

6-12-2008

ANALYSIS OF SUBGROUP EFFECTS IN RANDOMIZED TRIALS WHEN SUBGROUP MEMBERSHIP IS INFORMATIVELY MISSING: APPLICATION TO THE MADIT II STUDY

Daniel O. Scharfstein

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, dscharf@jhsph.edu

Georgiana Onicescu

Medical University of South Carolina

Steven Goodman

The Johns Hopkins School of Medicine

Suggested Citation

Scharfstein, Daniel O.; Onicescu, Georgiana; and Goodman, Steven, "ANALYSIS OF SUBGROUP EFFECTS IN RANDOMIZED TRIALS WHEN SUBGROUP MEMBERSHIP IS INFORMATIVELY MISSING: APPLICATION TO THE MADIT II STUDY" (June 2008). *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 170.
<http://biostats.bepress.com/jhubiostat/paper170>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Analysis of Subgroup Effects in Randomized Trials When Subgroup Membership is Informatively Missing: Application to the MADIT II Study

Daniel Scharfstein

The Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland U.S.A.

Georgiana Onicescu

Medical University of South Carolina, Charleston, South Carolina, U.S.A.

Steven Goodman

The Johns Hopkins School of Medicine, Baltimore, Maryland, U.S.A.

Summary. In this paper, we develop and implement a general sensitivity analysis methodology for drawing inference about subgroup effects in a two-arm randomized trial when subgroup status is only known for a non-random sample in one of the trial arms. The methodology is developed in the context of the MADIT II study, a randomized trial designed to evaluate the effectiveness of implantable defibrillators on survival.

Keywords: Bounds, Expert Opinion, Identifiability, Missing covariates, Selection Model

1. Introduction

Missing data is a hallmark of most clinical trials. Whether it appreciably affects the subsequent inference depends on the extent and pattern of the missing data, the reasons for missingness, and the statistical procedures used to estimate the effect of interest and its uncertainty. The greatest scientific challenge occurs when the data is missing informatively, as the estimates of effect and their uncertainty depend on assumptions not identifiable from the data. However, in biomedical settings, treating clinicians often can provide plausible explanations for the missingness. Sensitivity analyses that incorporate this clinical information can provide statistically and scientifically plausible bounds on effect estimates, facilitating the translation of clinical knowledge into quantitative inference.

In this paper we present an example of a high-stakes medical technology being considered for reimbursement by government agencies and other payors for which the critical clinical trial evidence had a large proportion of what was thought to be a key covariate missing partly by design. We present a statistical approach to incorporate plausible clinical/biological information to see if it could have been useful in this context to represent the uncertainty introduced by this missing data. We also describe what happened when simpler approaches of analysis were used to assess and communicate that uncertainty.

2. Clinical Background

According to the Center of Health Statistics, heart disease is the leading cause of death in the United States. More than half of these deaths are unexpected and occur within one hour after the onset of acute symptoms (Myerburg and Wellens, 2006). Sudden cardiac death

(SCD), as it is referred, claims the lives of roughly 325,000 Americans annually. Mechanistically, SCD is typically initiated by ventricular tachycardia (VT), where the ventricles of the heart beat at a very rapid rate with inadequate pumping, and then progresses to ventricular fibrillation (VF), where the coordinated beating of the heart disintegrates into uncoordinated, ineffective contractions (Winslow, Mehta, and Fuster, 2005). Death then results from cerebral hypoxia (Zipes, 2005). Risk factors for SCD include advancing age, male gender, African-American race, smoking, obesity, hypertension, and hereditary factors, such as hypertrophic cardiomyopathy (thickened heart muscle) and long-QT-interval syndrome (Myerburg and Wellens, 2006).

There is no known way to prevent the abnormalities that result in SCD. Pharmacologic means to reduce VT or VF are only partially successful. The only effective treatment is defibrillation via electrical shock to the heart at the time of an arrhythmic event. In the 1960s and 1970s an implantable, internal defibrillation device was developed and first placed in a human being in 1980. The implantable cardiac defibrillator (ICD) is a device that monitors the heart rate and rhythm continuously, is programmed to recognize VT and VF, and deliver corrective difibrillatory discharges when necessary (Mirowski, Mower, and Reed, 1980). The FDA first approved ICDs in 1985 and Medicare initiated coverage of ICDs in 1989, but only for extremely limited indications. In 1991 coverage was expanded to include those who had an episode of cardiac arrest with ventricular fibrillation (VF) which presumably would have resulted in SCD if defibrillation was not administered. In 1999, Medicare expanded coverage to include those who were deemed at high risk for SCD, defined as either (1) a history of sustained ventricular tachycardia, occurring spontaneously or induced by electrophysiological (EP) testing, or (2) familial or inherited conditions with high risk for VT (e.g. long Q-T interval measured by an electrocardiogram or hypertrophic cardiomyopathy) (Phurrough, Farrell and Chin, 2003).

A series of randomized clinical trials conducted in the 1990's and early 2000's attempted to establish the indications for implantation, i.e., to assess the existence and magnitude of benefit in different populations. The first of these trials assessed ICD benefit in patients who had experienced a near-fatal SCD event, and compared the ICD to medical (drug) management. The largest of these trials, published in 1997, halted after 1016 patients, with an observed 38% reduction in death rate in ICD patients after 18 months mean follow-up (AVID, 1997). Subsequent analyses of this trial (Domanski, 1999) suggested that the benefit was limited to subjects with poor cardiac function (Hallstrom, 2001).

Other trials evaluated the efficacy of ICD therapy in subjects without prior near-SCD episodes, but deemed to be at high risk. The key eligibility criteria for these trials was LVEF below a designated threshold, a documented history of VT, and a patient's susceptibility to an arrhythmia that could induced by electrophysiologic stimulation (EPS) and that was not responsive to drug therapy. EPS is an invasive procedure wherein leads are introduced into the heart during cardiac catheterization, and electrical shocks administered to produce a sustained arrhythmia (i.e. maintained after stimulation ceases) similar to that presumably experienced by a patient before SCD. Electrical stimulation or drug therapy is then used to return the heart to a normal rhythm. Patients in whom such an arrhythmia could not be induced were thought to be at low risk for SCD and were not eligible for these trials.

One of these trials was the Multicenter Automatic Defibrillator Intervention Trial (MADIT), sponsored by Guidant Corporation (Moss *et al.*, 1996). 196 patients were enrolled, and the ICD was shown to have a dramatic effect on all-cause mortality; reducing it from 39% to 16%, with a hazard ratio of 0.46 ($p=0.009$) over a 27-month mean follow-up period. On the basis of this and several other trials, a follow-up trial was initiated, called MADIT

II (Moss *et al.*, 2002). This trial differed from MADIT in that prior EPS testing (and therefore confirmation of an inducible arrhythmia) was not required, nor was a history of an arrhythmia. In addition, the threshold for LVEF was raised from 30% to 35% (i.e., better cardiac function). The sample size was 1232 patients, much larger than MADIT because of the anticipated lower mortality rate. The first patient was enrolled in July, 1997. The trial was monitored by a data safety monitoring board (DSMB) and was stopped by the DSMB due to efficacy in November 2001. With an average follow-up time of 20 months, the ICD group exhibited a statistically and clinically significant 31% reduction in the overall mortality hazard ($p=0.016$), with the proportion of patients who died in the ICD and control arms during the follow-up period being 14.2% and 19.8%, respectively.

