during analysis behave more like a separate level (or are similar to another level) rather than seeming to come from the other levels proportionally. Six months of 2005 data for Louisiana are somewhat sparse due to severe storms and an indicator variable flags this time period. There were four new SEER registries in 2000, there is no prior data for them.

Next we consider the Medicare data that is provided by NCI in their release of SEER-Medicare data.

Medicare data as Provided by the SEER-Medicare Program

Medicare data used in studies of SLNB consists of eligibility, inpatient stay, outpatient service, physician service, and durable medical equipment (which includes some cancer medications) information. Eligible individuals include qualifying individuals age 65 or greater, those with end stage renal disease, and certain people with disabilities. Note that this is administrative (billing) data for services that are paid for based on Medicare policy in effect at the time of service. This has implications for completeness, particularly for procedures not covered at the time of service. The pertinent information is obtained from four files: the hospital inpatient stay file (MEDPAR), which has one record per hospital stay, the hospital outpatient claims file (OUTSAF), which has one record per billable item, the non-hospital provider claims file (NCH), which has one record per billable item, and the durable medical equipment file (DME), which includes cancer medications.

Medicare data also has its share of missing data issues. Census data sometimes is missing, and the usual strategy is to use zip code based data in this case. There are still a (very small) number of missing after this strategy is employed although. Physicians tend to be erratic in billing for services they know Medicare won't pay for, so for uncovered services there will be 'missing' claims data. Similarly some services are 'bundled' with others and only the code for the more extensive procedure is provided. One needs to be familiar with Medicare policy during time period of interest and during some years the location of interest in order to finesse the complexities. Medicare was supposed to have the exact same coverage for services throughout the US but during the early 2000's some of the fiscal intermediaries had slightly different policies or effective dates for service coverage. E.g. different regions (primarily) of the country had different start dates for coverage of SLNB depending on which fiscal intermediary (FI) processed a hospital's claims. An additional issue is that, according to a Medicare data training workshop provided by RESDAC, a person always has the same FI processing their claims as when they first signed up. Thus, individuals that moved during the time period where there were differences in FI policies may have had the policies imposed based upon their previous residence's FI.

Concluding this section is a description of the SEER-Medicare data methodology.

SEER-Medicare data Methodology

SEER-Medicare methodology is as follows. From the SEER website, "Geographic areas were selected for inclusion in the SEER Program based on their ability to operate and maintain a high quality population-based cancer reporting system and for their epidemiologically significant population subgroups. The population covered by SEER is comparable to the general US population with regard to measures of poverty and education. The SEER population tends to be somewhat more urban and has a higher proportion of foreign-born persons than the general US population." Thus, it is not a probability sample and all generalizations to populations outside of the SEER regions must be model based. This data is then linked to Medicare eligibility data and a 94% linkage rate has been found for cases where the first cancer occurred at an age of 65 or greater. Cases where there is a successful link are kept.

The data file creation process starts with SEER data being transformed from the format of one record per incidence of primary cancer to a person level format. The data for the first ten cancers is retained, i.e. it is changed from tall to wide structure with truncating any incidences of cancer after the first ten. Variables are created for each of the (up to) ten cancers including all the different coding systems over time. This leads to many empty data elements and requires knowledge of which set of variables are to be used for any given incidence of cancer. This is determined by cancer number, date of diagnosis, and for a few variables cancer type.

This information is combined with demographic and eligibility information from the Medicare program to create the Patient Entitlement and Diagnosis Summary File (PEDSF) file. The name is a little misleading as it includes all of the (selected by NCI) SEER data, not just the diagnosis information. It also includes (de-identified and 'fuzzed') census tract and zip code level demographic (census) information.

In this next section Rogers theory of the diffusion of innovation is reviewed. It is a theory that has informed and framed much of the work done in modeling the use of SLNB.

1.1.4 Rogers Theory of the Diffusion of Innovation

Before looking at individual studies on the use of SLNB, it is useful to consider Rogers' (1962) theory on the diffusion of innovation. It explicitly is part of the conceptual basis of at least two papers on the use of SLNB (9; 44) and will be used in the development of the model in this work. He proposed that the uptake by a population of an innovation, be it a product, an idea, or a procedure, follows a consistent pattern. This pattern can be described by a S-shaped curve on a graph, with the proportion of a population that has adopted the innovation on the y-axis and time on the x-axis. It has aspects of a cumulative distribution function and is similar in shape to the logistic function, with the exception that it starts at zero and can eventually reach 100%. Thus the model should permit examining whether it is the case with the diffusion of SLNB.

The taxonomy for Rogers' five categories of adopter types is Innovators, Opinion Leaders, Early Majority, Late Majority, and Laggards. He suggests that the proportion of a population that falls into each category is, respectively, 2.5%, 13.5%, 34%, 34%, and 16% and it's around the point when 16% of the population has adopted the innovation that the function begins its steepest slope. I hypothesize that at the hospital level that one will be able to identify these different adopter types and thus the model should permit testing this as well. In the case of assessing the rate of use of a procedure over time, one way to identify the different categories is by graphical examination of individual hospitals' rate of use of over time. Numerically, a formula could be devised to categorize hospitals based upon their subject specific intercept terms and the subject specific polynomial if the effect of time is included in a model.

A limitation of Rogers model is that it only speaks to the time period when adoption is occurring and thus is a non-decreasing function over time. It also is limited to the choice of adopt or not adopt, so differential adoption isn't explicitly addressed. One could however easily imagine separate curves for different situations i.e. the curve is conditional upon some set of factors. For modeling the use of a procedure over a longer timespan it would be useful to consider the possibility of decreasing use, particularly for subpopulations. Possible reasons for decreasing use among a

(sub)population are, the procedure is found to have issues, it is supplanted by a newer technology, there is a substantial increase in cost to payer relative to alternatives (or benefit), and a change in incentives to the decision maker that reward different choices. Thus we need to extend Rogers' theory to model the lifecycle of a procedure where there is a decreasing component at some point. A good model will permit testing for these possibilities. In the case of a model that uses a global polynomial to measure the effect of time, it essentially requires the use of a cubic term.

In section 1.2, seven papers that examined the use of SLNB in early stage breast cancer are reviewed, starting with their methodologies, variables, and results in subsection 1.2.1 and then with a greater focus on the statistical models employed in subsection 1.2.2.

1.2 Previous Research in the use of SLNB in Early Stage Breast Cancer Treatment

With Rogers' theory in mind let's examine some of the work to date on the use of SLNB in breast cancer treatment.

1.2.1 Methodology, Variables and Results

Maggard et al. (2005) (35) used SEER data from 12 regions from 1998 to 2000, treating patient as the unit of analysis, and performed multivariable logistic regression to identify predictors of the use of SLNB during that time frame. Included were those cases that had an AJCC stage of I or II. Those cases with "Histologies corresponding to squamous cell, spindle cell, carcinoid, sarcoma, Paget's disease, and in situ tumors were excluded from the analysis."(35) They must have received definitive surgery, Lumpectomy or Mastectomy, and received SLNB, ALND, or both.

Variables included in the model were all categorical (reference levels are underlined) and included; SEER Registry, Tumor grade, Tumor stage (I, II), Age at diagnosis (<u>≤40</u>, 40-49, 50-59, 60-69, 70-79, and 80+), Race/ethnicity (<u>White</u>, Black, Hispanic, Asian, and Other), Marital status (Not married vs. <u>Married</u>), Year of diagnosis (1998, 1999, <u>2000</u>), and Surgery type (<u>Lumpectomy</u> vs. Mastectomy). While stating that they controlled for registry and tumor grade they did not give any information about how these variables were handled in the model. They found that older women and minority groups, as well as those receiving mastectomies were less likely to have a

SLNB performed then their respective reference groups.

Chen et al. (2008) (11) used National Cancer Database data from 1998 to 2005, treating patient as the unit of analysis, and performed several multivariable logistic regressions to examine the effect of patient, clinical, facility, and neighborhood characteristics, as well as year, on the receipt of SLNB. Inclusion criteria were; TNM staging of T1N0M0 or T2N0M0, received definitive surgery, received nodal staging, and no missing demographic data. They found all of the variables included in their models to be highly statistically significant with the exception of a few levels (mostly the 'missing' level) of a small number of variables. No attempt was made to do multi-level modeling despite the clearly multi-level nature of the data.

It should be noted that Urbach and Austin's (2005) (53) paper 'Conventional models overestimate the statistical significance of volume-outcome associations, compared with multilevel models', in the Journal of Clinical Epidemiology, points out just that, along with providing an example analysis using hospital procedure volume to predict an outcome. This of course would generalize to any higher level covariate in a multi-level structure, but is particularly of note since several of the SLNB papers use hospital procedure volume as a predictor variable including Chen et al.

Rescigno, Zampell, and Axelrod (2009) (45) used SEER data from 14 regions from 1998 to 2004, treating patient as the unit of analysis, and performed multivariable logistic regression to examine factors involved in the nodal staging procedures used. Inclusion criteria were; T1 to T3, N0 and N1, and M0 based on TNM staging, received definitive surgery, and status of nodal staging known. Variables included in the model were at the disease, patient, and neighborhood levels but no hospital level information was included. They found both appropriate and inappropriate use of SLNB and ALND, with significant effects of disease factors, age, Hispanic (but not African American) ethnicity/race, and neighborhood level demographics.

Carpenter et al. (2011) (9) used SEER-Medicare data from seven regions from 2000 to 2002 and employed a three level multilevel model where patient was nested within hospital, which was nested in SEER region, to examine the factors involved in the diffusion of the use of SLNB during

this time frame. The model was implented by use of a Generalized Linear Mixed Model with random intercept terms for hospital nested within SEER region and SEER region. Inclusion criteria were; 66 or older at the date of first diagnosis of primary breast cancer, 12 months Medicare parts A and B eligibility prior to diagnosis, 24 months Medicare parts A and B eligibility post diagnosis (or until death), no HMO coverage during the study period, were not identified via autopsy nor death certificate, received definitive surgery, and received nodal staging. Variables included in the model were at the disease, patient, neighborhood, hospital, and regional levels. They found that hospital level variables, other than year, had the largest effect, African Americans and older patients were much less likely to receive SLNB, and many variables previously reported to be significant were not found to be statistically significant.

Reeder-Hayes et al. (2011) (44) used the same data as Carpenter et al. but performed multivariable logistic regression with GEE estimation of within hospital correlations. Variables included in the model were similar as Carpenter et al. but included a variable for receipt of Medicaid, which was found to have a rather large (OR 0.61, C.I. 0.47,0.78) and significant effect.

Meyer et al. (2013) (38) used SEER-Medicare data from 2000 to 2005 but it is unclear what their model actually was. In the abstract they say they used a generalized linear model with generalized estimating equations while in the text they say they used SAS's PROC GLIMMIX which 'uses random effects and takes into account the clustering of patients within physicians and physicians within hospitals.' They also indicate that they used Maximum Likelihood with the Laplace method for numeric integration. Given the lack of a clearly stated model we are unable to assess it further.

Arrington et al. (2013) (3) used SEER data from 1998 to 2008 looking across all ages with a focus on urban vs. rural populations. They used multivariate logistic analysis but did not control for the hierarchical nature of the data.

Next we focus more in depth on the statistical models that the five papers with known models used for their analyses.

1.2.2 Statistical models

In this section we will look at the statistical models used in each of the papers, including the model assumptions. The three model types used are all variants of logistic regression. They are binomial logistic regression, a marginal model with generalized estimating equation (GEE) estimates of parameters including within cluster correlation, and a generalized linear mixed model (GLMM). To begin we consider the papers that used binomial logistic regression.

Binomial logistic regression - A type of generalized linear model

Four of the papers considered used binomial logistic regression for modeling the receipt of SLNB. There is a substantial body of literature on logistic regression but a standard reference is McCullagh and Nelder (1989) (37), whose notation and terminology is used in this section. For describing logistic regression McCullagh and Nelder use the framework of generalized linear models (GLM) originally developed by Nelder and Wedderburn (1972) (40). The main components of a GLM include the error distribution, a link function, and the systematic component. The variance function and the dispersion parameter stem from the choice of main components. Thus for the case, such as in the first three papers, where there are only categorical explanatory variables, binomial logistic regression can be described as follows:

Model description

Let N be the total number of patients observed in a study. $\mathbf{x}$ is a vector of length p whose elements consist of categorical explanatory variables. A covariate class, such as described in McCullagh and Nelder (1989) (37), is a distinct combination of covariate levels for the explanatory variables. There are n covariate classes and i is the index for the covariate classes such that $i = 1$ to n. Thus, a covariate class is all observations that have the covariate vector $(x_{i1}, x_{i2}, \ldots, x_{ip})$. m_i is the number of observations in the i^{th} covariate class.

Now, let $\mathbf{Y}$ be the response vector of length n consisting of counts of receipt of SLNB for each covariate class. $\boldsymbol{\pi}$ is a vector of length n consisting of the probabilities of SLNB for each covariate class. $\boldsymbol{\eta}$ is a vector of length n consisting of the log odds of SLNB for each covariate class. $\boldsymbol{\beta}$ is

the parameter value vector of length p. $\mathbf{X}$ is the design matrix of size $p \times n$. $\mathbf{m}$ is a vector of length n consisting of the covariate class counts.

Then using the GLM framework and the case of binomial logistic regression with categorical explanatory variables; we have error distributions that are binomial and independent, use of the logit for the link function, and a systematic component that is linear on the logit scale. These main components of the GLM as well as the corresponding variance function and dispersion parameter, $\boldsymbol{\phi}$ can be written for each covariate class as follows:

$$\text{Error distribution} \qquad Y_i \sim \text{Bin}(m_i, \pi_i), \text{ where } Y_j \perp Y_k \text{ and } j \neq k$$

$$\text{Logit link function} \qquad \eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right) \qquad \text{Systematic component} \qquad \eta_i = \mathbf{x_i^T}\boldsymbol{\beta}$$

$$\text{Variance function} \qquad \pi_i(1-\pi_i) \qquad \text{Dispersion parameter} \qquad \phi_i = \left(\frac{1}{m_i}\right)$$

The primary model assumption is that conditional upon the explanatory variables the binary outcomes are generated by a bernoulli process with each event independent and identically distributed. Another important assumption is that the model is correctly specified and all necessary independent variables are included in the model. Parameters can be estimated with maximum likelihood estimation although the EM algorithm is needed if there are missing values for any of the explanatory variables.

Parameter Estimation

Parameter estimation can be accomplished via either the Iteratively Reweighted Least Squares Algorithm, also known as Fisher Scoring, or by use of the Newton-Raphson Algorithm. Proc Logistic in SAS defaults to Fisher Scoring and it is described in the documentation as well as a more complete description in McCullagh and Nelder (1989) (37). Following is a description of the process using Fisher Scoring drawing heavily on McCullagh and Nelder (1989) section 4.4.

To begin note that the log likelihood for the binomial distribution is:

$$l\left(\boldsymbol{\pi};\mathbf{y}\right)=\sum_{i=1}^{n}\left[y_i\log\left(\frac{\pi_i}{1-\pi_i}\right)+m_i\log\left(1-\pi_i\right)\right]\text{ (McCullagh and Nelder's 4.11).}$$

and that it's derivatives with respect to π_i and β_r are:

$$\frac{\partial l}{\partial\pi_i}=\frac{y_i-m_i\pi_i}{\pi_i\left(1-\pi_i\right)}\qquad\qquad\frac{\partial l}{\partial\beta_r}=\sum_{i=1}^{n}\frac{y_i-m_i\pi_i}{\pi_i\left(1-\pi_i\right)}\frac{\partial\pi_i}{\partial\beta_r}$$

noting that:

$$\frac{\partial\pi_i}{\partial\beta_r}=\frac{d\pi_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_r}=\frac{d\pi_i}{d\eta_i}x_{ir}\qquad\text{and}\qquad\frac{d\pi_i}{d\eta_i}=\pi_i\left(1-\pi_i\right)$$

Then McCullagh and Nelder's equation 4.14:

$$\frac{\partial l}{\partial\beta_r}=\sum_{i=1}^{n}\frac{y_i-m_i\pi_i}{\pi_i\left(1-\pi_i\right)}\frac{d\pi_i}{d\eta_i}x_{ir}$$

in matrix notation becomes:

$$\frac{\partial l}{\partial\boldsymbol{\beta}}=\mathbf{X^T}\left(\mathbf{Y}-\boldsymbol{\mu}\right)$$

Fisher's information for $\boldsymbol{\beta}$ is given by McCullagh and Nelder's equation 4.15:

$$-E\left(\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right)=\sum_i\frac{m_i}{\pi_i\left(1-\pi_i\right)}\frac{\partial\pi_i}{\partial\beta_r}\frac{\partial\pi_i}{\partial\beta_s}=\sum_i m_i\frac{\left(d\pi_i/d\eta_i\right)^2}{\pi_i\left(1-\pi_i\right)}x_{ir}x_{is}=\left\{\mathbf{X^T W X}\right\}_{rs}$$

Where $\mathbf{W}$ (in reduced form) is a diagonal weight matrix:

$$\mathbf{W}=\text{diag}\left\{m_i\pi_i\left(1-\pi_i\right)\right\}$$

Let $\mathbf{Z}$, the adjusted dependent variable, have components:

$$z_i = \hat{\eta}_i + \frac{y_i - m_i\hat{\pi}_i}{m_i}\frac{d\eta_i}{d\pi_i}$$

McCullagh and Nelder's 4.16 is the equation that maximum likelihood estimates satisfy:

$$\mathbf{X^T W X}\hat{\boldsymbol{\beta}} = \mathbf{X^T W Z}$$

Parameter estimates can be obtained by starting with a value for $\hat{\boldsymbol{\beta}}$, say $\hat{\boldsymbol{\beta}}^{(\mathbf{0})}$, based upon initial estimates for $\hat{\boldsymbol{\pi}}^{(\mathbf{0})}$ using the data (with an appropriate adjustment for the cases where $\hat{\pi}_i^{(0)} = 0$ and $\hat{\pi}_i^{(0)} = 1$) and then solved iteratively using least-square methods.

The revised estimate is:

$$\hat{\boldsymbol{\beta}}^{(m+1)} = \left(\mathbf{X^T W X}\right)^{-1}\mathbf{X^T W Z}$$

With values on the right based upon the estimates from the previous iteration. The process is continued until convergence is obtained.

In the next section we look a paper that used a marginal model in an attempt to control for the within cluster correlation.

Logistic regression - Marginal model

Reeder-Hayes et al. (44) used a marginal (population averaged) model and GEE based estimates of the parameters with an unstructured correlation matrix. This model, unlike the previous, takes into account the fact that observations from the same hospital are likely to be correlated by treating the hospital as a cluster and modeling the within cluster correlation. It is called a marginal model to emphasize the fact that only the explanatory variables are used in generating the explanatory parameter estimates. The specifics of the model are as follows:

Model description

Let N be the total number of patients observed. There are n hospitals (clusters) and i is the index for the hospitals so that $i = 1$ to n. j indexes the patients within a hospital and there are m_i patients in the i^{th} hospital, so $j = 1$ to m_i. The response vector $\mathbf{Y_i} = (Y_{i1}, \ldots, Y_{im_i})^T$ consists of zeros and ones that indicate each patients (non)receipt of SLNB. The components of the model are as follows:

$$\text{Error distribution} \qquad Y_{ij} \sim \text{Bern}(\pi_{ij})$$

$$\text{Logit link function} \qquad \eta_{ij} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) \qquad \text{Systematic component} \qquad \eta_{ij} = \mathbf{x_{ij}^T}\boldsymbol{\beta}$$

$$\text{Variance function} \qquad \pi_{ij}(1-\pi_{ij}) \qquad \text{Correlation structure} \qquad \text{Corr}\,(Y_{ij}, Y_{ik}) = \alpha_{jk}$$

Note that in this type of model the expected value and the variance of Y_{ij} are modeled separately. Also note that since only the first two moments are given a likelihood function can not be specified. This leads to the use of quasi-likelihood estimates for the parameters. The model assumptions include that any missing outcome data is missing completely at random (MCAR) (46; 31).

Following is a description of the estimation process, it is based heavily on course notes from Professor Herring's longitudinal data analysis class which for this topic are based on Liang and Zeger (1986) (31).

Parameter Estimation

The working covariance matrix is defined as $\mathbf{V_i} = \phi_i \mathbf{A_i^5 R}\,(\alpha)\,\mathbf{A_i^5}$ where $\mathbf{A_i}$ is a diagonal matrix with the values of $v\,(\mu_{ij})$ on the diagonal and $\mathbf{R}\,(\alpha)$ is the working correlation matrix indexed by α.

Estimation of $\boldsymbol{\beta}$ is accomplished by the use of generalized estimating equations where it is the solution to:

$$\sum_{i=1}^{n} \mathbf{D_i^T V_i^{-1}}\,(\mathbf{Y_i} - \boldsymbol{\mu_i}) = \mathbf{0} \qquad \text{where} \qquad \mathbf{D_i} = \frac{\partial \boldsymbol{\mu_i}}{\partial \boldsymbol{\beta}}$$

A two stage iterative procedure is then used. 1. With the current estimates of $\boldsymbol{\alpha}$ and ϕ we obtain an estimate of $\boldsymbol{\beta}$. 2. Using the current estimate of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ are estimated using the standardized residuals:

$$r_{ij} = \frac{(Y_{ij} - \hat{\mu}_{ij})}{v\left(\hat{\mu}_{ij}\right)^{.5}}$$

The last model considered is a three level random effects model, found in Carpenter et al. (9).

Logistic regression - Three level random effects model

The statistical model employed in Carpenter et al. (9) uses a generalized linear mixed model with a logit link function. It is a three level random effects model with random intercept terms for hospital nested within SEER region at the second level and SEER region at the third level. There are fixed effect covariates at the person (first) level as well as at the hospital (second) level.

It can be formulated as follows:

Model description

Let N be the total number of patients observed. There are m third level units, SEER regions, which are indexed by k such that $k = 1$ to m. There are n_k second level units, hospitals, in each third level unit which are indexed by jk. Let l_{jk} be the count of first level units, patients, in each hospital, indexed by ijk. Thus, the j^{th} hospital in the k^{th} SEER region has l_{jk} patients. Let $\mathbf{Y}$ be the response vector of length N consisting of zeros and ones indicating each patients (non)receipt of SLNB. $\boldsymbol{\pi}$ is a vector of length N consisting of the expected value of Y, that is the probability of SLNB, for each patient. $\boldsymbol{\eta}$ is a vector of length N consisting of the log odds of SLNB for each patient. $\mathbf{x}^{(\mathbf{1})}_{\mathbf{ijk}}$ is a vector of length p whose elements consist of first (patient) level explanatory variables. $\mathbf{x}^{(\mathbf{2})}_{\mathbf{jk}}$ is a vector of length q whose elements consist of second (hospital) level explanatory variables. $\boldsymbol{\beta}^{(\mathbf{1})}$ and $\boldsymbol{\beta}^{(\mathbf{2})}$ are the corresponding fixed effect parameter value vectors of length p and q respectively. $\mathbf{b}^{(\mathbf{2})}$ and $\mathbf{b}^{(\mathbf{3})}$, are multivariate normal random variables with means of zero and $\mathbf{b} \sim MVN\left(\mathbf{0}, \mathbf{G}\right)$, where $\mathbf{G}$ is the covariance matrix.

The three levels of the systematic component for each patient are:

$$\text{First level model} \qquad \eta_{ijk} = \beta_{0jk} + \mathbf{x_{ijk}^{(1)T}}\boldsymbol{\beta^{(1)}}$$

$$\text{Second level model} \qquad \beta_{0jk} = \beta_{0k} + \mathbf{x_{jk}^{(2)T}}\boldsymbol{\beta^{(2)}} + b_{jk}^{(2)}$$

$$\text{Third level model} \qquad \beta_{0k} = \beta_0 + b_k^{(3)}$$

Thus, β_0 is the fixed effect that represents the mean of the logit across all SEER regions, β_{0k} is the mean value for the k^{th} SEER region, and β_{0jk} is the mean value for the j^{th} hospital in the k^{th} SEER region.

Parameters were estimated using maximum likelihood but required the use of quadrature in order to generate the estimates. This model assumes that any missing outcome variable data are missing at random (MAR) (46; 28; 49).

The following description of the estimation procedure draws heavily on the SAS documentation (47) for the procedure used to fit the models, Proc GLIMMIX (see chapter 38), as well as Professor Herrings 767 course notes.

Parameter Estimation

The joint probability density function, in general, is given by:

$$f\left(\mathbf{Y_i}|\mathbf{X_i}, \mathbf{b_i}\right) f\left(\mathbf{b_i}\right)$$

but since the $\mathbf{b_i}$ are unobserved the marginal likelihood function is used:

$$\prod_{i=1}^{N} \int f\left(\mathbf{Y_i}|\mathbf{X_i}, \mathbf{b_i}\right) f\left(\mathbf{b_i}\right) d\mathbf{b_i}$$

A two step procedure is used to get the maximum likelihood estimates, first obtain estimates

for $\boldsymbol{\beta}$ and $\mathbf{G}$ based on the marginal likelihood, using adaptive quadrature numerical integration for approximate estimates. Next, using these estimates generate predicted random effects values:

$$\hat{\mathbf{b}}_i = E\left(\mathbf{b_i}|\mathbf{Y_i}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{G}}\right)$$

SAS 9.2 implements adaptive quadrature as follows (47), the quadrature rule is:

$$\int_{-\infty}^{\infty} f(x)\, p(x)\, dx \approx \sum_{r=1}^{Q} w_r f(x_r)$$

where $p(x)$ is a probability density function, $f(x)$ is some function to be integrated against it, Q is the number of quadrature points, r is its index, and w_r are the quadrature weights. In our case $f(x)$ is the conditional distribution given the random effects, and $p(x)$ is the random effects distribution. When the number of quadrature points is not specified ahead of time then Proc GLIMMIX determines the number of quadrature points by evaluating the log likelihood at an increasing number of points until a tolerance is met. Additionally, and separately, 'the procedure centers and scales the quadrature points by using the empirical bayes estimates (EBEs) of the random effects and the Hessian matrix from the EBE suboptimization.' The manual goes on to state that this process improves the likelihood approximation 'by placing the abscissas according to the density function of the random effects.'

1.2.3 Limitations of Current Work

Looking at Carpenter et al. (9), being the work that most closely modeled the structure inherent in SEER-Medicare data of the papers considered, we find several limitations that impair the interpretability of the results.

Failure to account for repeated measures

One major limitation is the failure to take into account the repeated measures nature of the data, this is compounded by the collapsing to diagnosis year from diagnosis month. The data provided by the SEER-Medicare program provides the month of diagnosis rather than the exact date to help

ensure anonymity of the patients. This study, like the others, however collapses the data to year of diagnosis which both obscures within year trends and leads to having unidentified clusters, the hospital and diagnosis month combination, within each diagnosis year. It's been shown that failing to include all levels in a binary response multilevel model leads to biased estimates (51). Using a Laird and Ware repeated measures type model (28) treating the hospitals, nested within region, as the units and maintaining the uniqueness of the diagnosis month, would have permitted more accurate parameter and standard error estimates and provided more information about the shape of the outcome trajectory.

Failure to account for covariates measured with error

Despite the potential for measurement error in the covariates, as is customary in many fields, the covariates are treated as if they were perfect measurements of the construct of interest in all of the papers considered. It has been shown that treating stochastic variables in a GLMM as if they were non stochastic can lead to bias in parameter estimates and decreased precision of the estimates (57). Models for GLMMs with measurement error in the covariates have been described as generalized linear mixed measurement error models (GLMMeM) (58) and over the past 15 years or so, a variety of approaches have been developed for working with this class of models. Several methods have lately been proposed for addressing this issue primarily via the use of instrumental variables and alternatives to maximum likelihood for parameter estimation (30; 42). Other approaches include regression calibration, simulation extrapolation, likelihood based methods, and Monte Carlo EM (MCEM) algorithm estimation of the MLE. (64; 7).

In regression calibration the true value of the covariate measured with error is predicted by use of a regression model based upon the other available covariates. The predicted values are then used in place of the original measured with error values in the model of interest without any further adjustments. There are however issues with regression calibration for GLMMeMs with binary outcomes as pointed out in Carrol et al. (2006) (10). While in general substituting the predicted values would correctly specify the fixed effects structure in a GLMM (given that it was orthogonal to the random effects structure) it doesn't correctly specify the random effects structure.

Given that in the binary outcome case they are not orthogonal then both the random effects and the fixed effects estimates will be biased (58).

Simulation extrapolation is most suited for measurement errors that are additive (10) and gives an approximate but inconsistent estimator (30). While the numerical integration estimates of the (intractable) log likelihood suffer from high dimensionality of the random effects (64). The MCEM approach can have computational challenges such as non-convergence due to the random effects (64). Thus Li and Wang's (2012) (30) approach combining instrumental variables with a method of moment estimator (MME) using a simulation based approach in the cases (such as a GLMMeM with a binary outcome) where the moments are intractable may be a useful tool. Their simulation based approach also has the benefit of not requiring normally distributed random effects (29).

Limited Model Checking

Another limitation is that limited model checking was performed. While likelihood ratio tests indicated that the fixed and random effect terms in the final model provided a better fit than the models without them, no checking of model assumptions, particularly the normality of the random effects, was attempted. It is known that the maximum likelihood parameter estimates for binary outcome type Generalized Linear Mixed Models are sensitive to departures from normality for the random effects (21; 18; 34; 33). Methods have been proposed for testing for model misspecification but were not used (1). Perhaps it would be better to assume an alternate distribution or to use a nonparametric estimate of the random effect (27; 2; 29). Tests for normality of the random effects have been proposed (50; 18) and could have been used to test the need for non-normal random effects.

Methodological Limitations

Methodological limitations primarily were a function of the exclusion criteria for the study, they included: only examining a fairly short time span even though more data was available, using data from a limited number of SEER regions (seven), excluding those patients that did not receive any nodal staging, and excluding those cases whose staging information possibly did not meet the criteria for use of SLNB. While there were reasons for all of these decisions there also are

approaches, either statistical or methodological, that would permit use of a much larger and richer analytical data set that would provide both more power and greater insight into when SLNB was performed.

In the next section we consider estimators of the variance parameters of the random effects in Generalized Linear Mixed Models and some methods to compare thier relative efficiency.

1.3 Random Effect Parameter Estimation in Generalized Linear Mixed Models

The two primary methods for estimating the variance parameters of the random effects in Generalized Linear Mixed Models are Maximum Likelihood (ML) (with numeric integration) and Pseudo-likelihood (PL) (also called penalized quasi-likelihood). ML is the gold standard, particularly for binary outcomes, but is computationally intensive. The ML using quadrature algorithm that is implemented in SAS Institute's Proc Glimmix has been described in section 1.2.2.

1.3.1 GLMM specification

In this section we will use the following specification of the GLMM. Note that g^{-1} is the inverse link function and in this section $\boldsymbol{Z}$ is the design matrix for the random effects. Other terms are as defined previously in section 1.2.2.

$$E(\boldsymbol{Y}|\boldsymbol{b}) = g^{-1}(\boldsymbol{\eta})$$

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}$$

Where

$$\boldsymbol{b} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G})$$

and

$Var(\boldsymbol{Y}|\boldsymbol{b})$ comes from the exponential family

1.3.2 Pseudo-likelihood (PL)

The use of Pseudo-likelihood for estimation of variance parameters in a GLMM was proposed in two separate papers in 1993 (5; 63). Both approaches make use of the generalized mixed model

equations and iteratively solve for $\boldsymbol{\beta}$ and $\mathbf{b}$.

$$\begin{bmatrix} \mathbf{X^T W X} & \mathbf{X^T W Z} \\ \mathbf{Z^T W X} & \mathbf{Z^T W Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X^T W y^*} \\ \mathbf{Z^T W y^*} \end{bmatrix}$$

Where $\mathbf{y}^*$, the 'pseudo-data', is alternately used to estimate $\boldsymbol{\beta}$ and $\mathbf{b}$ and is estimated using their estimates from the previous iteration. Given its relative ease of computation it was the original method used in Proc Glimmix and remains the default method.

In the next section bias in these estimators is discussed.

1.3.3 Bias in Estimates

It has been shown that for GLMMs with binary outcomes that PL based random effect variance estimators are biased with the magnitude of the bias inversely related to cluster size (6; 32; 43). ML based estimates have been found to have less/little bias in this case with quadrature based approximations performing the best and Laplace based approximations somewhat less so (43).

In the next section one potential way of addressing parameter estimator bias is covered, the jackknife procedure.

1.3.4 Jackknife process

The Quenouille-Tukey jackknife is a procedure for nonparametrically estimating a function (statistic) of some unspecified distribution from the data at hand that should have less bias than some simpler estimator and is a special case of bootstrapping (16). Simply put, the jackknife estimator is the average of all possible instantiations of some estimator where a single data point is left out of its calculation.

Parameter Estimation

The following specifies the jackknife estimator, $\hat{\theta}_{(.)}$ for some estimator $\hat{\theta}$.

$$\hat{\theta}_{(i)} = \hat{\theta}(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{(i)}$$

Bias Estimation and correction

Bias estimation can be done by Quenouille's estimate of bias

$$\widehat{BIAS} = (n-1)(\hat{\theta} - \hat{\theta}_{(.)})$$

and the bias corrected estimator is simply

$$\tilde{\theta} = \hat{\theta} - \widehat{BIAS}$$

The next section covers three types of relative efficiency for multivariate estimators such as the estimators of covariance matrices.

1.3.5 Comparing estimators via D-, A-, and E-efficiency

This section heavily draws on section 8.6 of Jureckova, Sen, and Picek (2011) (23) where they discuss multivariate efficiency. In all three types of relative efficiency we will consider the reference is the ML estimator and it is the Fisher information from that which is compared to the asymptotic covariance matrix of the estimated random effects variance terms from the estimator which is under consideration. Essentially each type of efficiency is a different statistic of the eigenvalues of the product of asymptotic covariance matrix and the MLE Fisher's information. The specifics follow.

Let $\boldsymbol{T}_n$ be the estimator of interest of parameter $\boldsymbol{\theta}$, in our case the covariance matrix of the random effects, and let $\boldsymbol{v}_T$ be the dispersion matrix of $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta})$. Let $\mathcal{I}(\boldsymbol{\theta})$ be the Fisher information matrix and $\boldsymbol{D}^0$ is the diagonal matrix of the p eigenvalues of $\boldsymbol{v}_T\mathcal{I}(\boldsymbol{\theta})$.

Then we have:

D-efficiency

is the p^{th} root of the determinant of $(\boldsymbol{D}^0)^{-1}$

A-efficiency

is the mean of the eigenvalues of $(\boldsymbol{D}^0)^{-1}$

E-efficiency

is the largest eigenvalue of $(\boldsymbol{D}^0)^{-1}$

2 CHARACTERIZING INNOVATION ADOPTION

2.1 Introduction

Quality of care in the treatment of cancer is an important issue and awareness has been heightened ever since the Institute of Medicine released its recommendations in **Ensuring Quality Cancer Care** (39). One of the IOM's recommendations of primary importance to this work is to "Measure and monitor the quality of care using a core set of quality measures." Other recommendations included; "Services for the un- and underinsured should be enhanced to ensure entry to, and equitable treatment with, the cancer care system", "Studies are needed to find out why specific segments of the population ... do not receive appropriate cancer care", and "Cancer care quality measures should be used to hold providers ... accountable for demonstrating that they provide and improve quality of care." While the first and last call explicitly for a set of metrics, the other two implicitly require metrics in order to be achieved. The focus of this work is to generate the basis for a metric that can be used to measure, and ultimately improve, quality of medical care for all patients.

One indication that quality care is being provided is the use of new evidence based procedures. However, to date, there is no metric that captures the pattern of institutional adoption of a new evidence based procedure. We propose that Innovation Adoption Trajectories could be used as the basis for such a metric. The example we apply this approach to is the case of the adoption of Sentinel Lymph Node Biopsy (SLNB) for pathologic lymphatic staging of early stage breast cancer in the Medicare population. The statistical model used to generate these trajectories is a Logistic Normal Generalized Linear Mixed Model. In the next three sections we define and describe the theoretical basis for Innovation Adoption Trajectories, provide background information about Sentinel Lymph Node Biopsy, and give a brief description of how the model is used to generate

trajectories of the probability of use of SLNB.