Based on the results of MADIT II, the Centers for Medicare and Medicaid Services (CMS) was asked to expand Medicare coverage of ICDs to the MADIT II population. In including patients with better heart function (i.e., $30 < \text{LVEF} \leq 35$), and without prior history of arrhythmia or evidence of inducibility, this represented a substantial expansion of the eligible population. CMS wanted to be assured that the benefit seen in the MADIT II trial was not restricted to patients with inducible arrhythmias, since the observed effect ($\text{HR}=0.69$) was not as sizable as it was in the MADIT I trial ($\text{HR}=0.46$) in inducible patients. An analysis stratified by inducibility status was not possible, since the control patients did not undergo EPS testing as part of the trial, although 12 (2.4%) subjects in control group did have EPS results, which were obtained in 583 (79.3%) of the implanted patients. CMS obtained the MADIT II dataset to see if it was possible to estimate whether the effect of ICDs differed according to inducibility status, even though this covariate was missing for most of the control group.

Interestingly, it is likely that the missing inducible status in the implantation arm is informative. Many clinicians were reluctant to enroll their patients in MADIT II, already being convinced that ICDs benefited patients who met the original MADIT criteria, i.e., who were inducible. So some physicians performed EP testing, and if the patient was found to be inducible, implanted an ICD and did not enroll them in MADIT II. If patients were non-inducible, they were referred to MADIT II, although the preceding EP result was not recorded as part of the trial dataset. EP testing post-randomization (but prior to implantation) was likely to occur only in patients who had not previously had EP testing, making it likely that those who were not EP tested in the implantation group were relatively more likely to be inducible. Thus, it is plausible that missingness of inducibility status in the implantation group was related to inducibility, even after controlling for risk factors (Personal Communication with Dr. Hugh Calkins, Professor of Medicine, Division of Cardiology, Johns Hopkins University).

In this paper, we develop a methodology for evaluating the effectiveness of ICD therapy within inducibility strata that recognizes the fact that inducibility status may be informatively missing. Our approach is an extension of the work of Rotnitzky, Scharfstein and Robins (1998) and Rotnitzky *et al.* (2001) to informatively missing covariates.

The paper is organized as follows. In Section 2, we introduce notation and define the causal effects of interest. In Section 3, we discuss issues of identifiability and introduce our modeling assumptions. Section 5 discusses inference. Section 6 presents a re-analysis of the MADIT II study. The final section is devoted to a discussion, including an epilogue of CMS's decision and how the cardiology community current view of inducibility as an effect modifier.

3. Notation and Data Structure

Consider the following notation for an individual. Let I denote the indicator of inducibility (1 for inducible, 0 for non-inducible). Let T denote the indicator of assignment to the ICD group (1 for ICD, 0 for control). Further, let $Y(t)$ denote the indicator of dying during follow-up if, possibly contrary to fact, the individual had assigned been assigned to treatment group t , $t = 0, 1$. The observed outcome is $Y = Y(T)$. Let X denote the baseline prognostic factors, excluding inducibility. We let V denote a subset of X . Let M denote the indicator of missingness of inducibility status, defined for those in the ICD group.

We consider the observed data for an individual as $O = (T, Y, M : T = 1, I : M = 0, T = 1, X)$. We assume that we observe 1232 i.i.d. copies of O , $\mathcal{O} = \{O_j = (T_j, Y_j, M_j : T_j = 1, I_j : M_j = 0, T_j = 1); j = 1, \dots, 1232\}$.

4. Causal Estimand, Assumptions, and Identifiability

We are interested in estimating the relative risk of dying for ICD vs. control therapy, stratified by inducibility status. The causal estimand is

$$RR_i = \frac{P[Y(1) = 1|I = i]}{P[Y(0) = 1|I = i]},$$

where $i = 0, 1$. Since T is randomized, we know that

$$T \perp (Y(0), Y(1), I, X) \quad (1)$$

Thus,

$$RR_i = \frac{P[Y = 1|T = 1, I = i]}{P[Y = 1|T = 0, I = i]}$$

Since I is missing on a subset of patients in the ICD group and the entire control group (discarded information on 12 subjects in the control group with inducibility status), we need to make assumptions in order to draw inference about RR_i . We start with identification of $P[Y = 1|T = 1, I = i]$.

4.1. Point Identification of $P[Y = 1|T = 1, I = i]$

Through the laws of probability,

$$\begin{aligned} P[Y = 1|T = 1, I = i] &= \frac{\int_v \sum_{m=0}^1 P[I = i|Y = 1, T = 1, V = v, M = m] P[Y = 1, M = m|T = 1, V = v] dF(v|T = 1)}{\int_v \sum_{y=0}^1 \sum_{m=0}^1 P[I = i|Y = y, T = 1, V = v, M = m] P[Y = y, M = m|T = 1, V = v] dF(v|T = 1)} \end{aligned}$$

Notice that all probabilities in this equation are identifiable from the observed data except $P[I = i|Y = y, T = 1, V = v, M = 1]$, for $y = 0, 1$. To identify this quantity, we will assume the following pattern mixture model:

$$P[I = i|Y = y, T = 1, V = v, M = 1] = \frac{P[I = i|Y = y, T = 1, V = v, M = 0] \exp(\alpha i)}{c(y, v; \alpha)} \quad (2)$$

where

$$c(y, v; \alpha) = \sum_{i=0}^1 P[I = i | Y = y, T = 1, V = v, M = 0] \exp(\alpha i)$$

and α is a specified, non-identified constant. When $\alpha = 0$, this model assumes that, for patients in the ICD group with the same V and Y values, the distribution of inducibility status among those with missing inducibility data is the same as those who are not missing their inducibility data. When $\alpha > 0 (< 0)$, this model assumes that, for patients in the ICD group with the same V and Y values, there is a higher (lower) proportion of inducibility among those with missing inducibility data as compared to those who are not missing their inducibility data. The difference between these proportions increases as the absolute value of α increases. As $\alpha \rightarrow +\infty (-\infty)$, the model assumes that everyone with missing inducibility data is inducible (non-inducible). Since the probabilities on the right hand side of (2) are identifiable, specification of α , therefore, identifies $P[Y = 1 | T = 1, I = i]$. Dr. Calkins (see Section 2) makes the case that α is negative.

Using Bayes' rule, it can be shown that model (2) is equivalent to the following selection model:

$$\text{logit } P[M = 1 | T = 1, V = v, Y = y, I = i] = h(v, y) + \alpha i \tag{3}$$

where

$$h(v, y) = \text{logit } P[M = 1 | T = 1, V = v, Y = y] - \log c(y, v; \alpha)$$

Here, α is interpreted as the conditional log odds ratio of having missing inducibility status for an inducible vs. non-inducible patient in the ICD group. Under this formulation, $\alpha = 0$ implies that, within the ICD group, missingness of inducibility status, is under-related to the underlying inducibility status, given X and Y . $\alpha > 0 (< 0)$ implies that, even after adjusting for X and Y , patients who are inducible (non-inducible) are more likely to have missing inducible status than those who are non-inducible (inducible). Notice that, given α , $h(v, y)$ is identifiable.