We begin with a description of Innovation Adoption Trajectories.

2.1.1 Innovation Adoption Trajectories

We define innovation adoption trajectories as the model based estimates (or predicted values) of the probability of use of a clinical trial proven procedure over the period of adoption of the novel procedure by the medical community. It is proposed that these trajectories, derived from institutional claims and registry data, can be a useful tool for understanding promoters and barriers to evidence based medical practice. Characterization of these trajectories is motivated by Rogers' (1962) theory of adoption of innovation.

Rogers proposed that the uptake by a population of an innovation, be it a product, an idea, or a procedure, follows a consistent pattern. This pattern can be described by a S-shaped curve on a graph, with the proportion of a population that has adopted the innovation on the y-axis and time on the x-axis (Figure 2.1 - red dashed line). It is similar to the cumulative density function (CDF) of the normal distribution, with the exception that it starts at zero and can eventually reach 100%. He also proposed that there are five 'adopter types' in a population.

His taxonomy for these adopter types is Innovators, Early Adopters, Early Majority, Late Majority, and Laggards. He suggests that the proportion of a population that falls into each category is, respectively, 2.5%, 13.5%, 34%, 34%, and 16%, the distribution of which resembles the probability density function of the normal distribution (Figure 2.1 - blue solid line).

Note that in Rogers' theory it is assumed that an individual either uses or doesn't use an innovation. Our innovation adoption trajectories assume that an individual (e.g., institution) has multiple opportunities to make use of the innovation and does so with some probability that can vary over time. It is assumed prototypically that these trajectories are monotonically increasing functions over the time period that the medical community is adopting the innovation and have an asymptote of one. However, neither of these is required to be the case. Hypothetical prototypical

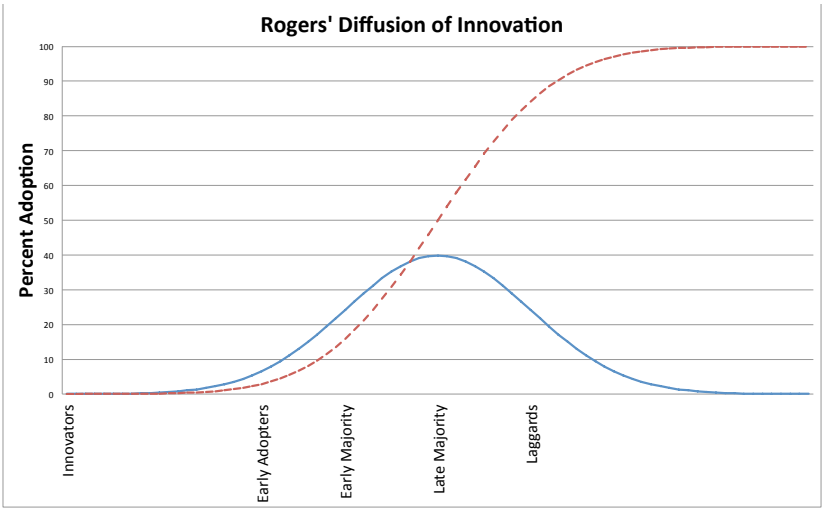

Figure 2.1: Distributions of Rogers' adopter types

trajectories corresponding to the different adopter types are shown in Figure 2.2. These hypothetical trajectories were generated using the CDFs of five Weibull distributions. The parameters for these distributions were selected to approximate a 0.8 probability of use of the innovation at the theoretical time point that an adopter type starts adopting an innovation and are simply meant to be illustrative. To the best of our knowledge this is a novel extension of Rogers' theory.

It is hypothesized that the trajectory of the medical procedure's use is determined by the culture of the institution as well as patient characteristics including morbidity. Institutional affiliations are proposed as a proxy for institutional culture.

Next we describe SLNB and give some background about it's adoption and describe why it is a preferred staging procedure, in some cases, relative to the previous standard. Previous research on it's adoption is also presented.

2.1.2 Adoption of Sentinel Lymph Node Biopsy

The example used in this work is the case of the adoption of sentinel lymph node biopsy (SLNB) in early stage breast cancer (BC) treatment in the Medicare population over the time period

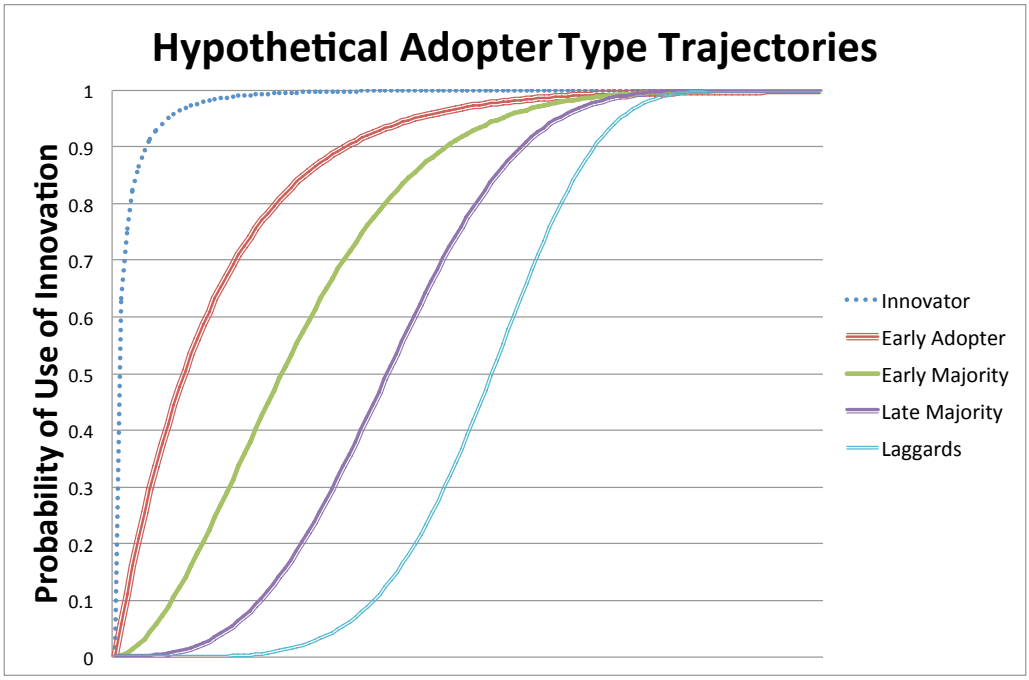

Figure 2.2: Hypothetical innovation adoption trajectories by adopter type.

of 1999 to 2007 when it was starting to be adopted by the medical community. Originally called sentinel lymphadenectomy (20), it was being used in place of Axillary Lymph Node Dissection (ALND) in some cases when appropriate (20; 8; 15; 19), although seemingly with disparities for at risk groups (22; 35; 11; 45; 9; 44; 38).

Lymph Node Staging Procedures

Axillary Lymph Node Dissection (ALND)

ALND involves the removal of all level one and two ipsilateral axillary lymph nodes. It is an extensive procedure and associated morbidities include, lymphedema, nerve paresthesias, axillary seromas, and infections. Its primary purpose is to provide pathological nodal staging information. It also serves the function of removing any metastatic tissue that is involved with the axillary lymph nodes.

Sentinel Lymph Node Biopsy (SLNB)

SLNB is a less invasive procedure than ALND, where a few lymph nodes are removed, typically 1 to 3, and intraoperatively assessed for metastatic involvement (8; 24). A standard version of the procedure during the time frame of this study started with injecting a blue dye and a radioactive tracer near the site of the tumor (36). After some time has passed both visual inspection and a Geiger counter are used to determine which lymph node(s) the lymph near the tumor drains to. These are excised and assessed for metastatic involvement. If it is determined that there is involvement then the next node in the lymph chain is excised and examined. Using this information it can then be determined whether to perform a completion ALND.

SLNB as Standard of Care for Early Stage Breast Cancer Treatment

Up through the late 1990's the standard practice had been to perform ALND for pathologic lymph node staging in cases of early stage BC. With the advent of SLNB in the mid 1990's and a growing body of evidence that it was equivalent to ALND in outcomes for pathologic node negative cases it became the standard of care (62; 13; 52; 14; 41; 54; 55; 61; 56). Medicare started reimbursing its use in 1999 while it was still in clinical trials since early evidence suggested it was a preferred treatment option in many cases. With the conclusion of the clinical trials SLNB became the documented standard of care for lymphatic staging of early stage breast cancer (26).

Previous Research

Factors that have been associated with the rate of SLNB's use over time include disease, patient, neighborhood, physician, institutional, and regional characteristics (22; 35; 11; 45; 9; 44; 38; 3). However, to date, the literature on this topic does not take into account both the multilevel and repeated measures over time aspects that are inherent in this data. Prior research on factors influencing the use of SLNB in BC treatment primarily used multivariable logistic regression (22; 35; 11; 45; 3). Two studies used intercept only random effects models (9; 38) to control for the multilevel nature of the data. One study used logistic regression with GEE based estimates

of the within hospital correlations (44). These last three studies essentially assumed that the correlations between any two patients within a given hospital were the same regardless the amount of time between their diagnoses. All of these works treated time (diagnosis year) as an unordered categorical variable.

The next section briefly describes the statistical model used and how trajectories are generated from it.

2.1.3 Use of Generalized Linear Mixed Model

The method we used to generate innovation adoption trajectories is to model the receipt of the novel procedure (0,1) at the patient level by applying a generalized linear mixed model (GLMM - logistic normal). This model is parameterized to capture the multilevel (patient, institution, and region) and longitudinal nature of the data. This type of model has been extensively described (28; 49; 17) as well as best practices for fitting (12; 1). We then make use of the estimated values over time, based upon the fixed effects, to generate regional trajectories. Generation of institutional trajectories is accomplished by calculating the predicted values over time for a given institution by combining the relevant region and hospital level fixed effects and the best linear unbiased predictors (BLUPs) of the random effects for that institution.

2.2 Methods

2.2.1 Data

We make use of the Surveillance, Epidemiology, and End Results (SEER)-Medicare database which has both cancer registry and procedure claims data, as well as some patient demographics and institutional data (59; 4; 25; 48). It is derived from 14 cancer registries covering 17 geographic regions across the US. Individual cases are linked to the corresponding Medicare claims and eligibility information. For those cases where the first primary cancer occurred at an age of 65 or greater there is a success rate of 94 percent in linking to Medicare data. It has been described extensively elsewhere, particularly in a supplement to Medical Care in 2002 (59; 4; 25; 48). What follows in a basic description of each of the components.

We start with the cancer registry data and how it's collected.

SEER Cancer Registry Program:

The National Cancer Institute's Surveillance Research Program contracts with state health programs and universities to operate incident cancer registries in geographic regions around the United States of America. Currently the registries cover about 28% of the US population. The 17 registries are each headed by a different investigator and each registry works differently. They do however use a standardized reporting system for data transmission to NCI. When data is received by NCI it is stored in a format of one record per incident cancer diagnosis. It should be noted that the program is for the most part facility based and thus information for services provided in a physician's office might be missing or incomplete. It should also be noted that for many variables there is a hierarchical structure, that is, some values supersede other values even though both are true (e.g., the maximum value is reported). The types of data included are: diagnostic, staging, treatment, and limited demographics.

Registry data can come from a variety of sources, but a case is first identified via a report from: a service provider where a cancer diagnosis was made, an autopsy report, or from a death certificate submitted to the cancer registry for the geographical region where the entity is located. In SEER-Medicare research it is a standard practice to exclude those cases that were only identified via autopsy or death certificate. Thus the reporting sources include at least one of: a Hospital Inpatient/Outpatient or Clinic, a Laboratory (Hospital or Private), a Physician's Office/Private Medical Practitioner, or a Nursing/Convalescent Home/Hospice.

There are a variety of missing data issues for SEER data. Registries don't capture 100% of all incident cases, although the expectation is at least 98%. Missing data is more likely to have come from patients who were primarily treated at physician's offices (rather than at a hospital or clinic). Pathological staging data 'overwrites' clinical staging data and if is unknown if this has occurred for breast cancer (prostate cancer has separate clinical staging variables). Changes in coding systems over time have led to some staging and treatment information being collected during some time frames but not others (or defined somewhat differently). Race data categories

depend on time period, SEER race data however is considered better than Medicare race data. Six months of 2005 data for Louisiana are somewhat sparse due to severe storms and an indicator variable flags this time period. There were four new SEER registries in 2000 and there is no prior data for them.

Finally it should be noted that SEER data is not a probability sample and all generalizations to populations outside of the SEER regions must be model based. From the SEER website, "Geographic areas were selected for inclusion in the SEER Program based on their ability to operate and maintain a high quality population-based cancer reporting system and for their epidemiologically significant population subgroups. The population covered by SEER is comparable to the general US population with regard to measures of poverty and education. The SEER population tends to be somewhat more urban and has a higher proportion of foreign-born persons than the general US population."

Next we consider the Medicare data that is provided by NCI in their release of SEER-Medicare data.

Medicare data as Provided by the SEER-Medicare Program:

Medicare data used in this study consists of eligibility, inpatient stay, outpatient service, physician service, and durable medical equipment (which includes some cancer medications) information. Medicare eligible individuals include qualifying individuals age 65 or greater, those with end stage renal disease, and certain people with disabilities. Note that this is administrative (billing) claims data for services that are paid for based on Medicare policy in effect at the time of service. This has implications for completeness, particularly for procedures not covered at the time of service.

Medicare data also has its share of missing data issues. Physicians tend to be erratic in billing for services they know Medicare won't pay for, so for uncovered services there will be 'missing' claims data. Similarly some services are 'bundled' with others and only the code for the more extensive procedure is provided. One needs to be familiar with Medicare policy during time period of interest and during some years the location of interest in order to finesse the complexities.

Medicare was supposed to have the exact same coverage for services throughout the US but during the early 2000's some of the fiscal intermediaries (FI) had slightly different policies or effective dates for service coverage. E.g. different regions (primarily) of the country had different start dates for coverage of SLNB depending on which FI processed a hospital's claims. An additional issue is that a person always has the same FI processing their claims as when they first signed up. Thus, individuals that moved during the time period where there were differences in FI policies may have had the policies imposed based upon their previous residence's FI.

An overview of SEER-Medicare data methodology follows.

SEER-Medicare data Methodology:

SEER-Medicare methodology is as follows. Selected SEER registry data elements (those elements with known poor reliability are excluded) from each incident cancer diagnosis, up to the first ten, are put into a format of one record per person. This data is then linked to Medicare eligibility data and a 94% linkage rate has been found for cases where the first cancer occurred at an age of 65 or greater. Cases where there is a successful link are kept. It is unknown whether there is any selection bias in the linkage process but assumed not.

This information is combined with demographic and eligibility information from the Medicare program to create the Patient Entitlement and Diagnosis Summary File (PEDSF) file. The name is a little misleading as it includes all of the (selected by NCI) SEER data, not just the diagnosis information. It also includes (de-identified and 'fuzzed') census tract and zip code level demographic (census) information. Census data sometimes is missing, and the usual strategy is to use zip code based data in this case. There are however, still a (very small) number of missing values after this strategy is employed.

Next is the inclusion criteria for this study.

Inclusion Criteria

The main inclusion criteria for this study are that the Medicare recipient is female with her first or only incident primary breast cancer occurring during the years of 1999 to 2007. A valid diagnosis month must be present and the reporting source must not be autopsy nor death certificate. This

cancer must have occurred at age of 66 or later and the basis for Medicare coverage must not be End Stage Renal Disease (ESRD). The individual must have had both parts A and B coverage from 12 months prior to diagnosis month until 12 months post diagnosis month (or until death) without HMO coverage during this time period. This is to ensure that the claims data is available for both inpatient and outpatient services. It also enables calculation of a comorbidity index for preexisting conditions. The inclusion criteria for this study are broad with respect to disease characteristics. This will permit examining both indicated as well as contraindicated use of SLNB. It also provides more information to base the models upon in comparison to sub-setting to some smaller population. Definitive treatment must have occurred within 12 months of diagnosis. This is necessary in order to be able to identify the institution where a SLNB may have been performed. After all criteria were applied there were 76478 patients included in the analysis and 2030 institutions where they received treatment.

Concluding this section is a description of the variables used in this work.

Variables included in the modeling process

Region level

SEER region, with rural GA and Atlanta combined, is a fixed effect. Time (month of diagnosis), scaled to 0 to 1 from start of study period, January 1999 to end of study period December 2007, Time^2, and Time^3, are fixed effects at the region level.

Hospital level

The continuous time variables are random effects at the hospital level. There were three hospital level fixed effects: ACOSOG, an indicator variable for institutional affiliation with the American College of Surgeons Oncology Group, a sponsor of a SLNB clinical trial, Co-op group, an indicator variable for other NCI cooperative groups having breast cancer research portfolios including; National Surgical Adjuvant Breast and Bowel Project, Cancer and Leukemia Group B, Southwest Oncology Group, and the Eastern Cooperative Oncology Group, and Teaching Hospital, an indicator variable for medical school affiliation. The interaction with the linear effect of time for all

hospital level fixed effects was also examined.

Person level

Demographics

There were three demographic person level fixed effects: Race being African American (AA), an indicator variable, with the reference group being all other races, Age, a categorical variable of patient age at diagnosis, with levels of: 66 to 69 (which is the reference group), 70 to 74, 75 to 79, and 80 plus, and Medicaid, an indicator variable for the patient being dual eligible during the year of diagnosis, a proxy for individual low income status. The interactions of the person level variables were explored as well as interactions with the linear effect of time.

Disease and treatment variables

Disease characteristic variables included the fixed effects of: Tumor grade, a four level categorical variable with the reference being 'well differentiated' and three indicator variables for the levels of 'poorly differentiated', 'moderately differentiated', and 'unknown or not assessed', Tumor size, we transformed the two continuous size variables (there are two due to a change in coding systems over time) into T staging categories from AJCC TNM staging. The reference category was T1c. This was the only staging variable used as it is the only constant staging data over time with changes in systems occurring during the timeframe of this study. Additionally, we do not know whether the N staging reported was clinical or pathologic, presumably for those who received nodal staging it was pathologic, while for those where there was no indication of staging it is unclear. Thus an inaccurate assessment of the effect of N staging is a concern, given that treatment decisions at that point in treatment are based upon the clinical results. The interaction between tumor grade and size was also included.