Under the selection model formulation, we can write the following alternative identification formula:

$$P[Y = 1 | T = 1, I = i] = \frac{E \left[\frac{(1-M)YI(I=i)}{(1+\exp(h(V,Y)+\alpha i))^{-1}} \mid T = 1 \right]}{E \left[\frac{(1-M)I(I=i)}{(1+\exp(h(V,Y)+\alpha i))^{-1}} \mid T = 1 \right]} \tag{4}$$

Also, note that

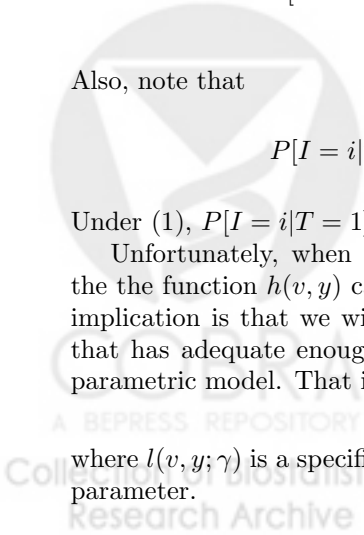
$$P[I = i | T = 1] = E \left[\frac{(1 - M)I(I = i)}{(1 + \exp(h(V, Y) + \alpha i))^{-1}} \mid T = 1 \right] \tag{5}$$

Under (1), $P[I = i | T = 1] = P[I = i | T = 0]$.

Unfortunately, when V is high-dimensional, the curse of dimensionality tells us that the the function $h(v, y)$ cannot be estimated well in small to moderate sized samples. The implication is that we will not be able to obtain an estimator of $P[Y = 1 | T = 1, I = i]$ that has adequate enough precision. As a result, we assume that $h(v, y)$ follows a fully parametric model. That is,

$$h(v, y) = l(v, y; \gamma^*) \tag{6}$$

where $l(v, y; \gamma)$ is a specified function of v, y , and $\gamma \in R^k$ and γ^* denotes the true unknown parameter.



Note that, under model (3), (6), with fixed α ,

$$E[(M - (1 - M)\exp(l(V, Y; \gamma^*) + \alpha I))\phi(V, Y) \mid T = 1] = 0 \quad (7)$$

for all $k \times 1$ -dimensional functions $\phi(V, Y)$. This latter formula will be used as the basis for estimating γ^* .

4.2. Identification Region for $P[Y = 1|T = 0, I = i]$

In the control group, there is no inducibility status data. However, we can identify the $P[Y = 1|T = 0]$ and, given α , the $P[I = i|T = 0]$ (from Section 4.1). By the law of total and conditional probability, we know that

$$P[Y = 1|T = 0] = P[Y = 1|T = 0, I = 1]P[I = 1|T = 0] + P[Y = 1|T = 0, I = 0]P[I = 0|T = 0] \quad (8)$$

Since the left hand side is identified and the mixing probabilities $P[I = i|T = 0]$, the quantities of interest $P[Y = 1|T = 0, I = i]$ are constrained. Without additional modeling assumptions, we can find the minimum and maximum solutions for $P[Y = 1|T = 0, I = i]$. By imposing additional constraints, we can further restrict the range between the minimum and maximum. In consultation with Dr. Hugh Calkins, we imposed the very conservative assumption that

$$0.05 \leq P[Y = 1|T = 0, I = i] \leq 0.50 \quad (9)$$

for $i = 0, 1$.

4.3. Identification Region for RR_i

For each α , the maximum (minimum) value of RR_i is $P[Y = 1|T = 1, I = i]$, identified in (4), divided by the minimum (maximum) solution for $P[Y = 1|T = 0, I = i]$ in (8) subject to constraint (9).

5. Estimation

For specified α and function $\phi(V, Y)$, we can estimate γ^* as the solution, $\gamma_n(\alpha; \phi)$ to the $E_n[U_\gamma(O; \gamma; \phi)] = 0$, where

$$U_\alpha(O; \gamma; \phi) = T(M - (1 - M)\exp(l(V, Y; \gamma^*) + \alpha I))\phi(V, Y)$$

and $E_n[\cdot]$ is the empirical expectation operator. When $\alpha = 0$, it is natural to estimate γ^* by maximum likelihood. We seek to choose $\phi(V, Y)$ so that when $\alpha = 0$, the above estimating equation will yield the maximum likelihood estimator. By taking $\phi(V, Y)$ equal to

$$\phi_n(V, Y) = \frac{l'(V, Y; \gamma_n^{ML})}{1 + \exp(l(V, Y; \gamma_n^{ML}))} \quad (10)$$

where γ_n^{ML} is the MLE for γ^* when $\alpha = 0$ and $l'(V, Y; \gamma)$ is the derivative of $l(V, Y; \gamma)$ with respect to γ , we obtain an estimating function that yields the MLE when $\alpha = 0$. For each α , we estimate γ^* using $U_\alpha(O; \gamma; \phi_n)$.

We estimate $P[Y = 1|T = 1, I = i]$ by

$$P_n[Y = 1, T = 1, I = i] = E_n \left[\frac{(1 - M)I(I = i)}{(1 + \exp(h(V, Y) + \alpha i))^{-1}} \mid T = 1 \right] \quad (11)$$

where $E_n[\cdot|\cdot]$ is the conditional empirical expectation operator.

We estimate the minimum and maximum of $P[Y = 1|T = 0, I = i]$ (found as the solution to (8) subject to constraint (9)), as the minimum and maximum solution to

$$P_n[Y = 1|T = 0] = P[Y = 1|T = 0, I = 1]P_n[I = 1|T = 0] + P[Y = 1|T = 0, I = 0]P_n[I = 0|T = 0] \tag{12}$$

subject to the constraint (9), where $P_n[\cdot|\cdot]$ is the conditional empirical probability operator. Denote the minimum and maximum as $\min P_n[Y = 1|T = 0, I = i]$ and $\max P_n[Y = 1|T = 0, I = i]$

The bounds on the relative risk of death for ICD vs. no treatment for inducibility status i is estimated by

$$\left[\frac{P_n[Y = 1|T = 1, I = i]}{\max P_n[Y = 1|T = 0, I = i]}, \frac{P_n[Y = 1|T = 1, I = i]}{\min P_n[Y = 1|T = 0, I = i]} \right] \tag{13}$$

The variability of our estimators are evaluated using non-parametric bootstrap.

6. Data Analysis

Before we can implement our methods, we must address the issues of missing baseline prognostic factors, V . In the pre-imputation columns of Tables 1 (age and general health), 2 (heart-related), and 3 (medications-related), we show the numbers missing and the observed means/percentages of various baseline characteristics. With the exception of inducibility status, the factors have low levels of missing data. In the ICD and control arms, 89% of patients had their baseline factors (except inducibility) completely recorded. To maximize the generalizability to the original population, we multiply imputed the missing baseline factors.

6.1. Imputation of Baseline Prognostic Factors

We utilized the sequential regression imputation method of Ragunathan *et al.* (2001) as implemented in IVEware (Ragunathan, Solenberger, and Van Hoewyk, 2002) The method assumes that the data are missing at random (Little and Rubin, 2002). The variables listed in Tables 1-3 and the indicator of death were modeled in the imputation procedure. For proper imputation, the joint conditional distribution of the missing variables (17 variables) given the completely observed variables (12 variables) must be specified. The IVEware procedure approximates this distribution by the product of conditional distributions of each variable with missing data given all the other variables, including those with and without missing data. Linear regression with normal errors is used for continuous variables; logistic regression for binary variables; and polytomous logistic regression for categorical variables.

The results are shown in the post-imputation columns in Tables 1, 2, and 3. They are based on the average of 5 imputed datasets. With the exception of inducibility status, the pre- and post- imputation means/percentages are, as expected, almost identical. Further, notice that the treatment groups are well balanced with respect to these factors.

Table 1. Baseline Characteristics, Pre- and Post-Imputation

Variable	Pre-Imputation		Post-Imputation	
	ICD	Control	ICD	Control
Age	64.4	64.6	64.4	64.6
Missing	0	0	-	-
Female	16.0	14.9	16.0	14.9
Missing	0	0	-	-
BMI	27.6	27.7	27.6	27.7
Missing	6	3	-	-
BUN	23.3	23.8	23.2	23.8
Missing	11	10	-	-
Diabetes	33.3	37.7	33.3	37.8
Missing	3	2	-	-
Smoking	79.0	81.9	79.5	81.9
Missing	5	3	-	-

Using the imputed inducibility status, we can, under the assumption of missing at random, estimate the inducibility subgroup effects. In fact, we find that the relative risk of dying for those treated with ICD vs. control is 0.54 (95% CI: [0.29,1.03]) for inducibles and 0.78 (95% CI: [0.56,1.09]) for non-inducibles. These relative risks and their associated confidence intervals were determined using Rubin's rules for combining results across imputed datasets.

6.2. Sensitivity Analysis

In what follows, we address the issue that inducibility status is unlikely to be missing at random. The imputed information on inducibility status and the inducibility status that is measured on the 12 patients in the control group was not used. We used imputed versions of the other baseline prognostic factors. We performed our analysis separately for each of 5 imputed datasets. Since there was essentially no variation across imputed datasets, we simply report the results for the first imputed dataset.

In our analysis, we first selected a model for $l(v, y; \gamma^*)$ in model (3,6) by setting $\alpha = 0$ and including all covariates in Tables 1 - 3 (except inducibility status) that were significant predictors of M at the 0.30 significance level. Death status was retained in the model regardless of its level of significance. The final logistic regression results (when $\alpha = 0$) are displayed in Table 4. Only use of ace inhibitors, presence of angina debicutus, and myocardial infarction were significant (at the 0.05 level) predictors of missing inducibility status in ICD group.

In our analysis, we assumed that α ranged between 0 and -2.0. That is, within levels of covariates listed in Table 4, we assumed that non-inducible patients in the ICD group had equal to 7.4 times the odds of having missing inducibility status as inducible patients in the ICD group. Figure 1 displays estimates (along with 95% pointwise confidence interval) of the probability of inducibility as a function of α for the first imputed dataset. The estimates range from a maximum of 40.0% (95% CI: [0.35,0.45]) at $\alpha = -2.0$ to a minimum of 35.6% (95% CI: [0.32,0.39]) at $\alpha = 0.0$.

In Figure 2, panels (a) and (b) display, respectively, present estimates (along with 95% pointwise confidence intervals) for the probability of dying under ICD for inducibles and non-

Table 2. Baseline Heart-Related Characteristics

Variable	Pre-Imputation		Post-Imputation	
	ICD	Control	ICD	Control
DBP	71.1	70.4	71.1	70.4
Missing	4	3	-	-
SBP	122.4	120.8	122.4	120.8
Missing	4	3	-	-
EF	23.1	23.2	23.1	23.2
Missing	0	0	-	-
Heart Rate	72.5	71.9	72.5	71.9
Missing	5	3	-	-
QRS Interval	0.12	0.12	0.12	0.12
Missing	11	8	-	-
CHFNYHA				
Grade 1	10.3%	12.1%	10.3%	12.0%
Grade 2	35.3%	33.8%	35.5%	33.8%
Grade 3	25.6%	23.1%	25.3%	23.1%
Grade 4	4.5%	4.2%	4.5%	4.2%
No CHF	24.5%	26.9%	24.4%	26.9%
Missing	10	10	-	-
Angina				
Grade 1	17.2%	16.8%	17.3%	16.7%
Grade 2	15.3%	17.4%	15.3%	17.3%
Grade 3	5.7%	6.2%	5.7%	6.2%
Grade 4	1.9%	1.2%	2.0%	1.2%
Decubitus	4.8%	2.9%	4.8%	3.0%
None	55.1%	55.5%	55.0%	55.6%
Missing	10	7	-	-
Atrial Arrhythmia	27.7%	25.4%	27.6%	25.5%
Missing	17	13	-	-
Hypertension	53.0%	53.5%	53.1%	53.6%
Missing	6	2	-	-
Inducible	36.0%	66.7%	36.2%	38.9%
Missing	159	478	-	-
Myocardial Infarction	12.4%	12.6%	13.0%	13.0%
Missing	40	31	-	-
Non CABG revascularization	45.1%	45.2%	45.1%	42.1%
Missing	11	6	-	-
Ventricular Arrhythmia	10.3%	13.5%	10.4%	13.6%
Missing	21	14	-	-

Table 3. Baseline Medications, Pre- and Post Imputation

Variable	Pre-Imputation		Post-Imputation	
	ICD	Control	ICD	Control
ACE inhibitors	77.4%	77.4%	77.4%	77.4%
Missing	0	0	-	-
Aspirin	67.8%	70.2%	67.8%	70.2%
Missing	0	0	-	-
Beta Blockers	63.2%	60.4%	63.2%	60.4%
Missing	0	0	-	-
Coronary bypass surgery	57.8%	55.9%	58.0%	56.1%
Missing	3	1	-	-
Calcium Channel Blockers	11.9%	14.3%	11.9%	14.3%
Missing	0	0	-	-
Digitalis	59.6%	56.5%	59.6%	56.5%
Missing	0	0	-	-
Diuretic agents	72.9%	77.6%	72.9%	77.6%
Missing	0	0	-	-
Lipid Lowering agents	66.3%	64.5%	66.3%	64.5%
Missing	0	0	-	-
Lipid Lowering Statins agents	63.5%	61.8%	63.5%	61.8%
Missing	0	0	-	-

Table 4. Inference about γ^* when $\alpha = 0$

Parameter	Estimate	Standard Error	95% Confidence Interval
Intercept	-1.14	0.69	(-2.49,0.22)
EF	0.02	0.02	(-0.01,0.06)
QRS	-3.41	2.85	(-9.00,2.18)
CHFNYHA - 1	-0.26	0.37	(-0.99,0.47)
CHFNYHA - 2	0.42	0.25	(-0.07,0.92)
CHFNYHA - 3,4	0.06	0.28	(-0.49,0.60)
Angina - 1	0.05	0.26	(-0.45,0.55)
Angina - 2	-0.06	0.27	(-0.59,0.46)
Angina - 3,4	0.61	0.33	(-0.04,1.25)
Angina - Decubitis	-1.11	0.62	(-2.33,-0.10)
Atrial Arrhythmia	-0.32	0.22	(-0.76,0.12)
Myocardial Infarction	0.65	0.27	(0.12,1.19)
Non CABG Revascularization	0.22	0.19	(-0.15,0.60)
Ace Inhibitors	-0.49	0.21	(-0.90,-0.08)
Beta Blockers	-0.27	0.20	(-0.65,0.12)
Lipid Lowering Agents	-0.75	0.66	(-2.03,0.54)
Lipid Lowering Statins	0.71	0.64	(-0.55,1.98)
Death	0.12	0.28	(-0.42,0.66)

Fig. 1. Estimates (along with 95% pointwise confidence interval) of the probability of inducibility as a function of α for the first imputed data set.