Treatment variables included surgery type and receipt of SLNB. Surgery was a three level categorical variable for the receipt of breast conserving surgery (BCS) and or mastectomy with the reference category being BCS only. The other two levels were mastectomy only and both BCS and

mastectomy. It was hypothesized that various factors that would influence surgery type (and subsequent surgeries) would also influence the use of SLNB. Receipt of SLNB (0,1) was the outcome variable and based upon both claims and registry data.

2.2.2 Proposed Model

The proposed model is a three level mixed effects logistic normal model with random intercept and cubic polynomial of time at the second level for the systematic component.

Let $\boldsymbol{Y}$ be the response vector of length N consisting of zeros and ones indicating each patients (non)receipt of SLNB. $\boldsymbol{\pi}\left(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\gamma}\right)$ is a vector of length N consisting of the expected value of Y, that is, the probability of the receipt of SLNB for each patient. $\boldsymbol{\eta}$ is a vector of length N consisting of the log odds. $\boldsymbol{X}$ is the design matrix with variables at the patient, hospital, and region levels. $\boldsymbol{\beta}$ is the set of parameter values associated with the respective variables in the design matrix. $\boldsymbol{Z}$ is a subset of $\boldsymbol{X}$ consisting of the design matrix for the random effect variables, in this case the polynomial of time at the second, hospital, level. $\boldsymbol{\gamma}$ is the random effect matrix.

The general form of the systematic component of this model is:

$$\boldsymbol{\eta} = logit\left(E(\boldsymbol{Y}|\boldsymbol{\gamma})\right) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} \tag{2.1}$$

While the error function is:

$$\boldsymbol{Y}|\boldsymbol{\gamma} \sim \text{Bern}(\boldsymbol{\pi}) \text{ where } Y_s \perp Y_t \text{ and } s \neq t \tag{2.2}$$

Note that:

$$\boldsymbol{Z} = [\boldsymbol{1}, \boldsymbol{time}, \boldsymbol{time^2}, \boldsymbol{time^3}] \tag{2.3}$$

and

$$\boldsymbol{\gamma} \sim MVN_4(\boldsymbol{0}, \boldsymbol{G}) \tag{2.4}$$

With an unstructured covariance matrix:

$$\boldsymbol{G} = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix} \tag{2.5}$$

Let N be the total number of patients observed in the M hospitals in the L regions. Each patient is observed only once and they are assumed to be i.i.d. within a hospital and over time conditional upon the patient, hospital, and regional covariates. There are L third level units, regions, which are indexed by k such that $k = 1$ to L. There are n_k second level units, hospitals, in each third level unit which are indexed by jk. There are o_{jk} patients within a given hospital indexed by ijk.

Noting that:

$$\mathbf{x}_{ijk}^{(11)\intercal} = [time_{ijk}, time_{ijk}^2, time_{ijk}^3]$$

The three levels of the systematic component are:

$$\eta_{ijk} = \alpha_{00jk} + \mathbf{x}_{ijk}^{(10)\intercal}\boldsymbol{\alpha}_{01jk} + \mathbf{x}_{ijk}^{(11)\intercal}\boldsymbol{\alpha}_{1jk} \tag{2.6}$$

$$\alpha_{00jk} = \beta_{00k} + \mathbf{x}_{jk}^{(2)\intercal}\boldsymbol{\beta}_{01k} + \gamma_{0jk}^{(2)} \tag{2.7}$$

$$\boldsymbol{\alpha}_{1jk} = \boldsymbol{\beta}_{10k} + \mathbf{x}_{jk}^{(2)\intercal}\boldsymbol{\beta}_{11k} + \boldsymbol{\gamma}_{1jk}^{(2)} \tag{2.8}$$

$$\beta_{00k} = \lambda_{00} + \mathbf{x}_k^{(3)\intercal}\boldsymbol{\lambda}_{01} \tag{2.9}$$

$$\boldsymbol{\beta}_{10k} = \boldsymbol{\lambda}_{10} + \mathbf{x}_k^{(3)\intercal}\boldsymbol{\lambda}_{11} \tag{2.10}$$

$\boldsymbol{x}_{ijk}^{(1)}$ is a vector of length p whose elements consist of first (patient) level explanatory variables. For notational convenience it is broken out into $\boldsymbol{x}_{ijk}^{(10)}$ consisting of the demographic and disease level characteristics that are assumed to have the same effect upon the intercept for the entire

population and $\boldsymbol{x}_{ijk}^{(11)}$ which consists of a cubic of time. Time corresponds to the patient's diagnosis month relative to the start and end of the study (normalized to a value between 0 and 1). Note that the patient level variables (other than time) are stochastic and time varying in this context relative to the higher levels. It is assumed that their values at any given time point are not related to the outcome values of other patients at prior time points. $\boldsymbol{x}_{jk}^{(2)}$ is a vector of length q whose elements consist of second (hospital) level explanatory variables which may be time varying. $\boldsymbol{x}_{k}^{(3)}$ is a vector of length r whose elements consist of third (region) level explanatory variables.

Equations 2.6 to 2.10 can be expressed as a single equation:

$$\begin{aligned}\eta_{ijk} =&\lambda_{00} + \mathbf{x}_{k}^{(3)\intercal}\boldsymbol{\lambda}_{01} + \mathbf{x}_{jk}^{(2)\intercal}\boldsymbol{\beta}_{01} + \mathbf{x}_{ijk}^{(10)\intercal}\boldsymbol{\alpha}_{01jk}+ \\ &\mathbf{x}_{ijk}^{(11)\intercal}\boldsymbol{\lambda}_{10} + \mathbf{x}_{ijk}^{(11)\intercal}\mathbf{x}_{k}^{(3)\intercal}\boldsymbol{\lambda}_{11} + \mathbf{x}_{ijk}^{(11)\intercal}\mathbf{x}_{jk}^{(2)\intercal}\boldsymbol{\beta}_{11k}+ \\ &\gamma_{0jk}^{(2)} + \mathbf{x}_{ijk}^{(11)\intercal}\boldsymbol{\gamma}_{1jk}^{(2)}\end{aligned} \tag{2.11}$$

Where the fixed effect intercept terms are on the first line of equation 2.11, the fixed 'slope' terms are on the second line, and the random intercept and slope is on the third line. Thus from the first line, λ_{00} is the overall intercept, $\mathbf{x}_{k}^{(3)\intercal}\boldsymbol{\lambda}_{01}$ are the regional deviations from the overall intercept, $\mathbf{x}_{jk}^{(2)\intercal}\boldsymbol{\beta}_{01}$ are hospital characteristic deviations from the overall intercept, and $\mathbf{x}_{ijk}^{(10)\intercal}\boldsymbol{\alpha}_{01jk}$ are patient level deviations from the overall intercept. From the second line we respectively, the overall slope, the regional deviations from the overall slope, and the hospital characteristics deviations from the overall slope. The random effects in the third line represent hospitals deviations from the overall intercept and slope since we are assuming that the random effects distributions are the same for all regions.

Concluding the methodology section is a description of the modeling process.

2.2.3 Modeling Process

A forward and reverse stepwise modeling process was used to obtain the best fitting model for this set of data. The decision criteria for retaining variables in the model were chosen as to not over fit the model and with an eye towards parsimony. Likelihood ratio tests were performed,

when appropriate, with the decision criteria of retaining terms whose inclusion generated either a p-value of less than .001 or an improvement in the BIC otherwise.

Preliminary models, looking at the time and random effects specification, started with a simple logistic regression of the use of SLNB with the fixed effect terms of SEER region and a set of indicator variables for time (year) and no other terms in the model. Next was a similar model except for using a cubic polynomial of time instead of the year indicator variables. The third model looked at separate cubic polynomials for each SEER region. The next tested the need for a random intercept of hospital(region). The last three models in the preliminary analysis added the random terms of time, time^2, and time^3 respectively, all with unstructured covariance matrices.

We then proceeded to build a model based on the random effect specification of a cubic polynomial of time with an unstructured covariance matrix for hospital(region). These models added in a stepwise fashion, the fixed effects of disease characteristics, surgery type, patient demographics, and institutional characteristics. Finally a reverse modeling process was employed to determine which variables interacted with linear time and the best fitting model was kept as the final specification of the fixed effects.

To confirm the appropriateness of the random effects specification a reverse process was used and four additional models were created that sequentially removed the highest order random effect term from the preceding model.

2.3 Results

2.3.1 Model Fitting Results

The results of the preliminary model fitting process are shown in Table 2.1. It includes the values for each model of: -2$\times$loglikelihood (-2LL), the likelihood ratio test significance (LRT), when applicable, for comparing the model to the preceding model, and the model Bayesian Information Criteria (BIC). The number of asterisks in the LRT column correspond to p-values of less than 0.05, 0.01, and 0.001.

The results indicate that relative to the standard (and each subsequent) model there are incremental improvements by: using a cubic polynomial of continuous time, having separate polynomials for each SEER region, including a random intercept for hospital, including a random linear effect of time, including a random quadratic of time, and including a random cubic of time. In each case (where applicable) the LRT had a p-value of less than 0.001. In the cases where the LRT was not applicable there was an improvement in the BIC.

Table 2.1: Preliminary Model Fit Statistics

Model	Description	-2LL	LRT	BIC
GLM 1	Fixed effects of Region with categorical time	91103	NA	91373
GLM 2	Fixed effects of Region and cubic of time	90975	NA	91178
GLM 3	Fixed effects of Region interacted with cubic of time	90756	***	91262
GLMM 0	GLM 3 + Random intercept for hospital(region)	82568	NA	82918
GLMM 1	GLMM 0 + Random linear time	81677	***	82156
GLMM 2	GLMM 1 + Random quadratic time	81427	***	81930
GLMM 3	GLMM 2 + Random cubic time	81356	***	81889

Table 2.2 shows the results for the forward stepwise model building process for the proposed fixed effects. All of the proposed main effects were found to improve the model fit with p-values of less than 0.001. The interaction of tumor size and tumor grade was found to be significant ($p<0.001$). The only demographic interaction found to be significant ($p<0.05$, with an improvement in the BIC) was between age and Medicaid status. The interactions of the hospital covariates were not included as they did not meet the inclusion criteria. The only interaction with time found to be significant ($p<0.001$) was that with the set of hospital covariates.

Table 2.2: Model Fit Statistics for Significance of Fixed Effects

Model	Description	-2LL	LRT	BIC
GLMM 4	GLMM 3 + Tumor size and grade	77841	***	78450
GLMM 5	GLMM 4 + Tumor size grade interaction	77711	***	78503
GLMM 6	GLMM 5 + Sugery type(s)	76554	***	77361
GLMM 7	GLMM 6 + Demographics	74651	***	75496
GLMM 8	GLMM 7 + Hospital covariates	74442	***	75310
GLMM 9	GLMM 8 + Demographic interactions	74426	NS	75371
GLMM 10	GLMM 8 + Age by Medicaid interaction	74432	*	75323
GLMM 11	GLMM 10 + Hospital covariate interactions	74416	**	75338
GLMM 12	GLMM 10 + Hospital covariates by time	74381	***	75295

Table 2.3 shows the results for the confirmation of the necessity of the random effects terms. These results were based on a reverse stepwise process with GLMM 13 being compared to GLMM 12, GLMM 14 being compared to GLMM 13, and so on. The results indicate that all random effects terms were highly significant (for the nested models) with p-values less than 0.001 and that excluding the random intercept substantially increased the BIC. Given the confirmation of the best fitting model including the cubic polynomial of time in the random effects, GLMM 12 was selected as the final model.

Table 2.3: Model Fit Statistics for Significance of Random Effects

Model	Description	-2LL	LRT	BIC
GLMM 13	GLMM 12 without random cubic of time	74476	***	75337
GLMM 14	GLMM 13 without random quadratic of time	74742	***	75580
GLMM 15	GLMM 14 without random time	75531	***	76353
GLM 4	GLMM 15 without random intercept	81947	NA	83151

In the next section we present the parameter estimates for the final model.

2.3.2 Final Model Parameter Estimates

Parameter estimates for the best fitting model are presented for the fixed effects in Tables 2.1 to 2.6 and for the random effects in Table 2.7. We start with the fixed effects.

Fixed Effects

Table 2.4: SEER Region Parameter Estimates

SEER Region	Intercept	time	$time^2$	$time^3$
1 - San Francisco	-3.35	4.31	2.79	-2.78
2 - Connecticut	-4.35	14.24	-15.26	6.32
20 - Detroit	-4.80	13.27	-14.89	7.86
21 - Hawaii	-4.09	6.38	-0.88	-0.12
22 - Iowa	-5.29	10.65	-4.70	0.08
23 - New Mexico	-3.31	2.38	10.23	-8.38
25 - Seattle	-2.98	11.24	-8.39	1.60
26 - Utah	-3.44	2.64	9.93	-8.05
27 - Atlanta and 37 - Rural Georgia	-3.83	9.94	-6.94	1.80
31 - San Jose	-3.09	5.31	2.46	-3.48
35 - Los Angeles	-3.79	10.55	-10.26	4.10
41 - Greater California	-3.29	6.40	-0.14	-1.84
42 - Kentucky	-4.00	11.56	-11.44	5.08
43 - Louisiana	-3.29	4.64	1.25	-1.81
44 - New Jersey	-3.25	9.83	-9.28	3.80

Table 2.5: Tumor Characteristics Parameter Estimates

Tumor Grade/Size	T1mic	T1a	T1b	T1c	T2	T3/4	Unkown
Well Differentiated	-1.07	-0.27	0.04	0.00	-0.33	-0.81	-1.33
Moderately Differentiated	-1.12	-0.53	-0.05	-0.01	-0.46	-1.21	-1.48
Poorly Differentiated	-0.71	-0.71	-0.30	-0.22	-0.62	-1.61	-1.67
Undifferentiated	-1.24	-1.54	-1.30	-1.14	-0.93	-1.41	-1.70
Unknown, Not Assesed	-0.94	-1.06	-0.87	-0.52	-0.68	-1.21	-1.85

Table 2.6 gives parameter estimates for the final model treatment, demographic, and hospital variables. Note that these are on the log odds scale as the presence of interactions would require calculations at specific levels to give meaningful odds ratios for most of the effects. The baseline group for this model is patients who: were non African American, received BCS only, had a well differentiated tumor, whose tumor size was T1c, were of age of 66 to 69 at diagnosis, not on medicaid, and whose hospital had no affiliations.

Thus relative to this group we can see that receipt of mastectomy, whether as the first surgery or subsequent surgery, led to a decrease in the odds of receipt of SLNB. Likewise for African Americans. For older women and those on Medicaid the story is a little more complicated given the significant interaction. The interaction can however, be interpreted as either an additional increase in the (negative) effect of age for those on Medicaid or an additional (negative) effect of Medicaid for older patients. Receiving treatment at a teaching hospital, at cooperative group affiliated hospital, and particularly an ACOSOG affiliated hospital, early in the time period examined, led to a increase in the odds of receipt of SLNB. Although this increase in odds diminished over the course of the study with the teaching hospital and ACOSOG effects being reduced to almost nothing. Interestingly, the cooperative group effect did not attenuate as much over time. Note that time effects were examined for the patient level covariates in a separate model (data not shown) whose results indicated that there was no change over time for the patient demographic variables.

Table 2.6: Treatment, Demographic, and Hospital Variables Parameter Estimates

Effect	Estimate	Std. Error	DF	t	p value
BCS Only	0.000	.	.	.	.
Mastectomy Only	-0.694	0.023	74346	30.23	<.0001
Both BCS and Mastectomy	-0.597	0.032	74346	18.71	<.0001
African American	-0.315	0.044	74346	-7.14	<.0001
65 to 69	0.000	.	.	.	.
70 to 74	-0.058	0.029	3301	-2.00	0.0453
75 to 79	-0.174	0.029	3301	-5.92	<.0001
80 plus	-0.905	0.029	3301	30.84	<.0001
On Medicaid	-0.344	0.061	74346	-5.62	<.0001
65 to 69 on Medicaid	0.000	.	.	.	.
70 to 74 on Medicaid	-0.111	0.082	74346	-1.35	0.1783
75 to 79 on Medicaid	-0.269	0.085	74346	-3.14	0.0017
80 plus on Medicaid	-0.142	0.084	74346	-1.70	0.0893
Teaching Hospital	0.334	0.132	74346	2.54	0.0111
Cooperative Group Affiliation	1.064	0.154	2014	6.89	<.0001
ACOSOG Affiliation	1.243	0.181	2014	6.85	<.0001
Teaching Hospital over time	-0.330	0.176	74346	-1.88	0.0601
Cooperative Group over time	-0.335	0.188	74346	-1.78	0.0756
ACOSOG over time	-1.068	0.212	74346	-5.05	<.0001

Random Effects

Table 2.7 gives the random effects variance components estimates for the final model in bold. The standard errors are in parentheses and the correlations are in the lower triangle of the matrix in italics.

Table 2.7: Variance Component Estimates

	Intercept	Time	Time^2	Time^3
Intercept	**3.3** (0.5)	**-7.0** (2.5)	**5.7** (5.0)	**-0.9** (3.0)
Time	*-0.43*	**82.9** (20.5)	**-159.1** (43.7)	**84.7** (26.2)
Time^2	*0.17*	*-0.93*	**354.8** (95.4)	**-206.0** (57.8)
Time^3	*-0.04*	*0.83*	*-0.98*	**125.6** (35.2)

2.3.3 Trajectories Based on Final Model

Figure 2.3 shows the trajectories for a typical hospital that has no affiliations in each SEER region for patients with the reference values for the disease and demographic variables. These trajectories are based on the fixed effect parameters of SEER region and their interactions with the cubic polynomial of time.

Figure 2.4 shows the trajectories of five example hospitals within the San Francisco SEER selected to demonstrate the amount of variability within a SEER region.

Figure 2.5 shows the trajectories for the different levels of variables that policy could presumably impact within the San Francisco SEER regioin. For example hospitals could recieve incentives to belong to a cooperative group or at risk populations could be targeted to help alleviate the disparity in services recieved in comparison to the not at risk population.

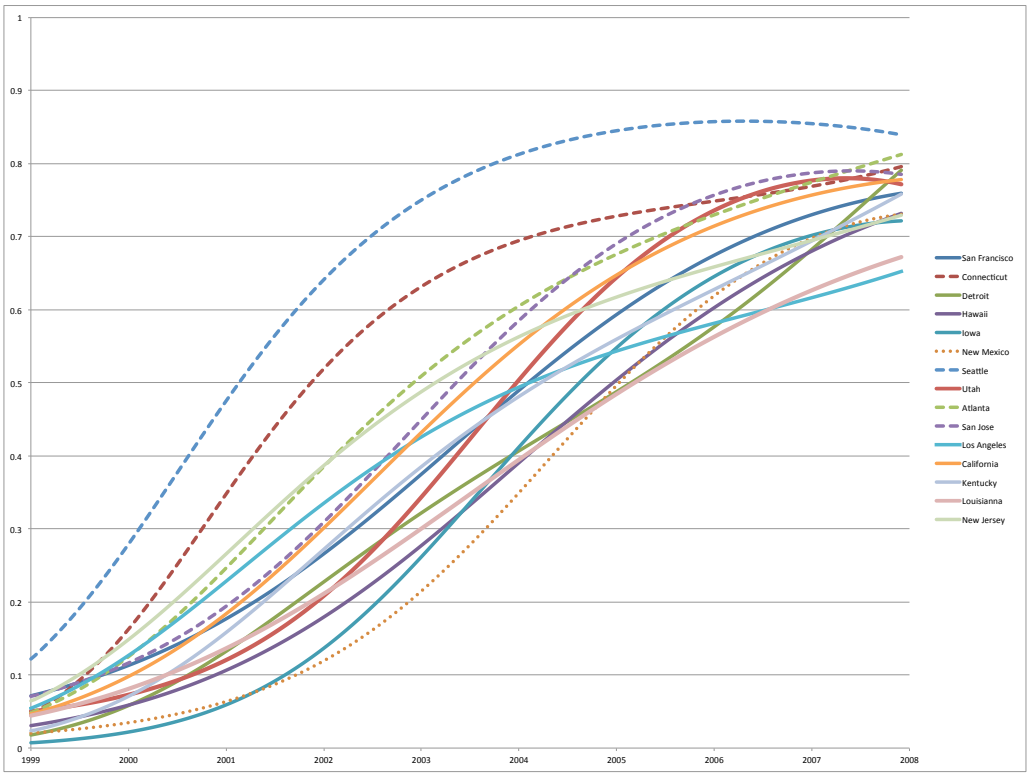

Figure 2.3: SEER Region Trajectories

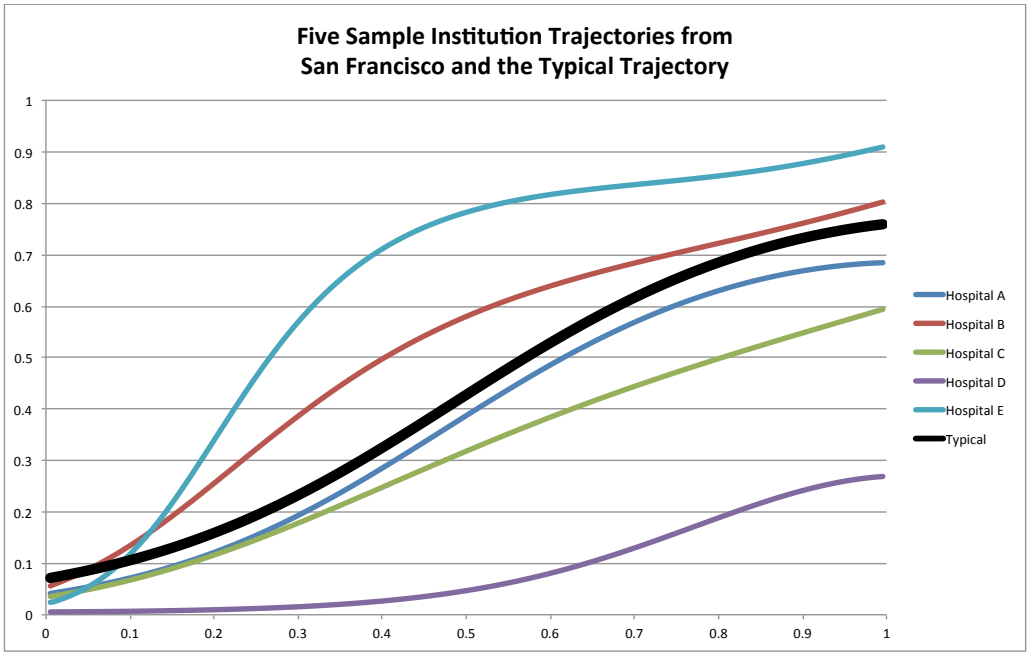

Figure 2.4: Trajectories of Sample Hospitals in San Francisco SEER Region

2.4 Discussion

The primary finding is that the Innovation Adoption Trajectories of SLNB do vary by institution and region as indicated by the the significance of the polynomial of time in the random and fixed effects. Thus they are a potential candidate for the basis of quality metrics that measure the adoption of evidence based procedures.

The utility of this type of model in comparison to standard models is three fold: it has the capability of looking at hospital specific outcomes over time, it is more robust to missing data, and it fits considerably better. Additionally, the estimates of the parameters and p-values should be both more precise and more trustworthy based on theoretical considerations (51; 53).

Although no comparison was made to the GEE based population average model some considerations would suggest that the GLMM might be preferred. The primary factor is that the GLMM

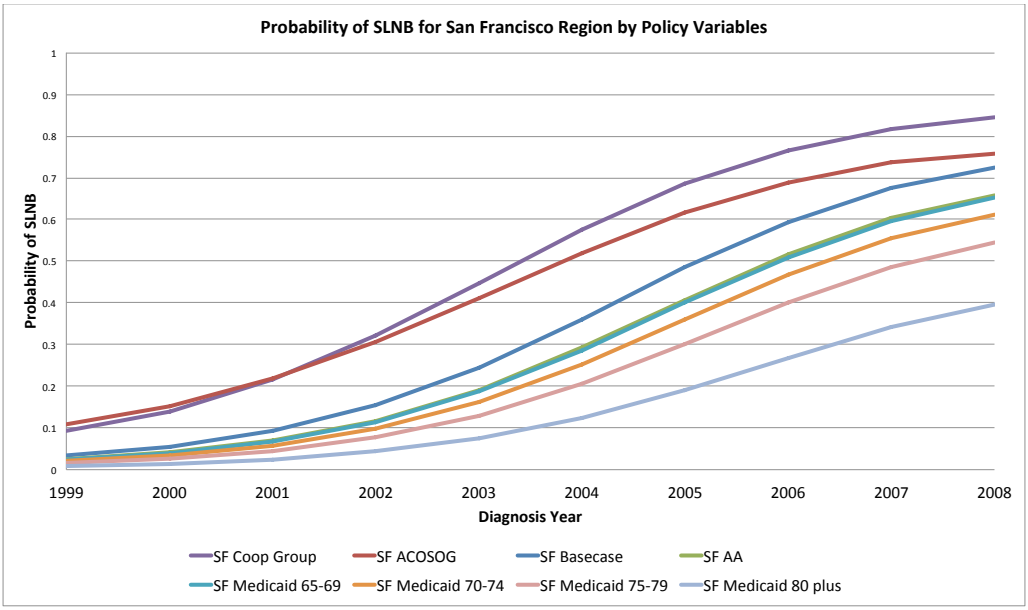

Figure 2.5: Effect of Policy Variables within San Francisco SEER Region

approach permits subject (hospital) specific estimates of the trajectories, something that is not possible with a GEE based approach. Another is that GEE based models depend on missing data being MCAR, something that may not be true of SEER-Medicare data, while the GLMM only requires MAR.

Previously reported disparities related to race and SES were found to continue throughout the study period as there was no significant interaction with time for these variables. Previously reported positive effects of NCI cooperative group affiliation were found to be present but their effect attenuated over time as indicated by the statistically significant interaction with time for these variables. It would seem that these affiliations were associated with a head start for those institutions but the other institutions caught up by the end of the study period.

Whether institutional culture is impacted by these affiliations or whether the culture of early adopter institutions drives having the affiliations can not be discerned by this associative study. It

should also be noted that the basic assumptions (1) - (4), albeit reasonable, may not be always very tenable, particularly when there are numerous covariables and when the assumption of multinormality in (3) may be questionable. Possible departures from these model assumptions may affect the p-values, particularly in the tail. This aspect needs to be studied in greater detail. Further research could include looking at whether some other random effects distribution would provide a better fitting model given the known sensitivity of the binary outcome GLMM to the random effects distribution specification (21; 18; 34; 33).

3 EVIDENCE BASED MEDICINE IMPLEMENTATION INDEX

3.1 Introduction

Clinical trial results enable evidence-based change in practice but their adoption may vary by institution and region. Innovation adopter types were originally proposed by Rogers' (1962) in his theory of the diffusion of innovation. Rogers proposed five categories of adopter types, innovators, early adopters, early majority, late majority, and laggards. He hypothesized that these types comprised 2.5, 13.5, 34, 34, and 16 percent of the population respectively. This work extends his original idea of the time until use of an innovation to a metric that characterizes the pattern of the rate of use of an innovation over time.

In order for an institutional metric to be useful it should be easy to understand, be generalizable among different conditions, control for case-mix, and allow fluctuation of caseload over time. To meet these objectives we propose a two step process. First institutional trajectories of the rate of use of an innovation are created that control for disease and patient characteristics. This is done by use of a generalized linear mixed model (GLMM). Next, the area under the curve (AUC) of each institution's trajectory is calculated. Early adopters will have larger values of this metric while those that adopt later will have smaller values.

As a test case for this metric, we considered the procedure of sentinel node biopsy (SLNB) and its adoption among hospitals participating in the SEER-Medicare registry from 1999 to 2007. Sentinel node biopsy is a surgical procedure adopted in the last 20 years as a less morbid alternative to axillary lymph node dissection (ALND) for determining lymphatic spread of tumor in early stage breast cancer. Prior to initiation of clinical trials of SLNB, standard of care for lymph node assessment required an axillary lymph node dissection (ALND) which conferred over a 20% risk of lymphedema and occasional long-term nerve injury.

The introduction of SLNB in the mid 1990's, followed by a growing body of level I and level II evidence demonstrating equivalent accuracy as ALND for pathologically node negative cases allowed SLNB to supplant ALND as the standard of care (62; 13; 52; 14; 41; 54; 55; 61; 56). Medicare started reimbursing its use in 1999 while it was still in clinical trials since early evidence suggested it was a preferred treatment option in many cases. With the conclusion of the clinical trials SLNB became the documented standard of care for lymphatic staging of early stage breast cancer (26).

We propose that variability in adoption of clinical trial results can be captured by quantifying innovation adopter type. In this work we make the assumptions that institutions have consistent cultures over time (albeit changeable) and that individual physicians both contribute to the culture and in general follow the treatment patterns of the prevailing institutional culture.

3.2 Methods

3.2.1 Data Description

The data used to generate the models of trajectories of rate of use of SLNB in indicated cases come from the NCI SEER-Medicare database, which is derived from 14 cancer registries covering 17 geographic regions across the US. Individual cases are linked to the corresponding Medicare claims and eligibility information. For those cases where the first primary cancer occurred at an age of 65 or greater there is a success rate of 94 percent in linking to Medicare data This data source is described extensively elsewhere (59; 4; 25).

The patient-level inclusion criteria are that the Medicare recipient is female with her first or only incident primary breast cancer occurring during the years of 1999 to 2007. A valid diagnosis month must be present and the reporting source must not be autopsy nor death certificate. This cancer must have occurred at age of 66 or later and the basis for Medicare coverage must not be End Stage Renal Disease (ESRD). The individual must have had both parts A and B coverage from 12 months prior to diagnosis month until 12 months post diagnosis month (or until death) without HMO coverage during this time period. This is to ensure that the claims data are available for both

inpatient and outpatient services. The inclusion criteria for this study were broad with respect to disease characteristics. This will permit examining both indicated as well as contraindicated use of SLNB over time. Definitive treatment must have occurred within 12 months of diagnosis. This is necessary in order to be able to identify the institution where a SLNB may have been performed. Patients that received neoadjuvant chemotherapy are excluded from the analysis. In total, 74,516 patients met the inclusion criteria for this study with an associated 2,004 institutions where surgery was performed.

At the hospital-level the inclusion criteria are that there must have been at least 15 cases that met the patient-level inclusion criteria treated at that hospital during the time period of the study. In order to be included in the final analysis all the hospital-level covariates needed to be present. 720 institutions were included after applying these criteria with 70,371 associated patients. Two institutions had missing hospital-level covariates and were excluded from the final hospital level analysis.

3.2.2 Use of GLMM to generate institutional trajectories

Generalized linear mixed models with a binary outcome are a standard way of generating trajectories for outcomes such as treat/no treat or use of procedure A vs. procedure B when there are either multilevel or longitudinal aspects to the data. It is an extension of logistic regression that handles cases that violate several assumptions of simple logistic regression. Given that institutions are routinely grouped by region or have registries they report to that service a particular region, as well as the measurement of use of a procedure at the patient level, it is natural to use a multilevel approach. Since we are interested in the rate of use of a procedure over time we also want to control for the repeated measures (at the institution level) aspect of the data. They also have the advantage that they handle missing or unbalanced data well given the assumption that data are missing at random. This approach, however, does require the use of data from a population of institutions, possibly from multiple regions in order to create the model.

Fixed effect terms in the multilevel longitudinal model included the SEER regions and their

interactions with time. These estimate the trajectory of the the typical hospital in each region. Additionally there were fixed effects for patient age and disease characteristics. Random effects of time at the hospital nested within SEER region level provide estimates of how a hospital's trajectory varied from the typical hospital within the SEER region. The reference group used in generating institutional trajectories consisted of T1/2, well-differentiated tumor, and aged 66 to 69.

3.2.3 AUC of trajectories

In using the AUC of trajectories to characterize adopter types we are assuming the trajectories are non-decreasing functions over the time span considered. This would imply the ordering of the AUC corresponds to the ordering of adopter type. The AUC of these model based trajectories is not simply the average rate of use of the procedure, but a function that weights all time points equally regardless of fluctuations in case load as well as controlling for case-mix.

The area under the curve of the trajectories was calculated based on the parameter estimates from the GLMM. For SEER regions this was based on the fixed effects of region and region by the cubic polynomial of time. For institutions they were based on the regional estimate plus (or minus) the institution specific random effects estimate.

3.2.4 Institution level analysis

In general we are proposing that the Evidence Based Implementation Index (EMII) be a function of a set of procedures whose AUC's are combined to form the metric. However, in this single procedure proof of concept, further analysis was performed using the EMII based upon the single procedure's AUC. Hospital level covariates such as whether it was a teaching hospital and what cooperative groups the institution was affiliated with were then examined to see if the EMII would vary by membership status. Simple descriptive statistics of the EMII were generated overall and by hospital characteristic. Additionally an ANOVA type approach was used, modeling the relationship between institutional variables and the EMII. This was done to control for the effects of SEER region and unequal cell sizes of the combinations of variables while estimating the magnitudes of effect of the hospital characteristic.

3.3 Results

3.3.1 GLMM results

All terms in the model were highly statistically significant and used in calculating the EMII.

3.3.2 Regional trajectories

Trajectories of the SEER regions are based upon the fixed effects of the GLMM and in these hospital specific models represent the trajectory of a hypothetical 'typical hospital' from the region. Figure 3.1 shows these trajectories; the difference between adoption patterns is evident with Seattle clearly being an early adoption region followed by Connecticut and then other SEER regions.