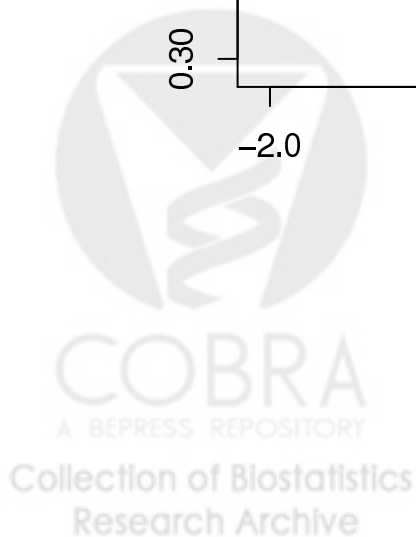
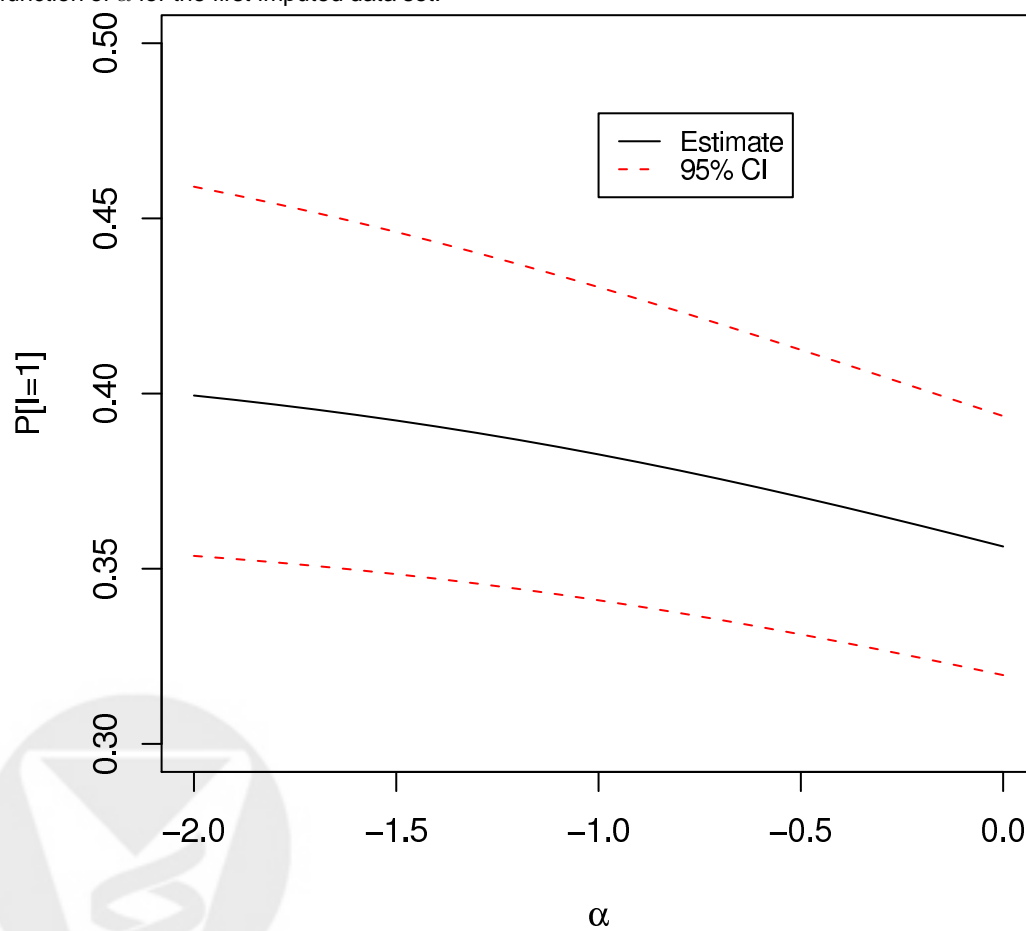
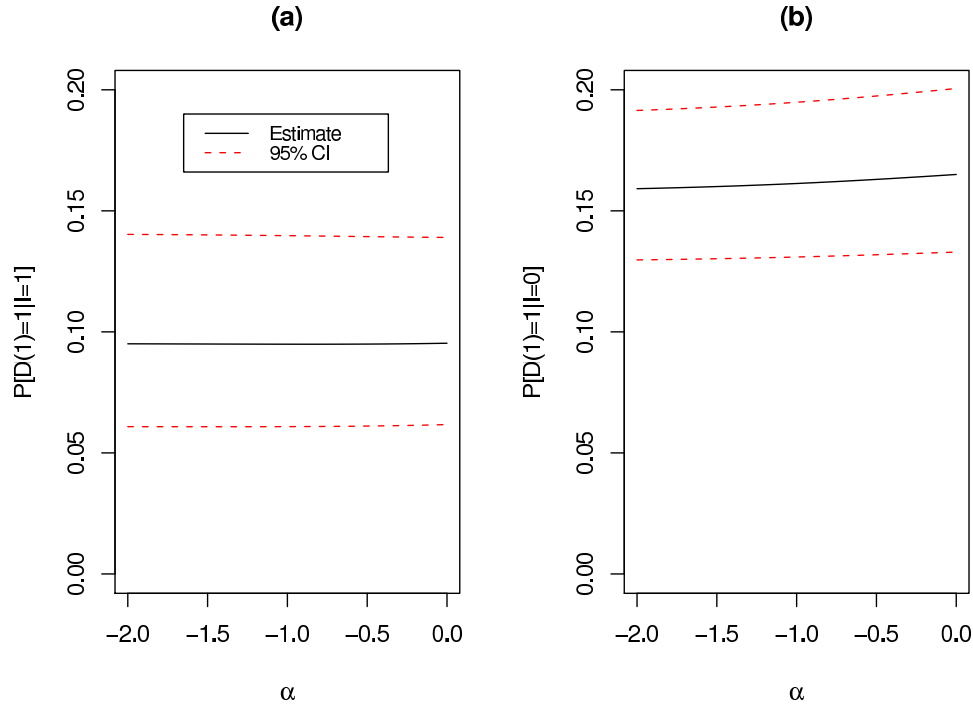


Fig. 2. Panels (a) and (b) display, respectively, estimates (along with 95% pointwise confidence intervals) of the probability of death under ICD for inducibles and non-inducibles, respectively, as a function of α for the first imputed dataset.



inducibles as a function of α . For inducibles, the estimates and confidence intervals for the probability of dying under ICD are constant functions of α up to the second decimal place. Specifically, the probability of death is 9.3% (95% CI: 6.1%,14.0%). For non-inducibles, the probability of dying under ICD varies slightly from a minimum of 15.9% (95% CI: [13.0%, 19.2%]) at $\alpha = -2.0$ to a maximum of 16.5% (95% CI: [13.3%, 20.1%]). It is important to highlight that the probability of dying under ICD is estimated to be lower for inducibles than non-inducibles.