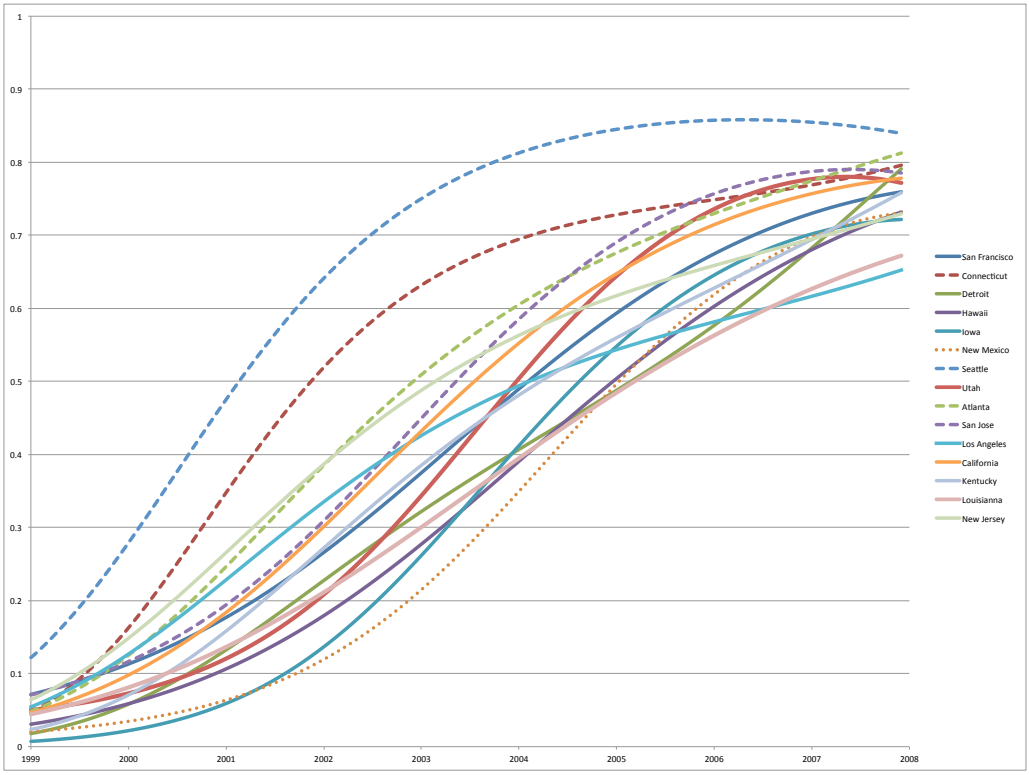

Figure 3.1: SEER Regional Trajectories of the rate of use of SLNB

3.3.3 Institutional Trajectories

Institutional trajectories are based upon the combination of the relevant fixed effects (i.e. the SEER region and its interaction with time) and the predicted random effect values for the cubic of time for that institution. Sample hospital trajectories from the San Francisco SEER region are

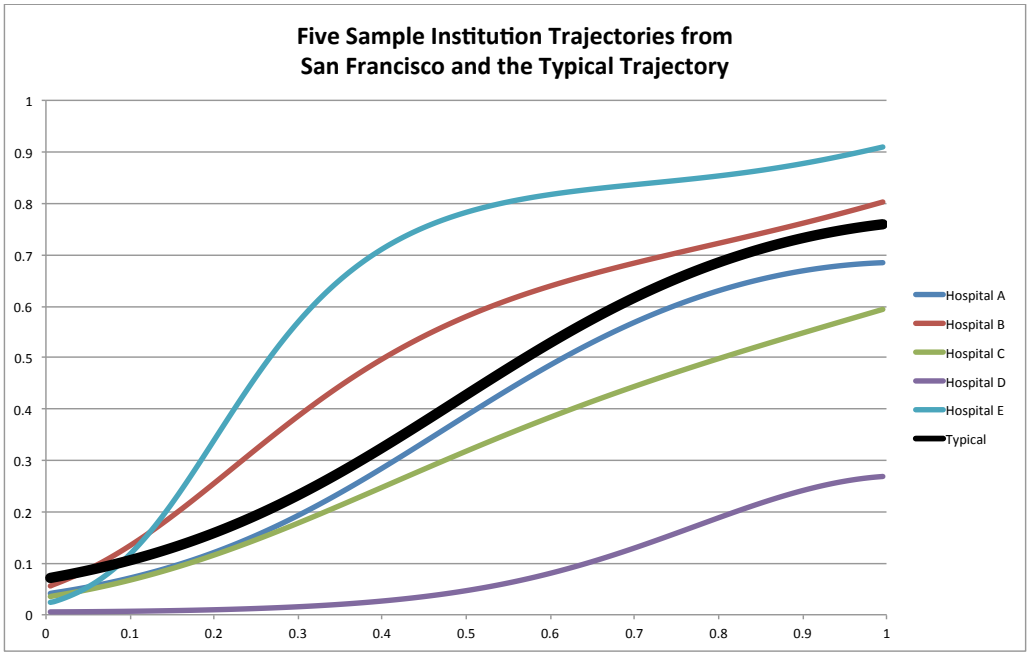

Figure 3.2: Trajectories of Sample Hospitals in San Francisco SEER Region

shown to demonstrate the variability found within a given region are shown in Figure 3.2

3.3.4 EMII

Region level values of the EMII are given in Table 3.1 in the 'typical' column. These represent the EMII of a typical hospital in the SEER region. Note that the typical value may not be the same as the mean value. The number of institutions included from each region in also given.

Table 3.1 also includes the mean, standard deviation, minimum, and maximum values at the hospital level within each SEER region and overall. Graphical examination of all 720 institutions indicates a bell shaped curve (data not shown). This would be consistent with Rogers' theory of a somewhat normal shaped distribution of adopter types. Table 3.2 gives the values for different hospital characteristics. Note that these groups are not mutually exclusive.

3.3.5 Model with EMII as outcome looking at hospital characteristics

The results of the exploratory analysis using the EMII as the outcome measure found that after controlling for regional variations there was strong evidence ($p < .01$ for the system of variables) that the variables of: ACOSOG affiliation, Teaching hospital, there being at least one Cooperative group affiliation besides ACOSOG (One Plus), the interaction of Teaching hospital and One Plus, and there being at least four cooperative group memberships other than ACOSOG (Four Plus), had significant association with the EMII. The effects of One Plus and Four Plus are additive which gives the net effect of there being three categories, no additional cooperative group memberships (the base case), one to three additional cooperative group affiliations, and four plus.

Table 3.1: Descriptive statistics of the EMII overall and by SEER region.

SEER Region	n	Typical EMII	Mean EMII	Std. Dev.	Min EMII	Max EMII
San Francisco	27	42	46	17	9	76
Connecticut	34	56	56	13	29	81
Detroit	34	36	40	18	11	73
Hawaii	14	35	38	17	13	79
Iowa	69	35	38	20	7	86
New Mexico	23	33	37	23	10	84
Seattle	33	66	66	13	41	87
Utah	24	42	45	15	17	69
Atlanta	24	49	51	16	25	81
San Jose	15	48	50	20	15	72
Los Angeles	60	41	45	20	12	77
Greater CA	152	45	48	19	4	87
Kentucky	60	40	43	18	14	82
Louisiana	64	35	39	22	7	86
New Jersey	87	47	49	18	11	83
ALL	720		46	20	4	87

Table 3.2: Descriptive statistics of the EMII by hospital characteristics

Characteristic	n	Mean EMII	Std. Dev.	Min EMII	Max EMII
No affiliations	282	37	19	4	86
ACOSOG	111	59	16	16	87
NCI Comprehensive Cancer Center	21	68	12	46	85
Teaching Hospital	284	51	19	8	87
1 to 3 Co-op Groups (other than ACOSOG)	261	51	17	10	86
4 or more Coop Groups (other than ACOSOG)	91	60	16	16	87

There was also moderate evidence (p=.07) that NCI comprehensive cancer centers had an additional effect upon the EMII beyond the effect of teaching hospital, ACOSOG, and number of cooperative group memberships. All of these institution level variables were associated with an increase in the EMII except for the interaction of Teaching Hospital and One Plus which was negative and indicated that the effect of One Plus was less for Teaching Hospitals than Non-Teaching Hospitals. Table 3.3 gives the estimated model adjusted values of the EMII for each region and their standard errors. Note that the region-level estimates correspond to the case where the institution is not a teaching hospital and has no affiliations. Table 3.4 gives the estimated effects of the institutional-level variables and their standard errors and p-values.

3.4 Discussion

We have demonstrated that the EMII differentiates uptake of new procedures (specifically, SLNB) among regions and institutions as well as between different hospitals types and institutional affiliations. While it doesn't show cause and effect, it provides a method to quantify which regions and types of institutions are early adopters of clinical trial results. This information, if made available, could be used by patients to select service providers (early adopters) as well as to drive institutional behaviors for early adoption.

Table 3.3: Estimated EMII values by region for non affiliated hospitals

SEER Region	Estimate	Std. Err.
San Francisco	35	3
Connecticut	44	3
Detroit	29	3
Hawaii	32	5
Iowa	32	2
New Mexico	32	4
Seattle	57	3
Utah	39	3
Atlanta	43	4
San Jose	42	4
Los Angeles	36	2
Greater CA	42	2
Kentucky	36	2
Louisiana	35	2
New Jersey	37	2

Table 3.4: Institutional affiliation estimated effects

Effect	Estimate	Std. Err.	p-value
ACOSOG	7	2	.0015
Teaching Hospital	5	2	.022
Plus One	12	2	$<$.0001
Teaching Hospital*Plus One	-7	3	.025
Plus Four	6	2	.013
NCI Comprehensive CC	8	4	.070

In the United States, more than $3 billion are spent in clinical trials research annually by the NIH alone, and these studies are only meaningful in their ability to change clinical practice and improve outcomes on a wide scale. Reasonable concerns that trials may yield positive results but nevertheless fail to drive practice change likely impact both current and future federal resource allocation into clinical trials. Prime examples of this may be found in prevention studies: NSABP P01 discovered almost a 50% reduction in breast cancer incidence with 5 years of tamoxifen use. However, this has not led to an increased uptake in chemoprevention, even 15 years after these

results were first reported. The reason underlying the low uptake is multifactorial, but an instrument such as the EMII could be of importance in tracking, reporting, and encouraging widespread adoption of clinical trials findings shown with a high level of evidence to likely to positively impact human health.

There are some important limitations of the study which merit discussion. First, the data we analyzed were limited to an older Medicare population so that its relevance to younger patients requires further testing. Moreover, there could be indications for and against use of the sentinel node biopsy procedure in individuals which may be difficult to capture. One example of this is in those women who were candidates for sentinel node biopsy, but in whom age and comorbidities were clinically deemed to preclude additional surgical procedures. One way in which this could be addressed is to evaluate the proportion of patients receiving axillary staging by ALND versus SLNB rather than evaluating the rate of SLNB uptake alone. However, this could create an additional confounder in that women with obviously positive nodes at diagnosis would be correctly treated with ALND instead of SLNB and thus not represent poor uptake of SLNB. Such issues will require individual consideration for each procedure measured. Nevertheless, this study provides an indication of how such a metric may be constructed, in the older Medicare population which carries the greatest cancer burden.

Further research is clearly needed to see if this approach holds across different procedures and different disease types as well as for populations other than Medicare patients. If to be used as a quality metric, benchmarks would need to be generated in order to assess where an institution is at in its process of adopting evidence based medicine. Finally, although it is assumed that the use of evidence based medicine improves outcomes, the correspondence between this metric and actual outcomes must be assessed.

4 ESTIMATION OF VARIANCE COMPONENTS

4.1 Introduction

In the previous two chapters we have seen that a Logistic Normal Mixed Model (LNMM) can be useful in characterizing the adoption trajectories of Sentinel Lymph Node Biopsy (SLNB) of hospitals serving women on Medicare with breast cancer using SEER-Medicare data. However it has been assumed that the use of Maximum Likelihood (ML) Estimators of the variance components of the random effects is preferable in contrast to pseudo-likelihood (PL) estimators. One reason for this assumption is that it has been shown that in a LNMM the PL estimator leads to biased results for the fixed effects (6; 32). ML with quadrature based estimation has been shown to have the least bias, while ML with Laplace estimation is somewhere in between (43).

One aspect that has not been considered previously is whether the bias inherent in the PL estimators could be reduced by use of a jackknife approach (16). Nor has the relative efficiency (23) of the estimators been compared in a LNMM. In order to investigate these aspects, a model employing random intercept and slope terms (and their covariance) was created with the subject being hospital nested within SEER region. It is assumed that patients are i.i.d. after controlling for all other factors and that any change over time in a hospital's use of SLNB is due to changes in physician/hospital practice rather than changes in the patient population. In order to increase precision, stratification by number of patients a hospital treated over the course of time the data covered (1999 to 2007) was performed.

Two types of PL estimators were considered, Maximum Subject Specific PL (MSPL) and Restricted Subject Specific PL (RSPL). MSPL (and it's stratified jack knifed version) was included as it's results might be more comparable to the ML estimator. RSPL was included as it is frequently

used. Random effects covariance matrix estimator efficiency was compared via A-, D-, and E- efficiency metrics which permitted multivariate comparisons in efficiency between ML and the four other estimator types, MSPL, RSPL, stratified jackknifed MSPL and stratified jackknifed RSPL. By use of the efficiency metrics we can quantify overall differences in the covariance matrices with respect to the stated precision of the estimates.

4.2 Methods

Five estimators were used to estimate the variance components of a LNMM with random intercept and slope at the second level of a three level heirarchical model. The dataset used was SEER-Medicare data which is briefly described in the next section.

4.2.1 Data

The data used to generate the models of trajectories of rate of use of SLNB in indicated cases come from the NCI SEER-Medicare database, which is derived from 14 cancer registries covering 17 geographic regions across the US. Individual cases are linked to the corresponding Medicare claims and eligibility information. This data source is described extensively elsewhere (59; 4; 25). A quick overview of SEER-Medicare data methodology follows.

Selected SEER registry data elements (those elements with known poor reliability are excluded) from each incident cancer diagnosis, up to the first ten, are put into a format of one record per person. This data is then linked to Medicare eligibility data and a 94% linkage rate has been found for cases where the first cancer occurred at an age of 65 or greater. Cases where there is a successful link are kept.

This information is combined with demographic and eligibility information from the Medicare program to create the Patient Entitlement and Diagnosis Summary File (PEDSF) file. The name is a little misleading as it includes all of the (selected by NCI) SEER data, not just the diagnosis information. It also includes (de-identified and 'fuzzed') census tract and zip code level demographic (census) information.

Fixed effect eariables used in the modeling process included SEER region, time (scaled 0 to 1

over the range of the data), SEER region by time, an indicator variable for institutional affiliation with the American College of Surgeons Oncology Group, a sponsor of a SLNB clinical trial, an indicator variable for other NCI cooperative groups having breast cancer research portfolios including; National Surgical Adjuvant Breast and Bowel Project, Cancer and Leukemia Group B, Southwest Oncology Group, and the Eastern Cooperative Oncology Group, and an indicator variable for medical school affiliation. Interactions with the linear effect of time for all hospital level fixed effects were also included.

There were three demographic person level fixed effects: Race being African American (AA), an indicator variable, with the reference group being all other races, Age, a categorical variable of patient age at diagnosis, with levels of: 66 to 69 (which is the reference group), 70 to 74, 75 to 79, and 80 plus, and Medicaid, an indicator variable for the patient being dual eligible during the year of diagnosis, a proxy for individual low income status.

Disease characteristic variables included the fixed effects of: Tumor grade, a four level categorical variable with the reference being 'well differentiated' and three indicator variables for the levels of 'poorly differentiated', 'moderately differentiated', and 'unknown or not assessed', Tumor size, we transformed the two continuous size variables (there are two due to a change in coding systems over time) into T staging categories from AJCC TNM staging. The reference category was T1c. This was the only staging variable used as it is the only constant staging data over time with changes in systems occurring during the timeframe of this study. Additionally, we do not know whether the N staging reported was clinical or pathologic, presumably for those who received nodal staging it was pathologic, while for those where there was no indication of staging it is unclear. The interaction between tumor grade and size was also included.

Treatment variables included surgery type and receipt of SLNB. Surgery was a three level categorical variable for the receipt of breast conserving surgery (BCS) and or mastectomy with the reference category being BCS only. The other two levels were mastectomy only and both BCS and mastectomy. It was hypothesized that various factors that would influence surgery type

(and subsequent surgeries) would also influence the use of SLNB. Receipt of SLNB (0,1) was the outcome variable and based upon both claims and registry data.

In the next section the specification of the LNMM is given.

4.2.2 Logistic Normal Mixed Model

In this section we will use the following specification of the Generalized Linear Mixed Model. Note that g^{-1} is the inverse link function and $\boldsymbol{Z}$ is the design matrix for the random effects.

$$E(\boldsymbol{Y}|\boldsymbol{b}) = g^{-1}(\boldsymbol{\eta})$$

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}$$

Where

$$\boldsymbol{b} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{G})$$

and

$Var(\boldsymbol{Y}|\boldsymbol{b})$ comes from the exponential family

For the LNMM:

Let N be the total number of patients observed. There are n hospitals (clusters) and i is the index for the hospitals so that $i = 1$ to n. j indexes the patients within a hospital and there are m_i patients in the i^{th} hospital, so $j = 1$ to m_i. The response vector $\mathbf{Y_i} = (Y_{i1}, \ldots, Y_{im_i})^T$ consists of zeros and ones that indicate each patients (non)receipt of SLNB.

$$\text{Error distribution} \qquad Y_{ij} \sim \text{Bern}(\pi_{ij})$$

$$\text{Logit link function} \qquad \eta_{ij} = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)$$

The next section describes the use of Pseudo-Likelihood in variance parameter estimation.

4.2.3 Variance Parameter Estimation - Pseudo-Likelihood

The use of Pseudo-likelihood for estimation of variance parameters in a GLMM was proposed in two separate papers in 1993 (5; 63). Both approaches make use of the generalized mixed model equations and iteratively solve for $\boldsymbol{\beta}$ and $\mathbf{b}$.

$$\begin{bmatrix} \mathbf{X^T W X} & \mathbf{X^T W Z} \\ \mathbf{Z^T W X} & \mathbf{Z^T W Z} + \mathbf{G^{-1}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{X^T W y^*} \\ \mathbf{Z^T W y^*} \end{bmatrix}$$

Where $\mathbf{y}^*$, the 'pseudo-data', is alternately used to estimate $\boldsymbol{\beta}$ and $\mathbf{b}$ and is estimated using their estimates from the previous iteration.

The following description of the ML estimation procedure draws heavily on the SAS documentation (47) for the procedure used to fit the models, Proc GLIMMIX (see chapter 38).

4.2.4 Variance Parameter Estimation - Maximum Likelihood with quadrature

The joint probability density function, in general, is given by:

$$f\left(\mathbf{Y_i}|\mathbf{X_i}, \mathbf{b_i}\right) f\left(\mathbf{b_i}\right)$$

but since the $\mathbf{b_i}$ are unobserved the marginal likelihood function is used:

$$\prod_{i=1}^{N} \int f\left(\mathbf{Y_i}|\mathbf{X_i}, \mathbf{b_i}\right) f\left(\mathbf{b_i}\right) d\mathbf{b_i}$$

A two step procedure is used to get the maximum likelihood estimates, first obtain estimates for $\boldsymbol{\beta}$ and $\mathbf{G}$ based on the marginal likelihood, using adaptive quadrature numerical integration for approximate estimates. Next, using these estimates generate predicted random effects values:

$$\hat{\mathbf{b}}_i = E\left(\mathbf{b_i}|\mathbf{Y_i}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{G}}\right)$$

Adaptive quadrature is implemented as follows (47), the quadrature rule is:

$$\int_{-\infty}^{\infty} f(x)\, p(x)\, dx \approx \sum_{r=1}^{Q} w_r f(x_r)$$

where $p(x)$ is a probability density function, $f(x)$ is some function to be integrated against it, Q is the number of quadrature points, r is its index, and w_r are the quadrature weights. In our case $f(x)$ is the conditional distribution given the random effects, and $p(x)$ is the random effects distribution. When the number of quadrature points is not specified ahead of time then Proc GLIMMIX determines the number of quadrature points by evaluating the log likelihood at an increasing number of points until a tolerance is met. Additionally, and separately, 'the procedure centers and scales the quadrature points by using the empirical bayes estimates (EBEs) of the random effects and the Hessian matrix from the EBE suboptimization.' The manual goes on to state that this process improves the likelihood approximation 'by placing the abscissas according to the density function of the random effects.'

The next section describes the jackknife procedure.

4.2.5 Jackknife procedure

In general, the jackknife estimator (16) is derived as follows: $\hat{\boldsymbol{\theta}}_n$(4.1) is the estimator based upon all the data, in our case $\boldsymbol{X}_k$ is the data from all the patients in the k^{th}hospital. $\hat{\boldsymbol{\theta}}^i_{n-1}$(4.2) are the estimators based upon the data less the i^{th} element, $\hat{\boldsymbol{\theta}}^i_J$(4.3) are the pseudovalues of the jackknife, and $\hat{\boldsymbol{\theta}}_J$(4.4) is the jackknife estimator.

$$\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \tag{4.1}$$

$$\hat{\boldsymbol{\theta}}^i_{n-1} = \hat{\boldsymbol{\theta}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1}, \boldsymbol{X}_{i+1}, \ldots, \boldsymbol{X}_n) \qquad (i = 1, \ldots, n) \tag{4.2}$$

$$\hat{\boldsymbol{\theta}}^i_J = n\hat{\boldsymbol{\theta}}_n - (n-1)\hat{\boldsymbol{\theta}}^i_{n-1} \qquad (i = 1, \ldots, n) \tag{4.3}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_J &= \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\theta}}_J^i \\ &= n\hat{\boldsymbol{\theta}}_n - (n-1)\bar{\hat{\boldsymbol{\theta}}}_{n-1}^i\end{aligned} \tag{4.4}$$

In the next section the stratification of the data that was done is described.

4.2.6 Stratified sampling

In order to reduce bias and to provide estimates for different size hospitals we have subdivided the data into three strata. The strata are: small institutions with 1 to 49 patients in the data, medium institutions with 50 to 199 patients in the data, and large institutions with 200 or more patients in the data. Respectively there are s, m, and l institutions in the strata and they are indexed by o, p, and q. Thus we have $\hat{\boldsymbol{\theta}}_{Js}, \hat{\boldsymbol{\theta}}_{Jm}$, and $\hat{\boldsymbol{\theta}}_{Jl}$ (4.5) as our jackknife estimators for the three strata. We then use a weighting function to combine these three into an overall estimate.

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{Js} &= \frac{1}{s}\sum_{o=1}^{s}\hat{\boldsymbol{\theta}}_{Js}^o \\ \hat{\boldsymbol{\theta}}_{Jm} &= \frac{1}{m}\sum_{p=1}^{m}\hat{\boldsymbol{\theta}}_{Jm}^p \\ \hat{\boldsymbol{\theta}}_{Jl} &= \frac{1}{l}\sum_{q=1}^{l}\hat{\boldsymbol{\theta}}_{Jl}^q\end{aligned} \tag{4.5}$$

In the next section the derivation of the jackknifed asymptotic covariance matrix is considered.

4.2.7 Jackknifing the asymptotic covariance matrix

At this point we note that there are two ways to derive the jackknifed asymptotic covariance matrix. The first is to define $\hat{\boldsymbol{\theta}}$ as the estimator of the covariance matrix and calculate $\boldsymbol{v}_J$(4.6) using a function of the pseudovalues and the jackknifed estimator.

$$\boldsymbol{v}_J = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{\boldsymbol{\theta}}_J^i - \hat{\boldsymbol{\theta}}_J\right)\left(\hat{\boldsymbol{\theta}}_J^i - \hat{\boldsymbol{\theta}}_J\right)^{\mathsf{T}} \tag{4.6}$$

The other is to define $\hat{\boldsymbol{\theta}}$ as the estimator of the asymptotic covariance matrix,$\hat{\boldsymbol{v}}$,and get the jackknifed estimator directly. This however calls for a modication of (4.3) and thus (4.4) to (4.8) and (4.9) via the transformations in (4.7). Note that r in (4.7) is the number of fixed effect parameters in the model and that we are essentially treating $\hat{\boldsymbol{v}}_n$ as the MSE in a regression model while $\hat{\boldsymbol{v}}_n^*$ is the corresponding SSE. Finally we must put the jackknifed estimate back on the original scale (4.10).

$$\begin{aligned}\hat{\boldsymbol{v}}_n &= \frac{1}{n-r}\hat{\boldsymbol{v}}_n^* \\ \hat{\boldsymbol{v}}_{n-1}^i &= \frac{1}{n-r-1}\hat{\boldsymbol{v}}_{n-1}^{*i}\end{aligned} \tag{4.7}$$

$$\begin{aligned}\hat{\boldsymbol{v}}_J^{*i} &= (n-r)\,\hat{\boldsymbol{v}}_n^* - (n-r-1)\hat{\boldsymbol{v}}_{n-1}^{*i} \qquad (i=1,\ldots,n) \\ &= (n-r)^2\,\hat{\boldsymbol{v}}_n - (n-r-1)^2\hat{\boldsymbol{v}}_{n-1}^i\end{aligned} \tag{4.8}$$

$$\begin{aligned}\hat{\boldsymbol{v}}_J^* &= \frac{1}{n}\sum_{i=1}^{n}\hat{\boldsymbol{v}}_J^{*i} \\ &= (n-r)^2\,\hat{\boldsymbol{v}}_n - (n-r-1)^2\bar{\hat{\boldsymbol{v}}}_{n-1}^i\end{aligned} \tag{4.9}$$

$$\hat{\boldsymbol{v}}_{J^*} = \frac{1}{n-r}\hat{\boldsymbol{v}}_J^* \tag{4.10}$$

Having done this for each of the three strata, we obtain a pooled optimally weighted estimate $\hat{\boldsymbol{v}}_{JP}$ where the strata are combined (4.11).

$$\hat{\boldsymbol{v}}_{JP} = \left(\hat{\boldsymbol{v}}_{Js}^{-1} + \hat{\boldsymbol{v}}_{Jm}^{-1} + \hat{\boldsymbol{v}}_{Jl}^{-1}\right)^{-1} \tag{4.11}$$

The next section reviews estimation of multivariate relative efficiency.

4.2.8 Efficiency of Estimators

This section draws on section 8.6 of Jureckova, Sen, and Picek (2011) (23) where they discuss multivariate efficiency. In all three types of relative efficiency we will consider the reference is the ML estimator and it is the fisher information from that which is compared to the asymptotic covariance matrix of the estimated random effects variance terms from the estimator which is under consideration. Essentially each type of efficiency is a different statistic of the eigenvalues of the product of asymptotic covariance matrix and the MLE fishers information. The specifics follow.

Let $\boldsymbol{T}_n$ be the estimator of interest of parameter $\boldsymbol{\theta}$, in our case the covariance matrix of the random effects, and let $\boldsymbol{v}_T$ be the dispersion matrix of $\boldsymbol{T}_n - \boldsymbol{\theta}$. Let $\mathcal{I}(\boldsymbol{\theta})$ be the Fisher information matrix and $\boldsymbol{D}^0$ is the diagonal matrix of the p eigenvalues of $\boldsymbol{v}_T\mathcal{I}(\boldsymbol{\theta})$.

Then we have:

D-efficiency

is the p^{th} root of the determinant of $(\boldsymbol{D}^0)^{-1}$ and is the geometric mean of the eigenvalues.

A-efficiency

is the mean of the eigenvalues of $(\boldsymbol{D}^0)^{-1}$ and is the arithmatic mean. It will always be equal to or greater than the value for the D-efficiency.

E-efficiency

is the largest eigenvalue of $(\boldsymbol{D}^0)^{-1}$ and will always be larger than both the D- and A-efficiency values.

4.2.9 Relative efficiencies to be calculated

The two types of pseudoliklihood we will be considering are Maximum Subject Specific Pseudolikelihood (MSPL) and Restricted Subject Specific Pseudolikelihood (RSPL). We will caclulate both jackknifed and unjackknifed versions of their D, A, and E relative efficiencies. Note that for both the non-jackknifed and the jackknifed estimates we use a pooled Fisher's information matrix (4.12).

$$\mathcal{I}(\boldsymbol{\theta})_P = \mathcal{I}(\boldsymbol{\theta})_s + \mathcal{I}(\boldsymbol{\theta})_m + \mathcal{I}(\boldsymbol{\theta})_l \tag{4.12}$$

Thus we end up with four sets of efficiencies stemming from: $\mathbf{\Lambda}_{MSPL}, \mathbf{\Lambda}_{RSPL}$ which are based on the non-jackknifed and pooled constituent elements, and $\mathbf{\Lambda}_{J-MSPL}, \mathbf{\Lambda}_{J-RSPL}$ which are based upon the jackknifed and pooled estimates.

4.3 Results

In the following two sections the estimates for the the values of the covariance matrix for each estimation type and the relative efficiencies will be given. We start with the covariance matrix parameter estimates for both the overall models as well as the stratified and jack knifed models.

4.3.1 Parameter Estimates

Table 4.1 gives the variance and covariance estimates for each of the estimation methods: Maximum Likelihood (ML), Maximum Subject Specific Pseudolikelihood (MSPL), Restricted Subject Specific Pseudolikelihood (RSPL), RSPL jack knife (JK), and MSPL jack knife. The standard errors are also provided for the non-jack knife estimation methods. The jack knife based estimates stem from the hospital group averages weighted by the number of hospitals in the group.

The intercept variance term estimates range from 1.98 to 2.49. The covariance term estimates range from -1.36 to -1.69. The slope variance term estimates range from 2.05 to 2.41. For the

intercept and the covariance estimates ML had the largest absolute values, while MSPL JK had the smallest absolute values. For the slope estimate RSPL JK had the largest estimate while MSPL had the smallest estimate.

Table 4.1: Covariance parameter estimates

Variable	ML	Std Err	MSPL	Std Err	RSPL	Std Err	RSPL JK	MSPL JK
Intercept	2.49	0.21	2.04	0.15	2.11	0.16	2.13	1.98
Covariance	-1.69	0.20	-1.45	0.16	-1.51	0.16	-1.53	-1.36
Slope	2.28	0.23	2.05	0.20	2.17	0.22	2.41	2.12

Table 4.2 gives the variance and covariance estimates for each of the estimation methods broken out by hospital group (strata). The number of hospitals (M) and patients (n) are given for each group. For ML there was a downward trend in the absolute value of the parameter estimates going from the hospitals with the fewest patients in the data to the hospitals with the most patients in the data. The two PL estimators had the smallest parameter estimates in the hospital group with the most number of patients with the midsize group having the highest values. Comparing the ML estimates to the PL estimates we find that in the '1 to 49' group the ML estimates were noticeably larger than the PL estimates. In the '50 to 199' group the estimates were similar across the five estimator types. In the '200+' group the ML and the MSPL estimates were similar while the RSPL estimates were somewhat higher. The combined estimates from the three strata shown in table 4.4 are based upon a weighting scheme using the inverse of the variance for each group and the number of hospitals.

Table 4.2: Covariance parameter estimates by hospital group

Group	M	n	Variable	ML	Std Err	MSPL	Std Err	RSPL	Std Err
1 to 49	1624	12738	Intercept	3.11	0.47	2.01	0.26	2.15	0.28
			Covariance	-2.04	0.52	-1.33	0.33	-1.49	0.35
			Slope	3.49	0.75	2.16	0.50	2.46	0.54
50 to 199	323	33291	Intercept	2.35	0.26	2.15	0.23	2.34	0.26
			Covariance	-1.85	0.26	-1.69	0.23	-1.87	0.26
			Slope	2.37	0.31	2.20	0.29	2.46	0.32
200+	99	30441	Intercept	0.89	0.15	0.87	0.14	1.10	0.20
			Covariance	-0.81	0.17	-0.79	0.16	-1.02	0.22
			Slope	1.18	0.22	1.15	0.22	1.51	0.30

Table 4.3: Asymptotic covariance matrices by hospital group

Group	Variable	MLE			MSPL			RSPL		
1 to 49	Intercept	0.217	-0.208	0.180	0.070	-0.075	0.076	0.077	-0.084	0.086
	Covariance	-0.208	0.268	-0.311	-0.075	0.110	-0.144	-0.084	0.124	-0.164
	Slope	0.180	-0.311	0.567	0.076	-0.144	0.255	0.086	-0.164	0.291
50 to 199	Intercept	0.068	-0.060	0.051	0.054	-0.048	0.041	0.066	-0.059	0.051
	Covariance	-0.060	0.067	-0.070	-0.048	0.055	-0.058	-0.059	0.067	-0.072
	Slope	0.051	-0.070	0.098	0.041	-0.058	0.083	0.051	-0.072	0.103
200+	Intercept	0.022	-0.022	0.022	0.021	-0.020	0.020	0.039	-0.038	0.038
	Covariance	-0.022	0.027	-0.033	-0.020	0.026	-0.031	-0.038	0.048	-0.058
	Slope	0.022	-0.033	0.050	0.020	-0.031	0.048	0.038	-0.058	0.089

Table 4.4: Covariance parameter estimates combined across groups

Variable	ML	MSPL	RSPL
Intercept	2.30	1.87	2.09
Covariance	-1.68	-1.33	-1.53
Time	2.47	2.01	2.35

4.3.2 D-, A-, and E-Efficiency Estimates for MLE vs. MSPL and RSPL

Table 4.5 shows the D-, A-, and E-efficiencies for the four pseudolikelihood based estimators relative to Maximum Likelihood. The Jackknifed RSPL values are the lowest, ranging from 0.68 to 0.75, followed by the RSPL values ranging from 0.79 to 0.86, and the two non-jack knifed versions being noticeably greater then the respective jack knifed estimators.

Table 4.5: D-, A-, and E-efficiency estimates

Type	RSPL	MSPL	J-RSPL	J-MSPL
D	0.792	1.190	0.682	0.839
A	0.793	1.191	0.683	0.843
E	0.860	1.270	0.754	0.912

Given the similarity of the D and A efficiencies an example case seems in order. Looking at the MSPL eigenvalues we have (1.135, 1.167, and 1.270). Plugging these values into the respective formulae for the three efficiency types as shown in section 4.2.8 we get the results as seen in Table 4.5.

4.4 Discussion

In this study we considered five different estimators of the covariance parameters in a LNMM. We looked at the overall estimates of the parameters, the stratified estimates, the combined stratified estimates and their relative efficiencies. While there were some differences across the overall estimates the most striking differences were found across the three strata. These differences were both in the parameter estimates and in the standard errors of the estimates. The differences in the parameter estimates across strata indicate that the larger institutions were more similar to each other

than the smaller institutions. The differences in the standard errors, particularly for the smaller institutions, across estimator types seem to indicate that the PL estimators underestimate the width of the confidence interval for the parameters. Taken together these two sets of results strongly suggest that stratification by institution size in addition to using maximum likelihood is useful for promoting both precision in the variance estimates and precision in the confidence intervals for these estimates.

The results of the relative efficiencies seem to correspond to the relationship between the standard errors as seen in table 4.2. That is the standard errors are consistently smaller for MSPL than ML while the standard errors for RSPL are larger than ML for one of the strata (200+). Given that ML is more reliable than PL this would seem to suggest that besides the parameter estimates being biased for PL that the confidence limits are biased as well. For MSPL for this data the confidence intervals would be too narrow regardless of strata, while the bias seems to be highly variable, both too wide and too narrow, for RSPL.

With respect to whether jackknifing would improve the PL estimates in some fashion it seems that the answer is, partially. As far as parameter estimate bias it does seem to reduce the difference (for the pooled estimates) between the ML and PL estimates. It also seems that the jackknifing procedure has the effect of increasing the standard errors for both types of PL. While this is a good thing for the MSPL standard error estimates it increases the bias for RSPL. It should be noted that the D-efficiencies are actually slightly lower than the A-efficiencies (as would be expected) but the difference is out at the third decimal.

These results overall strengthen the argument that ML should be used whenever possible when using a LNMM and also highlight that stratification should be considered when analyzing SEER-Medicare data.

5 SUMMARY AND CONCLUSION

In the last three chapters we have looked at: creating a quality metric based on a logistic normal model that characterizes adopter type, quantifing the adoption of innovation by institutions and regions by use of the area under the curve, and considered the effect of estimator type on the random effect variance components of the model. Here we recap the findings of chapters two through four as well as suggest some implications of these findings.

In chapter two we considered whether the proposed logistic normal model with hospital level random effects of a polynomial of time fit the data any better than models previously used in studies of the diffusion of SLNB. We also considered whether the trajectories based upon this model could be used as the basis of a metric that characterized adopter of innovation type. For both of these questions the answer is yes. As demonstrated by the forward model fitting process, which essentially recapitulated the type of models used over time, it was found that the random effects model fit the data better and that hospital level random effects of a cubic of time fit the data best among those models considered. It was also found that this model differentiated between: regions, hospital within regions, and covariates of interest, thus potentially could be used as the basis for a metric that measured adopter type.

Additionally, for those interested in more fully understanding the impact of the variables of interest, treating time as a continuous variable permits the simple examination, via a single interaction effect, of changes over time in the magnitude of the effect of a factor associated with the use of SLNB. For example, it was found that the magnitude of the effect of ACOSOG membership diminished over time. At the beginning of the time period examined in the analysis the effect of ACOSOG was rather large, being a ACOSOG member was strongly associated with greater rate of use of SLNB early in it's adoption period. Towards the end of the time period examined that

effect had diminished substantially. Likewise, there was some evidence (p=0.076) that teaching hospital affiliation had it's somewhat smaller effect completely disappear by the end of the study period. Being a member of one of the four cooperative groups considered had a slightly smaller effect than ACOSOG at the beginning but ended up having a larger effect at the end. Although, it's effect possibly (p=0.060) saw some attenuation.

In chapter three, building upon the logistic normal model developed in chapter two, a simple metric was proposed to quantify adopter type, the area under the curve of the trajectories generated by the logistic normal model. We called this metric the evidence based medicine implementation index (EMII). Using a logistic normal model that did not include hospital characteristics, the EMII was calculated for each hospital. To verify that the EMII of the hospitals did in fact differentiate between the categories of hospital characteristics, an additional model that used the EMII as the outcome and hospital characteristics as the independent variables showed that the EMII did vary consistently by hospital characteristic.

In chapter four we looked at whether the type of estimator of the random effect variance component affected either the parameter estimates or the estimate of the uncertainty of the parameter estimates (the asymptotic covariance matrix). Maximum likelihood was used as the reference estimator type against which several forms of pseudo-likelihood based estimators were compared. When looking at different sizes of hospitals it was found that the number of hospitals in the strata was associated with considerable variation in the standard error of the RSPL estimator. This suggests that a varying amount of bias in the standard errors of the variance parameter estimates is present for RSPL in comparison to maximum likelihood. For MSPL the bias was consistently in the direction of too narrow a confidence interval. This supports the preference of using maximum likelihood based estimation instead of pseudo-likelihood when possible.