Panels (a) and (b) in Figure 3 display the estimated bounds and 95% pointwise confidence intervals for the bounds on the probability of dying under standard care for inducibles and non-inducibles, respectively, for the first imputed dataset. At each α , the 95% confidence interval is interpreted as an interval that would wholly contain, under repeated sampling, the true interval 95% of the time. The interval was formed by using non-parametric bootstrap and a bi-section grid search. In panel (a), we see that the bounds are [6.0%,42.0%] (95% CI: [5.0%,49.5%]) at $\alpha = -2.0$ and [6.8%,46.5%] (95% CI: [5.0%,49.9%]) at $\alpha = 0$ for inducibles. So, for inducibles, our data do not further constrain the probability of dying under standard therapy, above and beyond the constraints imposed by Dr. Calkins. In panel (b), we see that the bounds for the probability of dying is estimated as [5.0%, 29%] (95% CI: [5.0%, 35%]) at $\alpha = -2.0$ and [5.0%,27%] (95% CI: [5.0%, 33.0%]). For non-inducibles, our data further constrain the upper bound (relative to the 50% conservative prior upper bound) on the probability of dying under standard therapy.

In panels (c) and (d) of Figure 3, we display the relative risks (ICD vs. standard therapy) of dying for inducible and non-inducible patients, respectively. It is estimated that the relative risk for inducibles lies between 0.23 and 1.60 (95% CI: [0.13, 2.47] at $\alpha = -2.0$ and between 0.20 and 1.40 (95% CI: [0.12, 2.36]) at $\alpha = 0$. For non-inducibles, the equivalent bounds are [0.55, 3.18] (95% CI: [0.41, 3.86]) and [0.61, 3.30] (95% CI: [0.45, 3.88]). For the first imputation and for all α , 100% of the bootstrapped samples had overlapping intervals for the relative risks for inducibles and non-inducibles.

These results demonstrate, in the context of the prior assumptions and our sensitivity analysis based model, that the data from MADIT II do not provide evidence of a differential effect of ICD therapy for inducible vs. non-inducible patients.

In discussions with Dr. Calkins, we also considered the additional constraint:

$$P[Y(0) = 1|I = 1] \geq P[Y(0) = 1|I = 0] \quad (14)$$

That is, prior to the MADIT II, it was generally believed that inducible patients were at higher risk of death than non-inducible patients, under standard therapy. Figure 4 shows, for the first imputed dataset, the results when this additional constraint is imposed. In panel (c), we see that the relative risk of dying for inducibles lies between 0.23 and 0.46 (95% CI: [0.13, 0.68]) at $\alpha = -2.0$ and between 0.20 and 0.45 (95% CI: [0.12, 0.67]) at $\alpha = 0.0$. These bounds are much tighter than ones that did not impose the above constraint (compare Figure 3c with Figure 4c). For non-inducibles, the bounds for the relative risk of dying are estimated as [0.84, 3.18] (95% CI : [0.63, 3.84]) at $\alpha = -2.0$ and [0.87, 3.30] (95% CI: [0.65, 3.88]) at $\alpha = 0$. These bounds are only slightly tighter than the ones without the above constraint (compare Figure 3d with Figure 4d). Here, for all α , less than 0.25% of the bootstrapped intervals for the relative risks of dying for inducibles and non-inducibles overlap. With the additional constraint, we would be able to conclude that there is strong evidence to suggest that ICDs are more effective for inducibles as compared to non-inducibles.

7. Discussion

In this paper, we demonstrated a general sensitivity analysis methodology for drawing inference about subgroup effects in a two-arm randomized trial when the subgroup status is only known for a non-random sample of one of the trial arms. Our methodology allows the incorporation of scientific constraints to increase the precision of the inferences. The methodology will be useful in settings where an intervention affords the opportunity to collect risk factors, which are not available otherwise. This may occur when the intervention is invasive.

We generally advocate the use of sensitivity analysis and bounds in settings where the parameters of primary of interest are not identifiable without strong untestable assumptions. We believe that such an approach is useful in conveying the additional uncertainty in the analysis that is generated above and beyond sampling variability.

7.1. CMS Process

The Medicare Advisory committee met on February 12 , 2003 in order to evaluate the scientific evidence of the effectiveness of the ICD's in Medicare patients, to analyze the external validity of MADIT II and to characterize the magnitude of the effect size for

Fig. 3. Panels (a) and (b) display the estimated bounds and 95% pointwise confidence intervals for the probability of dying under standard therapy for inducibles and non-inducibles, respectively as a function of α , for the first imputed dataset. Panels (c) and (d) displays display the estimated bounds and 95% pointwise confidence intervals for the relative risk of dying for ICD vs. standard therapy for inducibles and non-inducibles, respectively as a function of α , for the first imputed dataset.