One additional aspect of the data that was discovered, but not discussed in chapter four, is that the magnitude (and even significance) of the fixed effects seem to vary with the hospital size strata. For those interested in policy this might be an interesting aspect to explore. Previous work has

looked at the effect of hospital size with mixed results, but only the main effect and in some papers it was not found to be significant. The different fixed effect parameter values over the strata of hospital size would indicate that the interaction of hospital size with the other factors of interest should be considered.

Other possibilities for further research include: modeling other procedures whose adoption is of interest, investigating creation of a composite metric that is based upon multiple procedures, looking at whether the EMII, in conjunction with other quality indicators, predicts outcomes, patient satisfaction, or costs, and methodological topics such as the use of distributions other than normal in model construction.

A APPENDIX A

A.1 Data Description

Here is an overview of the data. Additional background material describing the structure of the SEER-Medicare database and the variables in each file can be found on the SEER-Medicare web site: http://healthcaredelivery.cancer.gov/seermedicare/

Contained in Patient Entitlement and Diagnosis Summary File (PEDSF)

- SEER Region
- Reporting Source
- Diagnosis Month and Year
- Age at diagnosis
- Diagnostic data
- Staging data
- Treatment data
- Eligibility data by month including HMO and medicaid status
- Demographics from SEER registries
- Demographics from CMS

Medicare data files

- Hospital inpatient stay file (MEDPAR) - one record per hospital stay
- Hospital outpatient claims file (OUTSAF) - one record per billable item
- Non-hospital provider claims file (NCH) - one record per billable item
- Durable Medical Equipment file (DME) - includes cancer medications

Hospital data (HOSPITAL) provided by NCI for 1996, 1998, and 2000 to 2006

- Healthcare cost report data (HCRIS)

Provider of Service data (POS)

Affiliations in 2002 and 2005

Derived variables:

Time is a scaled zero to one function of diagnosis month and the time span covered by this study.

Indicator variable for receipt of SLNB based on claims data and SEER registry data

Claims from OUTSAF and NCH files

From diagnosis month to 30 days post first surgery date

HCPCS codes of 38792 or 78195.

SEER variables used depended on year of diagnosis

1999 to 2002 SLNB was identified by a 1 or a 3 value of the variable sxscop1

2003 to 2007 a 2, 6, or 7 value of the sxscof1 variable indicated SLNB

Surgery type (BCS, Mastectomy, both) obtained from MEDPAR and OUTSAF claims

Hospital level variables for institution where first definitive surgery performed

Teaching hospital

NCI designated cancer center in 2002

Member of Cooperative group in 2002 (SWOG, CALGB, NSABP, or ECOG)

Member of ACOSOG 2002

A.2 Calculating Area Under the Curve in Chapter three

The AUC was calculated by use of an approximation that calculated the probability of receipt of SLNB for each of the 108 months in the study (using the midpoint). These probabilities were a function of the fixed effects plus (for individual hospitals) the best linear unbiased predictors. These probabilities were then summed and divided by 108. SAS macro is available upon request.

REFERENCES

[1] ABAD, A. A., LITIÈRE, S., AND MOLENBERGHS, G. Testing for misspecification in generalized linear mixed models. *Biostatistics 11*, 4 (October 01 2010), 771–786.

[2] AITKIN, M. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics 55*, 1 (Mar 1999), 117–128.

[3] ARRINGTON, A. K., KRUPER, L., VITO, C., YIM, J., KIM, J., AND CHEN, S. L. Rural and urban disparities in the evolution of sentinel lymph node utilization in breast cancer. *American Journal of Surgery 206*, 5 (Nov 2013), 674–681.

[4] BACH, P. B., GUADAGNOLI, E., SCHRAG, D., SCHUSSLER, N., AND WARREN, J. L. Patient demographic and socioeconomic characteristics in the seer-medicare database: Applications and limitations. *Medical care 40*, 8, Supplement: Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug. 2002), IV19–IV25.

[5] BRESLOW, N. E., AND CLAYTON, D. G. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*, 421 (1993), pp. 9–25.

[6] BRESLOW, N. E., AND LIN, X. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika 81* (1995), 81–91.

[7] BUONACCORSI, J. P. *Measurement error*. CRC Press, Boca Raton, 2010.

[8] BURAK, W. E., HOLLENBECK, S. T., ZERVOS, E. E., HOCK, K. L., KEMP, L. C., AND YOUNG, D. C. Sentinel lymph node biopsy results in less postoperative morbidity compared with axillary lymph node dissection for breast cancer. *The American Journal of Surgery 183*, 1 (1 2002), 23–27.

[9] CARPENTER, W. R., REEDER-HAYES, K., BAINBRIDGE, J., MEYER, A. M., AMOS, K. D., WEINER, B. J., AND GODLEY, P. A. The role of organizational affiliations and research networks in the diffusion of breast cancer treatment innovation. *Medical care 49*, 2 (Feb 2011), 172–179.

[10] CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A., AND CRAINICEANU, C. M. *Measurement error in nonlinear models: a modern perspective*, 2 ed., vol. 105. Chapman & Hall/CRC, Boca Raton, FL, 2006.

[11] CHEN, A. Y., HALPERN, M. T., SCHRAG, N. M., STEWART, A., LEITCH, M., AND WARD, E. Disparities and trends in sentinel lymph node biopsy among early-stage breast cancer patients (1998-2005). *J.Natl.Cancer Inst. 100*, 7 (April 2 2008), 462–474.

[12] CHENG, J., EDWARDS, L. J., MALDONADO-MOLINA, M. M., KOMRO, K. A., AND

MULLER, K. E. Real longitudinal data analysis for real people: building a good enough mixed model. *Statistics in medicine 29*, 4 (Feb 20 2010), 504–520.

[13] CLASSE, J. M., CURTET, C., CAMPION, L., ROUSSEAU, C., FICHE, M., SAGAN, C., RESCHE, I., PIOUD, R., ANDRIEUX, N., AND DRAVET, F. Learning curve for the detection of axillary sentinel lymph node in breast cancer. *European Journal of Surgical Oncology 29*, 5 (6 2003), 426–433.

[14] DOMENECH, A., BENITEZ, A., BAJEN, M. T., PLA, M. J., GIL, M., AND MARTIN-COMIN, J. Patients with breast cancer and negative sentinel lymph node biopsy without additional axillary lymph node dissection: a follow-up study of up to 5 years. *Oncology 72*, 1-2 (2007), 27–32.

[15] EDGE, S. B., NILAND, J. C., BOOKMAN, M. A., THERIAULT, R. L., OTTESEN, R., LEPISTO, E., AND WEEKS, J. C. Emergence of sentinel node biopsy in breast cancer as standard-of-care in academic comprehensive cancer centers. *Journal of the National Cancer Institute 95*, 20 (Oct 15 2003), 1514–1521.

[16] EFRON, B. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, 1982.

[17] FITZMAURICE, G. M., DAVIDIAN, M., VERBEKE, G., AND MOLENBERGHS, G. *Longitudinal data analysis*. CRC Press, Boca Raton, 2009.

[18] FITZMAURICE, G. M., LIPSITZ, S. R., AND IBRAHIM, J. G. A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics 63*, 3 (2007), 942–946.

[19] GIPPONI, M., BASSETTI, C., CANAVESE, G., CATTURICH, A., SOMMA, C. D., VECCHIO, C., NICOLO, G., SCHENONE, F., TOMEI, D., AND CAFIERO, F. Sentinel lymph node as a new marker for therapeutic planning in breast cancer patients. *Journal of surgical oncology 85*, 3 (Mar 2004), 102–111. LR: 20041117; JID: 0222643; RF: 57; ppublish.

[20] GIULIANO, A. E., DALE, P. S., TURNER, R. R., MORTON, D. L., EVANS, S. W., AND KRASNE, D. L. Improved axillary staging of breast cancer with sentinel lymphadenectomy. *Annals of Surgery 222*, 3 (Sep 1995), 394–9; discussion 399–401.

[21] HEAGERTY, P. J., AND KURLAND, B. F. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika 88*, 4 (2001), pp. 973–985.

[22] HUTCHINSON, J. R., CHAGPAR, A. B., SCOGGINS, C. R., II, R. C. G. M., CARLSON, D. J., LAIDLEY, A. L., EL-EID, S. E., MCGLOTHIN, T. Q., NOYES, R. D., LEY, P. B., TUTTLE, T. M., AND MCMASTERS, K. M. Surgeon and community factors affecting breast cancer sentinel lymph node biopsy. *The American Journal of Surgery 190*, 6 (12 2005), 915–919.

[23] JURECKOVA, J., SEN, P. K., AND PICEK, J. *Methodology in robust and nonparametric statistics*. CRC Press, Boca Raton, 2013.

[24] KANE, J. M., EDGE, S. B., WINSTON, J. S., WATROBA, N., AND HURD, T. C. Intraoperative pathologic evaluation of a breast cancer sentinel lymph node biopsy as a determinant for synchronous axillary lymph node dissection. *Annals of Surgical Oncology 8*, 4 (05/01 2001), 361–367.

[25] KLABUNDE, C. N., WARREN, J. L., AND LEGLER, J. M. Assessing comorbidity using claims data: An overview. *Medical care 40*, 8, Supplement: Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug. 2002), IV26–IV35.

[26] KRAG, D. N., ANDERSON, S. J., JULIAN, T. B., BROWN, A. M., HARLOW, S. P., COSTANTINO, J. P., ASHIKAGA, T., WEAVER, D. L., MAMOUNAS, E. P., JALOVEC, L. M., FRAZIER, T. G., NOYES, R. D., ROBIDOUX, A., SCARTH, H. M., AND WOLMARK, N. Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the nsabp b-32 randomised phase 3 trial. *The lancet oncology 11*, 10 (Oct 2010), 927–933.

[27] LAIRD, N. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association 73*, 364 (1978), pp. 805–811.

[28] LAIRD, N. M., AND WARE, J. H. Random-effects models for longitudinal data. *Biometrics 38*, 4 (12/01 1982), 963–974.

[29] LI, H., AND WANG, L. A consistent simulation-based estimator in generalized linear mixed models. *Journal of Statistical Computation and Simulation* (06/30; 2012/07 2011), 1–19.

[30] LI, H., AND WANG, L. Consistent estimation in generalized linear mixed models with measurement error. *J Biomet Biostat S7*, 007 (2012).

[31] LIANG, K.-Y., AND ZEGER, S. L. Longitudinal data analysis using generalized linear models. *Biometrika 73*, 1 (April 01 1986), 13–22.

[32] LIN, X., AND BRESLOW, N. E. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association 91* (1996), 1007–1116.

[33] LITIÈRE, S., ALONSO, A., AND MOLENBERGHS, G. Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics 63*, 4 (2007), 1038–1044.

[34] LITIÈRE, S., ALONSO, A., AND MOLENBERGHS, G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine 27*, 16 (2008), 3125–3144.

[35] MAGGARD, M. A., LANE, K. E., O'CONNELL, J. B., NANYAKKARA, D. D., AND KO, C. Y. Beyond the clinical trials: How often is sentinel lymph node dissection performed for breast cancer? *Annals of Surgical Oncology 12*, 1 (01/01 2005), 41–47.

[36] MARTIN, R. C. G., EDWARDS, M. J., WONG, S. L., TUTTLE, T. M., CARLSON, D. J., BROWN, C. M., NOYES, R. D., GLASER, R. L., VENNEKOTTER, D. J., TURK, P. S., TATE, P. S., SARDI, A., CERRITO, P. B., MCMASTERS, K. M., AND FOR THE UNIVERSITY OF LOUISVILLE BREAST CANCER STUDY GROUP. Practical guidelines for optimal gamma probe detection of sentinel lymph nodes in breast cancer: Results of a multi-institutional study. *Surgery 128*, 2 (8 2000), 139–144.

[37] MCCULLAGH, P., AND NELDER, J. A. *Generalized Linear Models*, second ed., vol. 37. Chapman & Hall, 1989.

[38] MEYER, A. M., REEDER-HAYES, K. E., LIU, H., WHEELER, S. B., PENN, D., WEINER, B. J., AND CARPENTER, W. R. Differential receipt of sentinel lymph node biopsy within practice-based research networks. *Medical care 51*, 9 (Sep 2013), 812–818.

[39] NATIONAL CANCER POLICY BOARD AND INSTITUTE OF MEDICINE AND NATIONAL RESEARCH COUNCIL. *Ensuring Quality Cancer Care*. The National Academies Press, 1999.

[40] NELDER, J. A., AND WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society.Series A (General) 135*, 3 (1972), pp. 370–384.

[41] OLLILA, D. W., BRENNAN, M. B., AND GIULIANO, A. E. The role of intraoperative lymphatic mapping and sentinel lymphadenectomy in the management of patients with breast cancer. *Advances in Surgery 32* (1999), 349–364.

[42] ORAL, E. Binary regression with stochastic covariates. *Communications in Statistics: Theory and Methods 35*, 8 (2006), 1429–1447.

[43] PINHEIRO, J. C., AND CHAO, E. C. Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics 15* (2006), 58–81.

[44] REEDER-HAYES, K. E., BAINBRIDGE, J., MEYER, A. M., AMOS, K. D., WEINER, B. J., GODLEY, P. A., AND CARPENTER, W. R. Race and age disparities in receipt of sentinel lymph node biopsy for early-stage breast cancer. *Breast cancer research and treatment 128*, 3 (Aug 2011), 863–871.

[45] RESCIGNO, J., ZAMPELL, J., AND AXELROD, D. Patterns of axillary surgical care for

breast cancer in the era of sentinel lymph node biopsy. *Annals of Surgical Oncology 16*, 3 (03/01 2009), 687–696.

[46] RUBIN, D. B. Inference and missing data. *Biometrika 63*, 3 (1976), pp. 581–592.

[47] SAS INSTITUTE INC. *SAS/STAT 9.2 User's guide.* SAS Institute Inc., 2008.

[48] SCHRAG, D., BACH, P. B., DAHLMAN, C., AND WARREN, J. L. Identifying and measuring hospital characteristics using the seer-medicare data and other claims-based sources. *Medical care 40*, 8, Supplement: Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug. 2002), IV96–IV103.

[49] STIRATELLI, R., LAIRD, N., AND WARE, J. H. Random-effects models for serial observations with binary response. *Biometrics 40*, 4 (1984), pp. 961–971.

[50] TCHETGEN, E. J., AND COULL, B. A. A diagnostic test for the mixing distribution in a generalised linear mixed model. *Biometrika 93*, 4 (2006), pp. 1003–1010.

[51] TEN HAVE, T. R., KUNSELMAN, A. R., AND TRAN, L. A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in medicine 18*, 8 (1999), 947–960.

[52] TEW, K., IRWIG, L., MATTHEWS, A., CROWE, P., AND MACASKILL, P. Meta-analysis of sentinel node imprint cytology in breast cancer. *British Journal of Surgery 92*, 9 (2005), 1068–1080.

[53] URBACH, D. R., AND AUSTIN, P. C. Conventional models overestimate the statistical significance of volume-outcome associations, compared with multilevel models. *Journal of clinical epidemiology 58*, 4 (/4 2005), 391–400.

[54] VERONESI, U., PAGANELLI, G., VIALE, G., GALIMBERTI, V., LUINI, A., ZURRIDA, S., ROBERTSON, C., SACCHINI, V., VERONESI, P., ORVIETO, E., CICCO, C. D., INTRA, M., TOSI, G., AND SCARPA, D. Sentinel lymph node biopsy and axillary dissection in breast cancer: results in a large series. *Journal of the National Cancer Institute 91*, 4 (Feb 17 1999), 368–373.

[55] VERONESI, U., PAGANELLI, G., VIALE, G., LUINI, A., ZURRIDA, S., GALIMBERTI, V., INTRA, M., VERONESI, P., ROBERTSON, C., MAISONNEUVE, P., RENNE, G., CICCO, C. D., LUCIA, F. D., AND GENNARI, R. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *The New England journal of medicine 349*, 6 (Aug 7 2003), 546–553.

[56] VERONESI, U., VIALE, G., PAGANELLI, G., ZURRIDA, S., LUINI, A., GALIMBERTI, V., VERONESI, P., INTRA, M., MAISONNEUVE, P., ZUCCA, F., GATTI, G., MAZZAROL, G., CICCO, C. D., AND VEZZOLI, D. Sentinel lymph node biopsy in breast cancer: ten-year

results of a randomized controlled study. *Annals of Surgery 251*, 4 (Apr 2010), 595–600.

[57] WANG, N., AND DAVIDIAN, M. A note on covariate measurement error in nonlinear mixed effects models. *Biometrika 83*, 4 (1996), pp. 801–812.

[58] WANG, N., LIN, X., GUTIERREZ, R. G., AND CARROLL, R. J. Bias analysis and simex approach in generalized linear mixed measurement error models. *Journal of the American Statistical Association 93*, 441 (Mar. 1998), 249–261.

[59] WARREN, J. L., KLABUNDE, C. N., SCHRAG, D., BACH, P. B., AND RILEY, G. F. Overview of the seer-medicare data: Content, research applications, and generalizability to the united states elderly population. *Medical care 40*, 8, Supplement: Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug. 2002), IV3–IV18.

[60] WARREN, J. L., VIRNIG, B., AND YABROFF, R. Seer-medicare data training workshop (powerpoint presentation), march 8-9 rockville, maryland, 2010.

[61] WHITE, JR, R. L., AND WILKE, L. G. Update on the NSABP and ACOSOG breast cancer sentinel node trials. *The American Surgeon 70*, 5 (May 2004), 420–424.

[62] WILKE, L. G., AND GIULIANO, A. Sentinel lymph node biopsy in patients with early-stage breast cancer: status of the national clinical trials. *The Surgical clinics of North America 83*, 4 (Aug 2003), 901–910.

[63] WOLFINGER, R., AND O'CONNELL, M. Generalized linear mixed models a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation 48*, 3-4 (1993), 233–243.

[64] WU, L. *Mixed effects models for complex data*, vol. 113. Chapman & Hall/CRC Press, Boca Raton, 2010. Lang Wu.; Includes bibliographical references and index.; Monographs on statistics and applied probability ; 113.