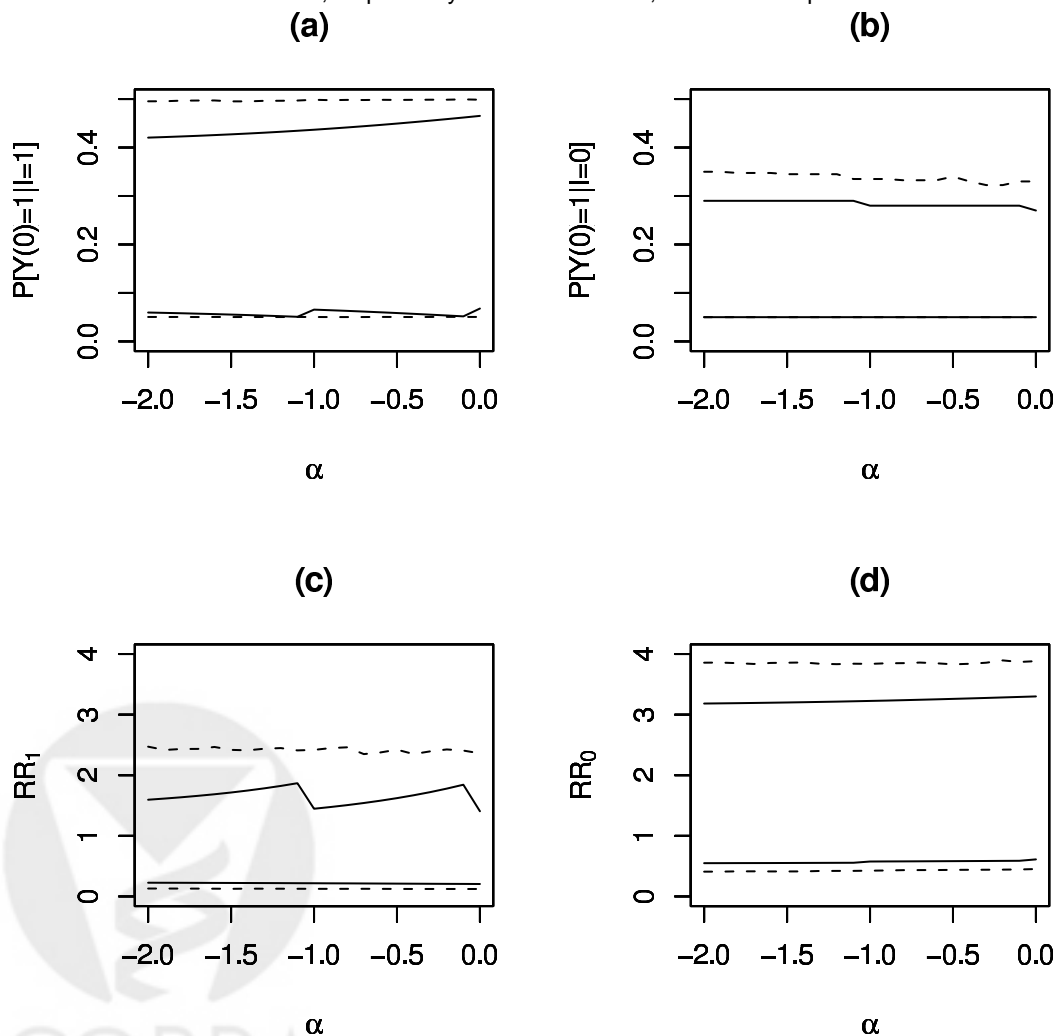
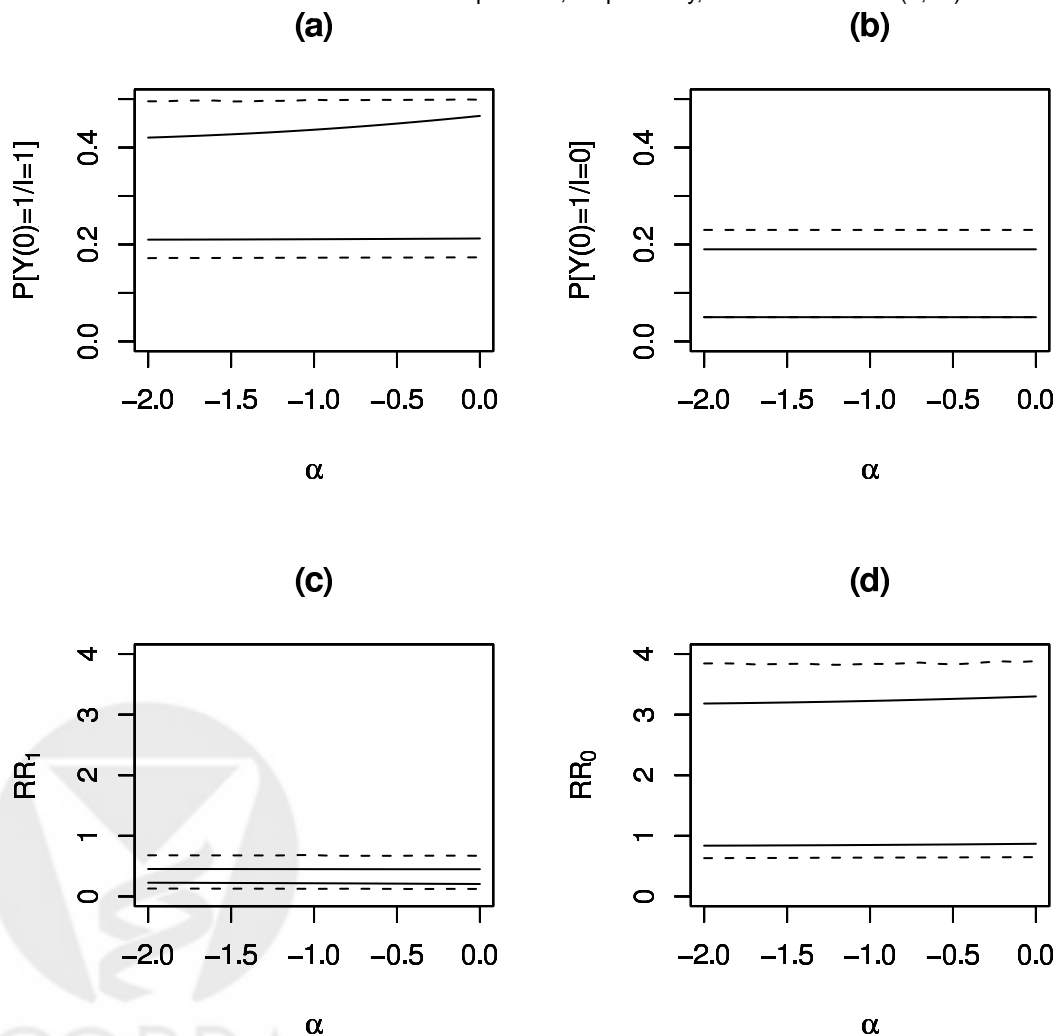


Fig. 4. Panels (a) and (b) display the estimates and 95% pointwise confidence intervals for the probability of dying under standard therapy as a function of α , for the inducible and non-inducible groups, respectively, under constraints (9,10). Panels (c) and (d) display the estimated bounds and 95% pointwise confidence intervals for the relative risk of dying for ICD vs. standard therapy as a function of α for inducible and non-inducible patients, respectively, under constraints (9,10).



different populations of patients. The goal of the discussion and decisions was to provide the evidence to CMS who had to make a coverage recommendation for the ICD. CMS was particularly interested in whether there were differential effects of ICD's by inducibility status. One of the authors (SNG) was asked to assist in the subgroup analysis. His analysis used data on patients who had EP testing to develop a prediction model for inducibility given baseline factors. He found that NYHA functional class, ejection fraction, heart rate, and blood urea nitrogen (BUN) were the strongest predictors, albeit with a ROC curve AUC of 0.65, indicating that the predictors of inducibility status were not highly informative. He used this model to multiple impute inducibility status for those in the control group and those in the ICD group without EP testing results. His analysis assumed that missingness of inducibility status was unrelated to death or inducibility status conditional on NYHA functional class, ejection fraction, heart rate, and blood urea nitrogen (BUN). The results of this MAR analysis showed an absolute decrease in mortality of -9.5% (95% CI, -16.9% to -2.2%) in the inducible group and -3.6% (95% CI, -9.1% to 2.0%) in the non-inducibles, from a baseline mortality of about 20% in each. The difference between the two effects was estimated as -5.9% (-15% to 3.1%). Based on this, he concluded

“that the data strengthened the finding from MADIT I that inducible patients experience a substantial benefit from ICDs, that it provides weak to moderate evidence that the ICD effect is greater in inducible than in non-inducible patients, and that it would be an error to interpret the results as indicating no benefit in the non-inducible group. The adjudged strength of the evidence for an ICD effect in non-inducibles must come from a qualitative, biologic judgment about the similarity of the physiologic mechanism producing the treatment effect in the two types of patients. If similar, the evidential strength and treatment effects lie somewhere between the separate and combined results.” (Anderson, 2003)

Our most restrictive analysis (based on assumptions 9 and 14) is consistent with this conclusion; however our less restrictive analysis (based on assumption 9) is less so. Using the approach developed in this paper would require the members of the scientific advisory board to specify their beliefs regarding assumptions 9 and 14.

Although Goodman's analysis and interpretation supported the company position, it was widely misperceived. A Washington Post editorial written several months after the meeting, Berger (2003) condemned this analysis, saying, “I attended the proceedings and was appalled by what I witnessed. CMS presented a specious analysis to discredit the rigor of the study's landmark findings. The agency was manipulating data in an effort to limit access to the therapy.”

Guidant Corporation argued that coverage should not be based on inducibility status, and made available at the hearing previously undisclosed pre-enrollment inducibility status on 257 patients, 113 randomized to standard therapy and 114 to the ICD arm (Anderson, 2003). For this possibly non-random subset of patients, they reported that 19.5% of inducibles died under conventional therapy versus only 9% in the ICD group, yielding a relative risk of death for non-inducibles of 0.54. No confidence intervals were reported in the meeting minutes. The point estimate is consistent with our bounds generated under constraint (9) and inconsistent with the bounds under constraints (9,14).

Panel members tended to find the analysis of the pre-randomization inducibility data compelling, and unanimously voted that the “evidence is adequate to draw conclusions about health outcomes in patients identical to the patients enrolled in the MADIT II trial.”

Risk stratification based on inducibility status was not part of their recommendation (Anderson, 2003).

On June 6, 2003, a CMS decision memorandum stated CMS's intent to expand the 1999 coverage criteria to include patients with previous MI, LVEF less than or equal to 30% and a QRS duration greater than 120 milliseconds, a subset of the MADIT II population (Phurrough, Farrell, and Chin, 2003). Inducibility was not part of their coverage decision. It has been argued that QRS duration, a factor that was collected routinely in MADIT II, was added as an restriction to reduce the financial impact of allowing coverage for all MADIT II-type patients (Reynolds and Josephson, 2003) until further clinical trial information could be obtained. The restriction was removed by CMS in September, 2004 after the release of findings from a subsequent trial (SCD-HeFT) that supported the MADIT II results, although subsequent CMS coverage was made contingent on the institution of a registry of all ICD users, with the ostensible purpose of facilitating further research and refining ICD indications in the future (Phurrough *et al.*, 2004).

7.2. Final Thoughts

This story highlights the difficulty of both calculating and communicating quantitative results that reflect both sampling and non-sampling uncertainty in a policy arena. Those who most need to understand and process the quantitative information often have difficulty doing so. The methodology proposed in this paper would require individuals to think about the underlying assumptions in the analyses, to formalize their prior beliefs, and hopefully understand better the impact of these priors on conclusions, as well as present conclusions that reflect current biologic and clinical understanding.

References

- Anderson, J. (2003). Medicare coverage advisory committee meeting for implantable cardioverter defibrillators (icds) (cag-00157n). <http://www.cms.hhs.gov/mcd/viewmccac.asp?id=39> (Accessed March 7, 2006).
- Berger, R. (2003). Life-and-death federal meddling. *Washington Post*, page A29.
- Domanski, M., Sakseena, S., Epstein, A., Hallstrom, A., Brodsky, M., Kim, S., Lancaster, S., Schron, E., and Investigators, A. (1999). Relative effectiveness of the implantable cardioverter-defibrillator and antiarrhythmic drugs in patients with varying degrees of left ventricular dysfunction who have survived malignant ventricular arrhythmias. *Journal of the American College of Cardiology*, 34:1090–1095.
- Hallstrom, A., McAnulty, J., Wilkoff, B., Follmann, D., Raitt, M., Carlson, M., Gillis, A., Shih, H.-T., Powell, J., Duff, H., Halperin, B., and Investigators, A. (2001). Patients at lower risk of arrhythmia recurrence: A subgroup in whom implantable defibrillators may not offer benefit. *Journal of the American College of Cardiology*, 37:1093–1099.
- Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. John Wiley, New York.
- Meyerbrug, R. and Wellens, H. (2006). Epidemiology of cardiac arrest. In Priori, S. and Zipes, D., editors, *Sudden Cardiac Death*, pages 3–20. Blackwell Publishing, Malden, Massachusetts.

- Mirowski, M., Mower, M., and Reid, P. (1980). The automatic implantable defibrillator. *American Heart Journal*, 100:1089–1092.
- Moss, A., Hall, W., Cannom, D., Daubert, J., Higgins, S., Klein, H., Levine, J., Saksena, S., Waldo, A., Wilber, D., Brown, M., and Heo, M. (1996). Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *New England Journal of Medicine*, 335:1933–1940.
- Moss, A., Zareba, W., Hall, W., Klein, H., Wilber, D., Cannom, D., Daubert, J., Higgins, S., Brown, M., and Andrews, M. (2002). Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *New England Journal of Medicine*, 346:877–883.
- Phurrough, S., Farrell, J., and Chin, J. (2003). Centers for medicare and medicaid services decision memorandum: Implantable cardioverter defibrillators (cag 00157n). <http://www4.cms.hhs.gov/mcd/viewdecisionmemo.asp?id=39>.
- Phurrough, S., Salive, E., Farrell, J., and Chin, J. (2004). Centers for medicare and medicaid services decision memorandum: Implantable cardioverter defibrillators (cag 00157r2). <https://www.cms.hhs.gov/mcd/viewdraftdecisionmemo.asp?from2=viewdraftdecisionmemo.asp&id=139&>.
- Raghunathan, T., Lepkowski, J., Van Hoewyk, J., and Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27:85–95.
- Raghunathan, T., Solenberger, P., and Van Hoewyk, J. (2002). Iweware: Imputation and variance estimation user guide. Technical report, Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan, <http://www.isr.umich.edu/src/smp/ive/>.
- Reynolds, M. and Josephson, M. (2003). Debate: Risk stratification, costs, and public policy: madit ii (second multicenter automated defibrillator implantation trial). *Circulation*, 108:1779–1783.
- Rotnitzky, A., Robins, J., and Scharfstein, D. (1998). Semiparametric regression for repeated outcomes with non-ignorable non-response. *Journal of the American Statistical Association*, 93:1321–1339.
- Rotnitzky, A., Scharfstein, D., Su, T., and Robins, J. (2001). A sensitivity analysis methodology for randomized trials with potentially non-ignorable cause-specific censoring. *Biometrics*, 57:103–113.
- Winslow, R., Mehta, D., and Fuster, V. (2005). Sudden cardiac death: Mechanisms, therapies, and challenges. *Nature Clinical Practice: Cardiovascular Medicine*, 2:353–360.
- Zipes, D. (2005). Epidemiology and mechanisms of sudden cardiac death. *Canadian Journal of Cardiology*, 21:37A–40A.
- Zipes, D., Wyse, D., Friedman, P., Epstein, A., Hallstrom, A., Greene, H., Schron, E., and Domanski, M. (1997). A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. *New England Journal of Medicine*, 337:1576–1584.